executive insights into the world of large-scale data management

WinterCorp

scaling up

WINTER 2006

IN THIS ISSUE

(Navigation Hint: Click on any title below to jump to that article)

HOW TO SELECT THE BEST DATABASE PLATFORM TO MEET GROWING REQUIREMENTS

by Rick Burns

Taking a fact-based, engineering-oriented approach to the product evaluation process can lower project risk, drive sound purchasing decisions, and maximize ROI.

INFORMATION: The more we get, the more we want. Fueled by the internet, data capture technologies, automated business processes, and other sources, our business enterprises can now access huge volumes of fine-grained, real-time data. And still our appetite for information is insatiable—often pushing the limits of today's data management products.

This corporate craving demands a database platform that can grow or "scale" as data volume increases. But meeting ever-expanding data requirements is only part of the solution. Choosing a truly scalable platform—one that your business won't outgrow—is actually a multi-faceted challenge. That's because data volume is just one of many factors—or "dimensions"—that affect a platform's ability to scale. Each of these dimensions must be carefully considered during the platform selection process. In fact, failure to do so is one of the leading reasons why product evaluations fail.

PLATFORMS, continues on page 3

INDUSTRY NEWS BRIEFS

by Robert Dorin

Teradata Attempts to Raise the Bar in 2006

Teradata defined the Active Data Warehouse (ADW) in October 2003 as "an integrated information repository to drive strategic and tactical decision support within an organization"—taking operational and analytical capabilities to the next level with automated, event-based triggers. In 2005, Teradata presented the integration of the ADW into the Real-Time Enterprise (RTE), where data latency is minimized and business processes operate in near real-time—and developed a Reference Architecture for Real-Time Enterprise Data Warehousing based on Service-Oriented Architecture (SOA). The latest Teradata vision is Enterprise Intelligence driving RTE enabled by ADW. To Teradata's credit, it has a lead in tackling the hardest, most complex problems in data warehousing. But the application of the blueprint in any particular customer environment is a challenge. Customers

INDUSTRY BRIEFS, continues on page 6

WinterCorp

Scaling Up is an independent, vendor-neutral newsletter for executive-level owners and sponsors of terabyte-scale data management systems that must support business-critical objectives and enable enterprise growth.

Advertising is not accepted.

PUBLISHER
WinterCorp

EXECUTIVE EDITOR
Richard Winter

CONTRIBUTING WRITERS
Rick Burns
Robert Dorin

MANAGING EDITOR
Trish Baumann

ART DIRECTOR
Kimberly Foss
Ampers&® Studio

FREE SUBSCRIPTIONS
Log onto www.wintercorp.com and click "Subscribe now" or call 781-642-0300, x132.

COMMENTS, QUERIES & SUGGESTIONS
Email ScalingUp@wintercorp.com or call 781-642-0300, x132

ABOUT WINTERCORP
WinterCorp is an independent consulting firm that specializes in technology near, at and beyond the frontier of database scalability. Since its inception in 1992, WinterCorp has architected many of the largest and most challenging databases in production today. Its client services help organizations define business-critical database solutions, engineer their implementations, and manage their scalability to support growth and optimize their business value.

www.wintercorp.com

FROM THE EDITOR

Are all database products created equal? Or are some better than others?

INDUSTRY PUNDITS WOULD LEAD YOU TO BELIEVE that database management has reached commodity status, with little or no clear, functional difference from one product to the next. While this may be true for small workgroup-class database applications, the facts speak otherwise for multi-terabyte data warehouses and other enterprise-class systems.

Richard Winter

Time after time, tests have shown that the capabilities of different product implementations can differ markedly at the multi-terabyte level—and change significantly as scale continues to grow.

In short, all database products are not created equal—at least in the large scale arena. Some serve many concurrent users well. Others handle large databases well. Some do a better job with complex queries. Others are better at parallelism. Still others may excel at mixed workload management. But no one product is "best" for everything. And all have their scaling limits.

This is a crucial tenet for executive owners and sponsors of terabyte-scale data management systems. Such systems are major financial and engineering commitments with the potential to yield up to billions in business value. The platform (hardware, operating system and database engine) you choose can have a huge impact on the time, cost and effort needed to reach your goals. It can dictate your system's ability to accelerate time to value…perform as needed to meet business objectives …adjust to unexpected developments along the way…and scale to support growing analytical requirements. It can mean the difference between success and failure.

