



scaling up

FALL 2005

IN THIS ISSUE

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

Growing (Business) Pains? Is it your Data Management Architecture | 1

Industry News Briefs:
on IBM... | 1
on Oracle... | 5
on Teradata... | 5
on Microsoft... | 7

Welcome! | 2

Data Management Humor | 4

News from WinterCorp | 5

Database Growth Hits the Fast Track | 6

Queries | 7

Calendar of Upcoming Data Management Events | 8

GROWING (BUSINESS) PAINS? IS IT YOUR DATA MANAGEMENT ARCHITECTURE?

by Megan Santosus

As businesses grow, new requirements can strain data management architecture—dragging down performance, constricting analysis, and impeding crucial operations. Industry leaders take a proactive approach to ensure that their large-scale databases keep pace with changing needs.

IF THERE IS A SINGLE WORD that summarizes the mandate of today's enterprise, it's *agility*. Faced with heightened global competition, businesses need to be more responsive and proactive. To achieve such flexibility, companies typically turn their attention to developing adaptable workforces and processes. What's often overlooked in the mix—yet no less important—are “scalable” IT systems that are designed to adapt, change and grow in step with the needs of the business.

Building such flexible IT systems requires a data management architecture that can scale as business needs dictate. Critical to such an architecture are the data warehouses containing information about customers, products and suppliers that power most businesses. A scalable architecture not only allows companies to

GROWING PAINS, [continues on page 3](#)

INDUSTRY NEWS BRIEFS

IBM tackles enterprise information management

AT A RECENT ANALYST BRIEFING, IBM presented its vision of Information on Demand and described how it has expanded beyond traditional relational data management to the burgeoning domain of content management, including unstructured content, data integration, and information as a service. While DB2 remains the foundation of its Information Management (IM) strategy, the company is taking a broad view of IM. Unlike other providers, IBM claims the agility to swap out various vendor technologies within the layers of its IM framework. Scalability remains critical at all layers, including information integration, master data management, and application development. Each layer must scale well if IBM's vision is to succeed at the ever-expanding frontier of large-scale data management. IBM appears to agree based on the results of its XML development work, which show significant performance gains when accessing relational

INDUSTRY BRIEFS, [continues on page 5](#)



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

Rick Gilbert

Megan Santosus

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

WELCOME TO SCALING UP!

To Our Readers...

WHETHER YOU'RE PART OF THE MANAGEMENT TEAM that runs your company's information technology (IT) department, or help manage an area such as Marketing or Finance that relies on the benefits of IT, you're part of an inescapable and universal phenomenon in which enterprises are striving to digest astronomical amounts of data. And this organizational hunger for data shows no sign of letting up. The world's largest databases are now in the hundreds of terabytes—and growing at an exponential rate.



Richard Winter

The challenge, of course, lies not in curbing our corporate appetite for information—but in managing these growing oceans of data in order to extract their business value, accomplish mission-critical objectives, and achieve strategic goals.

While this challenge is rich in rewards, it is also laced with risk. Experience shows that the larger or more complex the database, the greater the threat of performance and scalability problems. At their best, such problems "merely" constrict an organization's analytic and decision-making capabilities. At their worst, they can bring data

management systems to a standstill, corrupt data warehouses, shut down revenue-driving operations, disrupt supply and distribution networks, or do other damage that can really hurt your business.

Scaling Up is a resource for executive-level owners and sponsors of large-scale data management systems responsible for supporting business-critical objectives and enabling enterprise growth. The goals of the newsletter are to help you address today's data management challenges, make sound technology decisions, and maximize the return on your investment. To meet these goals, each issue will share the experiences of others who have realized the business value of terabyte-scale database management systems...shed light on new technologies and practices for creating scalable data management infrastructures... and explore the relevance of industry advancements in this rapidly changing field. In addition, our *Queries* column provides expert answers to questions posed by readers about their specific data management challenges.

Thank you for reading this premier edition of *Scaling Up*. I'm confident that you'll find it valuable in meeting your data management challenges, and welcome your comments and suggestions for future issues.

Sincerely,

Richard Winter

Executive Editor

richard.winter@wintercorp.com

effectively manage all this information; it also enables them to readily integrate new data sources, support more users, handle bigger workloads, and launch new services—all of which is fundamental to business growth.