Given these high stakes, don't be fooled by those who make sweeping generalizations that minimize the impact of platform selection on data management success. Because the differences in database engines grow so large at higher levels of scale and complexity, platform selection is a strategic decision—one that is disruptive and costly to change. All database products are not equal because all large scale databases do not have the same engineering and business requirements. When it comes to enterprise-class systems, it pays to take a thoughtful approach to platform selection like the one described in this issue's feature article.

As always, thanks for reading *Scaling Up*. Your comments and suggestions are always welcome.

Sincerely,

Richard Winter
Executive Editor
richard.winter@wintercorp.com

Press Clippings

Check out Charles Babcock's article ("Data, Data, Everywhere") in the January 9th issue of ***Information Week*** for tips on managing mega-terabyte databases. The article contains ideas from the data masters at Wal-Mart, EBay and Nielsen Media Research—as well as insights from Richard Winter of WinterCorp. The article is available at **www.informationweek.com/news/showArticle.jhtml?articleID=175801775**.

SCALABILITY, from page 1

GET TO KNOW THE DIMENSIONS OF SCALABILITY

Before setting out to discover the best platform that can scale to your growing needs, assess the following dimensions of scalability as they relate to your business objectives.

First, be sure to understand your ***data volume*** requirement. This is the most obvious measure of scalability and one which clearly illustrates the scope of the scalability challenge. Today's largest databases contain more than 100 terabytes (TB) of user data according to the 2005 WinterCorp TopTen™ Program, an independent survey of the world's biggest and most heavily used databases. This finding continues a longstanding trend of rapid growth which shows no signs of slowing down. In fact, the WinterCorp survey found that the world's largest databases have posted an annual compounded growth rate of approximately 75% since 1995.

Next, evaluate the ***complexity of your data***. This can have as big an impact on scalability as database size. Some very large databases have relatively simple schemas. For example, large telecommunications firms mine repositories of call detail records for consumer calling patterns. In this case, the bulk of the data is in a single very large table. But companies more frequently want to cross-reference data from disparate sources. This leads to more complex schemas—often with thousands of tables and many complex relationships among them. Such sophisticated schemas are increasingly typical at financial and insurance institutions.

Then anticipate the impact of new business initiatives on your ***user population***—and plan accordingly. The size of the user community climbs as new data sources support the initiatives of new users—and as companies open their databases to customers and suppliers. Enterprise-class databases commonly support thousands of concurrent users.

Finally, today's competitive pressures demand sophisticated processes to extract more granular and subtle information from the accumulated data. The winning edge goes to the "early bird"—the business that can react more quickly to new circumstances. To compete successfully, companies rely on a ***higher volume and complexity of queries***, combined with reduced ***data latency*** (i.e., faster updates to ensure near-real time data). These additional dimensions play a big role in determining scalability because they contribute to the explosive growth in database workload—a fact which makes it crucial to factor them into the platform selection process.

• • •

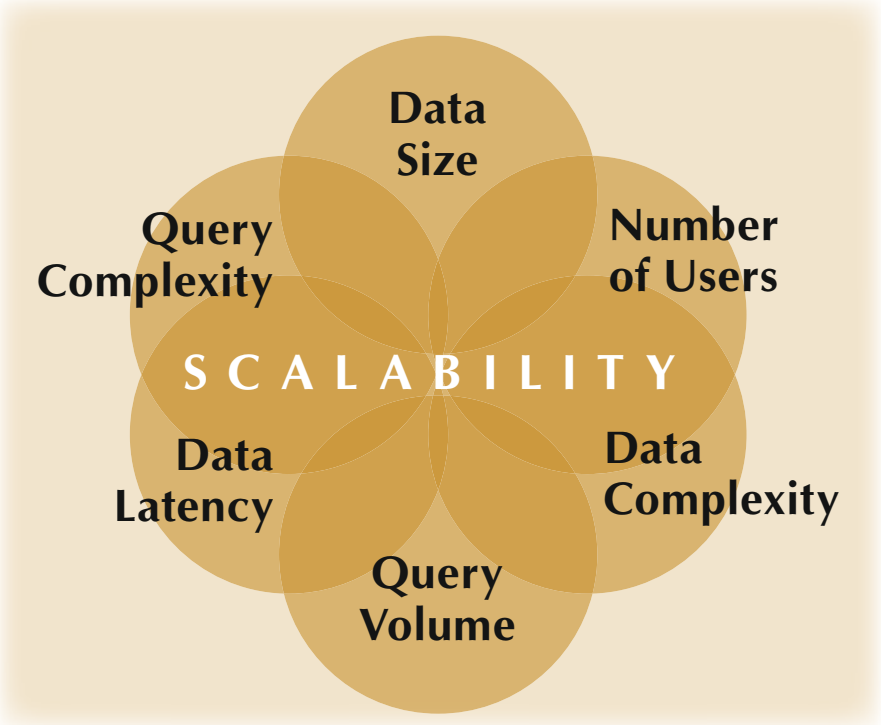
Doubling your data, users and queries while reducing latency may increase scalability demands more than ten times—dragging database performance to a standstill.