“Having a scalable database architecture and sound data management strategy means your business can grow without disruption,” says Richard Winter, president of WinterCorp. In addition, says Winter, such an agile architecture “presents opportunities for a wider range of business strategies and tactics.” In other words, a scalable architecture opens doors that would otherwise be nailed shut by data management limitations.

LET BUSINESS GOALS SHAPE YOUR ARCHITECTURE

For some companies, the approach to scalable data management is dictated by both their operating models and business strategies. Take an online company that sells books, music and movies, for example. The company serves as a conduit between thousands of independent sellers and buyers. The sellers make their inventory information available on the company’s website. All told, detailed data about nearly 50 million titles is posted online. When a buyer purchases one or more items, it’s up to the company to handle logistics and consolidate billing between that buyer and the respective sellers. The data management challenge is to update the inventory quickly and accurately so that a DVD isn’t sold twice or a book doesn’t go unsold because the database fails to present timely and relevant information to the buyer. “The success of the business,” says Rick Burns, vice president of engineering for WinterCorp, “depends on getting lots of detailed information correct and current right up to the last sale.”

For the online company, sub-par data management performance can

lead directly to reduced sales and higher inventory costs. To improve data management and scalability, the first step is to view the challenge in a business context rather than a technical one, according to Burns. When engaged with a client on a project, Burns says, “we begin the requirements phase by defining exactly how the business goal is realized.”



Scalability eliminates obstacles to integrating new data sources, supporting more users, handling bigger workloads, and launching new services—which means your business can grow without disruption.



THE SOLUTION ISN’T JUST ABOUT THE DATA

In the case of the online company, the business goal—to sell books from independent partners—results from a tight coordination of multiple data sources. But it’s not a data issue alone. To build a scalable architecture, companies also have to consider how employees interact with that data. “If the business goal is to rapidly develop targeted marketing campaigns and it takes a month to slice and dice the data, then your system is not up to the demands,” Burns says. “You need to consider how to scale your database architecture so that employees can build an effective campaign in days rather than weeks.”

In terms of scalability, therefore, Burns says it’s important to assess how employee activities relate to the business goal, how employees use

data to achieve the goal, and how that usage affects the database architecture. For example, if the goal is to launch a new promotional campaign in three days, start by envisioning the steps a marketing manager would take to accomplish that objective.

Use that business-focused approach as the foundation for planning and implementing the architecture—rather than simply going out and getting a bigger database system or more processing power. “If you want to double your business, you don’t want to just double your number of processors because that may not work,” Burns says. What companies need to do is evaluate their key applications, and then map the demand those applications place on their IT systems. Only then, Burns says, can companies assess how best to build a scalable architecture.

From Richard Winter’s perspective, a scalable architecture is one that’s modular. Then as business needs change, it’s easier to scale up to support new tasks. While these tasks will vary from company to company, Winter says a scalable architecture will drive their successful completion by supporting greater amounts of more detailed data—along with more complex, insightful analysis.

RETAILERS SCALE UP TO TARGET BIG PROFIT OPPORTUNITIES

Such scalability can yield data management strategies that significantly improve business operations. A great example can be found in the retail industry. Ten years ago, most retailers tracked sales data at a fairly high level, such as by store, product or time period. That big-picture view was honed down to a granular level when it became possible to leverage

GROWING PAINS, [continues on page 4](#)

point-of-sale data. Retailers with the foresight to capture such information in a data warehouse realized they could perform much more detailed analysis. Instead of merely identifying broad market trends, they could track what an individual consumer purchased. The concept of “market basket analysis” was born.

In practical terms, market basket analysis enables retailers to develop precise marketing and merchandising strategies. By understanding who buys what products, for example, stores can target promotions to specific shoppers. And by identifying which products tend to be purchased together, retailers can develop store merchandise displays that encourage cross-selling.

To illustrate the business impact of market basket analysis, Winter cites the case of a retail chain which used its point-of-sale data to discover that many shoppers who bought greeting cards also bought lipstick. The retailer turned this nugget of information into sales without incurring any additional

• • •
A scalable architecture opens doors to new business initiatives that would otherwise be nailed shut by data management limitations.

• • •
advertising or promotional expenses. How? By merely moving the two displays together in its stores, sales of both items increased 30 percent—adding millions of dollars in profit to the bottom-line.

THE POTENTIAL FOR PHARMACEUTICALS

Retailers who have earned a reputation for their targeted marketing prowess have scaled their architectures to slice and dice huge volumes of ever-changing data to increasingly granular levels of detail. It’s a strategy that is now receiving increased attention in the

pharmaceutical industry. According to a recent article in *Pharmaceutical Executive* magazine, the co-authors contend that pharmaceutical companies are missing out on marketing opportunities because they track sales by prescription rather than by patient. Even if patient identity remained anonymous, patient-level data could allow pharmaceutical firms to analyze what drugs are used in combination with others. The results of such analysis could help these companies more effectively drive development, marketing and sales, the co-authors conclude.

AT FEDEX, SCALABILITY DELIVERS CUSTOMER SERVICE

Another industry that has bought into the concept of scalability and data management-enabled business strategies is package delivery, according to Rick Gilbert, executive director of data management consulting at WinterCorp. Unlike the business model focus of the online bookseller, or the marketing potential of the

GROWING PAINS, [continues on page 6](#)



“We have lots of information technology. We just don’t have any information.”

© 2003 The New Yorker Collection from cartoonbank.com. All Rights Reserved.

How to scale your database to support business growth

1. Identify your business goals.
2. Define the activities and data needed to achieve these goals.
3. Quantify the impact of these activities and data on your architecture.
4. Define a solution that can be shown to meet these quantified requirements, leaving a margin of safety to handle requirements which haven’t yet surfaced.

• • •

and non-relational data using XML optimizations built into DB2. IBM's IM framework is designed to be used over the next decade and beyond—and is well worth tracking.

Oracle gets a sense of humor—and record benchmark results

At Oracle Open World in September, the keynotes were focused almost entirely on the recent acquisition of PeopleSoft and on Oracle's Fusion Architecture. But in technical conference sessions, Oracle continued to march ahead with the implementation of grid architecture across its infrastructure product line. Record 3TB TPC-H benchmark results were announced with Oracle Database 10gR2. In addition, some significant advances in scalability—including new parallel algorithms for sorting and aggregation—were announced for 10gR2. Oracle's push into Linux was also evident in its conference sessions as well as in the 2005 WinterCorp TopTen™ Program, an independent survey of the world's largest databases (check out the survey results at www.wintercorp.com). Larry Ellison, Chuck Phillips and friends displayed a heretofore uncommon sense of humor as they let SNL's Dana Carvey lampoon their imperial ways in a pre-conference keynote.

Teradata scores with large-scale announcements

Large-scale data management was spotlighted in two Teradata announcements made at the company's recent *Partners* user group meeting in Orlando, Florida. The company introduced Teradata Active System Management (TASM), a set of software management products designed to help database administrators better understand how their Teradata warehouse is behaving, simplify its management, and achieve optimal resource utilization and performance.

INDUSTRY BRIEFS, [continues on page 7](#)

NEWS FROM WINTERCORP

World's biggest & busiest databases

THE RESULTS ARE IN. Find out what products, platforms and architectures drive today's largest and most heavily used databases. Like a data warehouse that tips the scale at a record-breaking 100 TB.¹ And an OLTP system with the muscle to handle a torrential 1.1 billion SQL statements per hour (yikes, and we thought we were busy). It's all part of the 2005 WinterCorp TopTen™ Program, an independent survey of the world's biggest and hardest working databases. Log onto www.wintercorp.com to view the complete survey results—or to slice the results by usage, operating system, or metric (e.g., database size, uncompressed data volume, number of rows, and peak workload). WinterCorp will provide detailed analyses of the research findings in publications available in early 2006.



System running out of headroom?

All data management architectures have their scaling limits when faced with today's growing data volumes, user communities, and workloads. The trick is to anticipate these limits before your database platform bumps up against them—resulting in heavyweight scalability problems that can knock out system performance, cause staggering downstream costs, and render your business unconscious. A new *Scalability Assessment* service from WinterCorp enables you to proactively manage scalability issues throughout your system's lifecycle. Using a quantitative, requirements-driven approach, WinterCorp evaluates the scalability of your large-scale database platform in light of emerging requirements...defines and quantifies its risk factors...and identifies how to resolve lurking scalability issues before they impact business success.

For more information, email us at info2@wintercorp.com.