• • •

Any one of these scalability dimensions can significantly ramp up database requirements. But combining them exerts a multiplying effect. For example, experience proves that doubling the database size, user population and query volume while reducing latency may increase scalability demands more than ten times. Without anticipating and managing such scalability challenges, companies put their data management systems at risk—along

SCALABILITY, continues on page 4

DIMENSIONS OF DATA WAREHOUSE SCALABILITY



SCALABILITY, from page 3

with all the business-critical operations and strategic initiatives they support.

WHY DATABASE PRODUCT EVALUATIONS FAIL

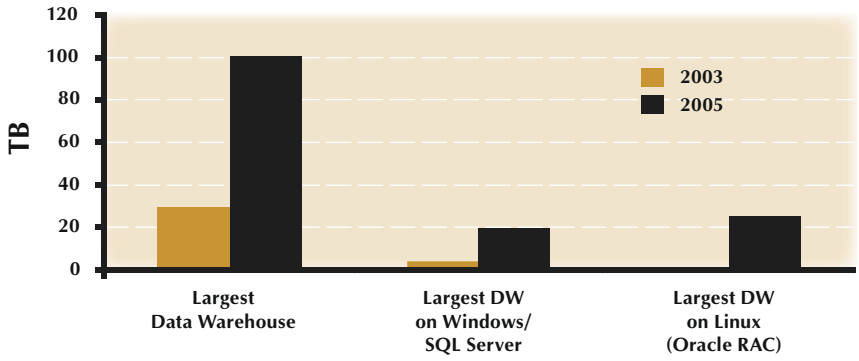
So how do you go about selecting the best database platform to meet your growing needs? The myriad technology choices available make it a challenging proposition. This is even more true when the scalability requirements of your project extend beyond the ordinary experience of any given computing environment—whether UNIX, Windows, Linux or mainframe. When properly conducted, however, database evaluations can be successfully navigated and can ultimately create or extend strategic vendor partnerships.

But many evaluations fail, usually for one of the following reasons:

- The evaluation does not detect that the products or configurations being considered cannot meet the long-term business requirements of the project. This may be because the evaluation team did not understand the impact of business objectives on data volume, data complexity, user population, query volume, query complexity and data latency—the dimensions of scalability which can potentially strain system performance and impact business-critical operations. Or it may be because the evaluation failed to sufficiently account for the effects of scale on database performance. Either case leads to unanticipated and costly upgrades once scalability issues arise in production—or worse, prevents your system from meeting key business objectives.
- Product configurations are not comparable, so price comparisons are meaningless.
- The evaluation does not lead the bidding vendors to commit to solving a specific business problem, leaving the evaluation team with little or no leverage.
- The evaluation does not elicit meaningful information about specific product capabilities or product directions.

NEW PLATFORMS FOR LARGE DWs EMERGE IN 2005

The 2005 WinterCorp TopTen™ Program noted the emergence of SQL Server and Linux as platforms for large data warehouses. While the largest data warehouse in the survey grew to more than 100 TB (from 29 TB in 2003), the largest SQL Server database grew from 1.6 TB to 19.5 TB. Linux appeared for the first time, with the largest Linux data warehouse as a 24.8 TB Oracle cluster.



• • •

A modest investment in a fact-based, engineering-oriented approach to product selection pays off handsomely in lowering project risk and ensuring a scalable solution that meets requirements.

• • •

TAKE A FACT-BASED APPROACH TO PLATFORM SELECTION

Each of these potential failures can be traced to a common cause—lack of a fact-based, engineering-oriented approach to evaluating database products.

This approach is grounded in a crucial tenet of platform selection: Success hinges on the ability to derive quantitative scalability requirements from the business objectives of the project. To get a sense of the size and structure of the data, ask where it comes from, what do the data structures look like, how are they related, how much data is there, how fast does it change, and how fast is it growing. To understand the workload implications of the project, find out what questions need to be asked of the data, who asks them, how often they ask, and how usage patterns will change with time. To determine availability requirements, find out how various types of system failure affect your users.