Expert help for enterprise-scale data integration

Moving data from multiple sources to a single target—the dreaded extract, transform and load (ETL) process—is among the most complicated, time-consuming and costly steps in a large-scale data management project. New million-dollar ETL power tools offer distinct advantages. But practical experience shows that it takes more than a tool to keep the ETL process on time and on budget and to deliver true business intelligence. So WinterCorp recently launched a new *Data Engineering* service designed to streamline ETL from A to Z, meet time and budget constraints, and improve data quality in order to yield the greatest business value. It features a methodology based on the company's years of experience architecting some of the world's largest database management systems and working with industry-leading ETL tools. Its seasoned *Data Engineering* team is headed by Bob Vecchione (formerly of Xchange, Strategic Technologies & Systems, Epsilon and Thinking Machines), who has led over 50 client engagements dealing with large-scale data management challenges.

For more information, email us at info2@wintercorp.com. •

¹Validated measure of the total volume of user data. Actual disk capacity could exceed 500 TB. Participation in the WinterCorp TopTen Program survey is voluntary.

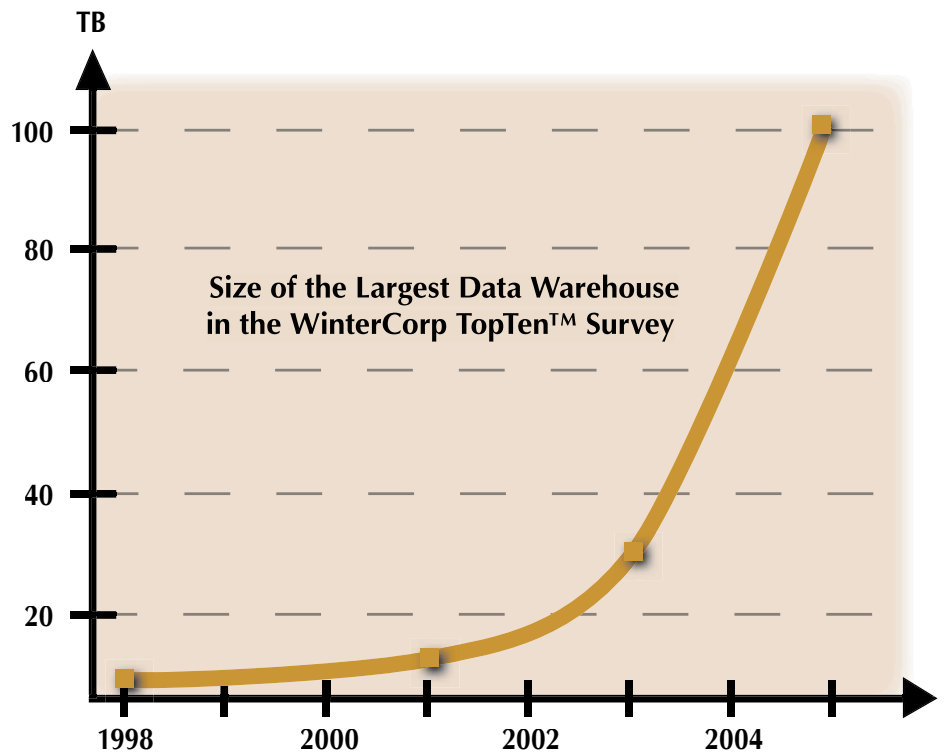
IN OUR NEXT ISSUE ...

How to select the best database platform to meet growing needs

The myriad technology choices available make platform selection a challenging proposition. What's more, the capabilities of different product implementations vary markedly. Some database products serve many concurrent users well. Some serve large databases well. But few serve both well. And all have their scaling limits.

Find out why many product evaluations fail—and how yours can succeed—in our next issue. (Hint: Taking a fact-based, engineering-oriented approach to the product evaluation process can lower project risk, drive sound purchasing decisions, and maximize ROI.)

DATABASE GROWTH HITS THE FAST TRACK



The size of the world's largest databases has tripled every two years since 2001 according to the recent WinterCorp TopTen™ Program, an independent survey of the world's biggest and most heavily used databases.

GROWING PAINS, from page 4

pharmaceuticals industry, shipping companies are motivated by increased competition and customer demands. "Customers want detailed information about their shipments from pick-up to delivery," says Gilbert. "So every customer transaction must be tracked each step along the way." FedEx has responded by giving customers an integrated view into their various transactions across separate shipping options, an approach made possible by a scalable technology platform that accommodates a central data

• • •
"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."