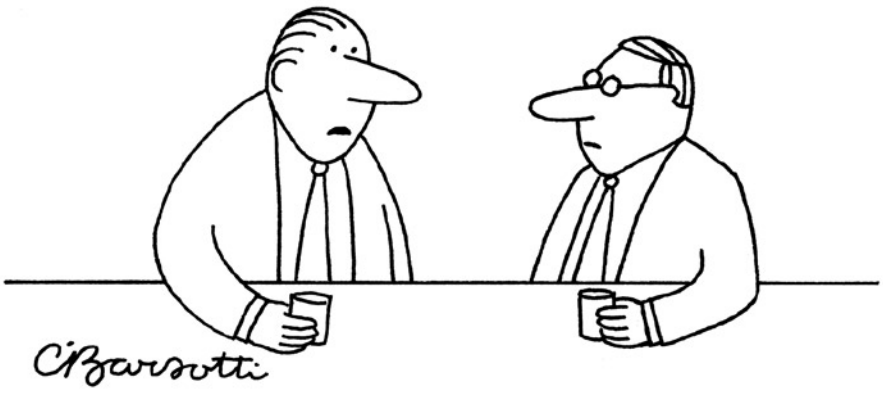
Once you have developed these technical requirements, you have the information in hand to qualify bidders. You can then ask all qualified vendors to propose configurations that satisfy these requirements. And you can insist that bidders prove that their proposed configurations will meet the requirements.

SCALABILITY, continues on page 6

CALENDAR OF UPCOMING IT EVENTS

FOCUS	DATE	EVENT
Technical	March 1 – 3, 2006	Oracle OpenWorld 2006, Tokyo, Japan
Executive	March 5 – 7, 2006	ComputerWorld Premier 100: Mastering IT Agility, Accelerating Business Innovation, Palm Desert, CA
Technical	March 23 – 24, 2006	IDUG 2006 Australia, International DB2 User Group Conference, Sydney, Australia
Technical	March 28 – 30, 2006	Linux World, Sydney, Australia
Executive	April 2 – 5, 2006	Teradata Universe Business Conference, Madrid, Spain
Technical	April 23 – 27, 2006	DAMA International Symposium & Wilshire Meta-Data Conference, Denver, CO
Technical	April 23 – 27, 2006	COLLABORATE 06, Oracle Technology & Application Forum, Nashville, TN
Technical	April 24 – 27, 2006	EMC World & Technology Summit, Boston, MA
Technical	May 7 – 11, 2006	IDUG 2006 North America, International DB2 User Group Conference, Tampa Bay, FL
Technical	May 14 – 19, 2006	TDWI World Conference - Spring 2006 Chicago, IL
Technical	June 20 – 22, 2006	Teradata Universe Technical Conference, Salzburg, Austria
Technical	June 21-23, 2006	Data Warehousing & Business Intelligence Summit 2006, Rome, Italy

DATA MANAGEMENT HUMOR



"I can't explain it—it's just a funny feeling that I'm being Googled."

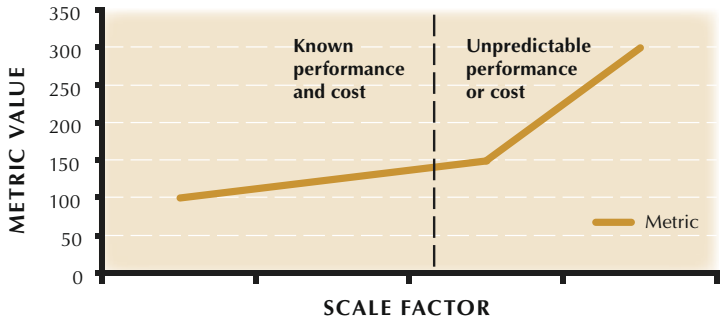
SCALABILITY, from page 4

GET PROOF BEFORE YOU COMMIT

Proof can come in many forms. The best proof is a convincing demonstration of product performance and scalability through a full-scale benchmark using your queries and your data. And while you're testing, don't forget to test the proposed availability solution.

If you can't conduct a benchmark (and many times circumstances preclude one), other types of proof can be obtained. Your own experience with the database products on similar projects is very valuable. Product references (i.e., detailed explanations of how other users solved similar problems at similar scale) are also helpful. The softest, but still quite useful proof consists of detailed architectural explanations by the vendor of how the product handles specific queries or solves specific scalability problems.