• • •
warehouse. "This strategy preserves the unique operational advantage of each service, but offers customers

a common set of service, sales and technology interactions," Gilbert says. As businesses contemplate their data management strategies, they need to first assess how such strategies will support their business goals. Given that approach, there's no one-size-fits-all plan for scalability. Says Richard Winter: "How you plan and implement your architecture—and how you engineer it for scalability, performance and cost effectiveness—has to be based on your own business needs." •

Scheduled for release as part of Teradata Warehouse 8.1 later this year, TASM consists of enhanced versions of Teradata Manager and Dynamic Workload Manager, along with the new Teradata Workload Analyzer. Our opinion? TASM has the potential to significantly advance the scalability of Teradata warehouses, especially for complex workloads. Teradata also took advantage of its user conference to announce data warehouse support for the Linux platform in early 2006.

Coming up: Microsoft SQL Server 2005

Microsoft's long awaited SQL Server 2005 promises significant advances in database scalability in its commercial release scheduled for November 7. Its key capabilities for large-scale data management include an initial capability for database partitioning and a new integrated extract, transform and load (ETL) facility. In our recent WinterCorp TopTen™ Program, there were three Microsoft sites with SQL Server 2005 already in production. We are awaiting the product's release to see how it actually performs in the field. •

QUERIES

Expert answers to your large-scale data management questions



As a newly appointed marketing executive at a large corporation, I am relying on the delivery within six months of a campaign management system running on our enterprise data warehouse. My staff tells me that we need about 15 terabytes of new stored data incorporated into the system to accomplish our marketing goals. The data warehouse team seems competent, but they haven't done anything on this scale before. How can I tell if they are on track to deliver what I need?

—Name withheld by request



You have a lot at stake here. What's more, such a large and time-sensitive project entails risk, especially if the implementation team is not experienced with the large-scale data warehouse.

You need to ensure that there is a concrete technical approach and plan in place to achieve your goal. This requires some hard thinking and serious analysis. Schedule a thorough and frank review to define the key engineering issues involved in your project. Here are the key questions to investigate with respect to these issues:

1. What are the database requirements? These requirements should be concrete and quantified. They ought to cover database macro structure, scale, workload, response time and operating schedule.¹
2. How do we know that the planned solution (i.e., platform, configuration and database design) will satisfy the requirements? There should be a quantitative case that the requirements will be met. This case should be supported by a performance analysis and preferably backed by test data.
3. What reviews, tests or other steps are planned to confirm that the requirements are being met as implementation proceeds?
4. What actions will be taken if a gap between planned and actual performance is revealed?

This review should be conducted by people who are knowledgeable and experienced with large-scale data warehouse architecture, design and performance. If these resources exist internally, put them to work. Insist that they take decisive actions to address all four questions. If you don't get concrete and convincing answers, seek outside expertise to do the job. Pay attention to how the schedule for questions three and four relate to your timeframe for delivering business benefits.

Based on the outcome of this review, you'll know if your team is on a successful track—or if you need to shift gears by: A) modifying your implementation schedule (e.g., accept delivery in nine months instead of six); B) investing in additional resources (e.g., purchase additional hardware, bring in additional expertise to help your team, etc.); or C) rethinking your marketing program in the short term around a more modest data warehouse solution. •

¹Take a look at *Defining Your Data Warehouse Workload*, which is available at no charge at <http://wintercorp.com/rwintercolumns/archives.html> for more information about these database requirement categories.

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.

CALENDAR OF UPCOMING DATA MANAGEMENT EVENTS

DATE	EVENT
October 30–November 4, 2005	TDWI World Conference, Orlando, FL
November 4, 2005	MIT's International Conference on Information Quality, Cambridge, MA
November 6 – 9, 2005	Business Objects Insight International User Conference, Orlando, FL
November 7, 2005	Microsoft SQL Server 2005 scheduled for general release
November 7–10, 2005	DAMA International Europe Conference, London, UK
November 12, 2005	BI Forum, Chicago, IL
November 13–15, 2005	IBM Information Integration Live, Las Vegas, NV
November 14, 2005	Business Intelligence Forum, Chicago, IL
November 16, 2005	TDWI/DATAlegro Webinar: The Pros and Cons of Data Warehouse Applications
November 17, 2005	Sybase/HP/Intel Seminar: Taming the Data Explosion, New York, NY
November 22, 2005	France Teradata User Group, Paris, FR
November 30, 2005	WinterCorp TopTen™ Program Teleconference



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.