Scalability by Metric: *To avoid surprise, require proof of scalability at your requirement levels.*



INVEST THE TIME TO ENSURE A BIG PAY-OFF

In short, an effective product evaluation strategy will:

- Quantify project requirements by assessing all the dimensions of scalability as they relate to your business objectives
- Compare product capabilities to quantified requirements
- Insist on proof

This strategy will avoid mid-project scalability surprises, provide a basis for comparing proposals from different vendors, and allow you to pick the right platform to meet your specific project needs.

Despite the clear benefits of these product evaluation techniques,

SCALABILITY, continues on page 7

INDUSTRY BRIEFS, from page 1

may need help in unscrambling the acronyms and details in order to reap the benefits of Teradata's vision.

Will Appliances for Data Warehousing Take Hold in 2006?

As data volumes continue to grow and data warehousing evolves from an optional enabler for competitive advantage to a mandatory prerequisite for survival, companies are struggling more and more to control the time and resources required to provide advanced business intelligence capabilities. Several new vendors, notably Netezza and Datallegro, are offering an integrated data warehouse appliance as a low-cost alternative to traditional data warehouse platforms. In 2005, Datallegro introduced data warehouse appliance products with costs as low as $90,000 per terabyte. And Datallegro has recently added encryption technology in response to new data security legislation and the increase in data theft. Such appliances claim to manage data warehouse complexity, performance and scale with greater ease-of-use and lower acquisition and administrative costs—but the claims are largely unproven. The value proposition is seductive. We await the validation of these claims.

Consolidation and Acquisitions Will Continue in 2006

The big get bigger—and the trend shows no sign of letting up. Last year IBM moved to an impressively broad strategic vision of information management—and acquisitions played an important part. In 2005, IBM acquired Ascential for data integration and DWL for customer data management—adding to its 2004 acquisitions of Trigo for product data management and Alphablox for analytics. Oracle, the enterprise applications company, added Siebel as its 2005 blockbuster acquisition to PeopleSoft (and JD Edwards), the 2004 blockbuster. Oracle, the database company, added new technologies to its stable in 2005, specifically the in-memory database TimesTen and the open source database engine InnoDB. The challenges of merging technologies and corporate cultures are substantial, with many examples, the most visible being the Digital/Compaq/HP evolution. But IBM and Oracle have handled their acquisitions with care. In today's technology business environment, there is no reason to expect an acquisition slowdown in 2006. Only time will tell if this trend will diminish competition—or heighten it among the data management gorillas who survive. •

SCALABILITY, from page 6

companies often short-change the evaluation process in the face of time and budget constraints. At WinterCorp,

> "If you want to double your business, you don't want to just double the number of processors you have because that may not work."

we have seen organizations invest tens of millions of dollars in a new data warehouse, only to have it fall immediately into disuse. Why? The vendor never had to commit to specific database performance goals because the engineering requirements were never quantified during the evaluation.

We've seen other projects canceled after hundreds of labor years of effort because the selected database, which worked well at proof-of-concept size, could not scale to meet the data volume and workload needs. The lesson is clear: A modest investment in an empirical, engineering-oriented evaluation approach pays off handsomely in lowering project risk and ensuring a scalable solution that meets business-critical requirements. •

Steps to Successful Product Evaluation

1. Identify scalability requirements based on business objectives
2. Quantify those requirements
3. Compare product capabilities to quantified requirements
4. Insist on proof

NEWS FROM WINTERCORP

Facts from the Frontier

How big are today's largest databases? What organizations have implemented them? Which products, platforms and architectures are they using to do the job? And what new trends are shaping the future of large-scale data management? Learn the answers to these and other related questions in WinterCorp's newest research report—your guide to the expanding frontier of database size and power. The report is based on the findings of the 2005 WinterCorp TopTen™ Program, the world's only validated survey of the largest and most heavily used databases in operation today. Since their inception in 1995, the WinterCorp TopTen survey programs have gained widespread recognition as barometers of database scalability and practice. The 2005 TopTen Program discloses the best practices and leading technology solutions of 175 large-scale database practitioners who span the globe and represent all major industries. Order the research report at www.wintercorp.com—and get the facts you need to succeed at, near, or beyond the frontier of database scalability.

Forget to TiVo our TopTen Teleconference?

Good news for those who missed our live teleconference highlighting the research findings and award winners from the 2005 WinterCorp TopTen Program. Now you can download the slides and listen to the 40-minute replay at your convenience simply by logging on to www.wintercorp.com. The teleconference profiles today's top ten databases and their infrastructures by metric (database size, data volume, number of rows, and peak workload) and usage (data warehousing, online transaction processing, and scientific/archive/other). It also provides a glimpse into the future by projecting the growth in database size and workload over the next three years.

WinterCorp Seminar: Insights into Today's Large-Scale Data Warehouse Products

WinterCorp surveys the large-scale data warehouse landscape in a new one-day course designed for your database architects, project managers, and senior technical staff. The course explores how today's large-scale data warehouse products work…how they differ from one another…and how those differences affect database scalability, performance and availability. It covers well-established data warehouse solutions from IBM, Microsoft, Oracle, Sybase IQ and Teradata, as well as newer data warehouse appliances and other emerging products from companies such as Netezza and Datallegro. Available on-site for up to 25 students, it provides the ideal foundation for selecting a new data warehouse platform or giving your team a sound working knowledge of the large-scale data warehouse solutions that might be used in your enterprise today or considered in the future. For more information, log on to www.wintercorp.com.

WinterCorp's Richard Winter Addresses International BI Summit

They're certainly not built in a day, but Rome is the site of an international conference on large-scale data warehouses this summer. Guest speaker Richard Winter will lead two conference seminars—"The Role of Data Warehouse Appliances" and "Building the Large-Scale Data Warehouse"—drawing on practical techniques, case studies, and experience from WinterCorp's client engagements. Sponsored by Technology Transfer, the Data Warehousing and Business Intelligence Summit 2006 is scheduled for June 21-23. For more information, log on to www.technologytransfer.it/en. •

WinterCorp
411 Waverley Oaks Rd, Ste 327
Waltham, MA 02452

FOR MORE INFORMATION:

WinterCorp is an independent consulting firm specializing in the performance and scalability of terabyte-scale data management systems. For more information on our consulting services, please log on to www.wintercorp.com, call **781-642-0300 x132**, or email info2@wintercorp.com.

QUERIES

Expert answers to your large-scale data management questions

Q: ***As head of risk management for a large financial firm, I'm sponsoring a project to expand our data warehouse so we can detect fraud more effectively. The system will be larger than any other BI system we've built. Our architecture team suggested a benchmark to help us choose the right database platform. This would increase our schedule and budget. How can I tell if it's necessary?***

—Name withheld by request

A: IT IS FITTING that you're in risk management, because that is exactly what benchmarks are about. There are three reasons to do a competitive benchmark:

1. To determine if the proposed platforms can meet your requirement set—and what to do if they fall short;
2. To measure differences in their performance, scalability, availability and price/performance; and
3. To capture information that allows you to negotiate the best terms for your chosen platform.

It's essential to benchmark if: (a) your project is high risk; (b) your platform options are very costly; or (c) you're considering new technology that hasn't been widely proven, like data warehouse appliances. A benchmark increases your power as a buyer. It says to vendors, "If you want my business, show me in a measurable way how well you can solve my problem."

Mainstream. If you are doing something that has been done many times before, it's not high-risk. You can learn how it has been done from independent research studies and case studies (such as those produced by WinterCorp) or from users who have implementations that meet requirements similar to yours. You do not need a benchmark.

High Risk. If you are at, near or beyond the frontier of industry experience, you're dealing with considerable risk. You are near the frontier if there are not a substantial number of successful production implementations meeting requirements as tough as your—or tougher. What is a substantial number? Twenty-five or more is a good rule of thumb. In applying this test, consider data volume, workload, query complexity, schema complexity, data latency and workload predictability. If there is little or no comparable industry experience, it's essential to do a benchmark.

High Cost. A very large data warehouse platform can cost tens of millions of dollars. The total investment including application development, data acquisition/cleansing and other services can exceed a hundred million dollars. In most organizations, a platform expenditure of over five million dollars in five years would cost-justify a benchmark. At this level, you're likely to recover the cost of the benchmark during platform acquisition and implementation. That's because a good benchmark will provide the information needed to define what configurations are comparable; negotiate with vendors effectively; identify the most appropriate configuration, and improve—and perhaps accelerate—system design and implementation. Each of these benefits has a financial value. In a large acquisition, they will likely provide a good return on your benchmark investment.

For more details, please contact us. •

SEND US YOUR QUERIES

Scaling Up *welcomes your questions about large-scale data management challenges. To submit a question, email us at ScalingUp@wintercorp.com. Our expert consultants will answer it via email as soon as possible, based on the volume of queries received. In addition,* ***Scaling Up*** *will publish a submitted question and its answer in each issue.*