

Change, Challenge and the Data Warehouse

What you need to know about 5 trends that are reshaping the data warehouse landscape

[John Edwards](#), Contributing writer, IBM Data Management magazine

Date: 19 Oct 2009

Level: Introductory

Comments: 0 ([Add comments](#))

[Tag this!](#) [Update My dW interests](#) ([Log in](#) | [What's this?](#))

★★★★★ Average rating (based on 4 votes)

Mike Randolph, vice president and senior technology manager for Bank of America, has seen data warehouse trends come and go. "Change has been constant over the years," says Randolph, who supervises a 22-node, IBM DB2-driven warehouse that supports the bank's credit card operations. "You either learn to adapt to the changes or get swallowed up by them."

These days, five major trends—exploding data growth; end-user demands for greater data analysis, granularity, and speed; requester and source proliferation; the growing popularity of prefabricated appliances/data models; and the challenge of working with unstructured data—are reshaping the data warehouse landscape, challenging adopters across all industries. But Randolph isn't shrinking from the task. "You need to face change head on with the knowledge that any temporary disruptions created will be more than compensated for by better performance and the addition of new capabilities," he says.

For data warehouse managers like Randolph who are willing to embrace emerging trends, change and challenge present an opportunity to excel, says Warren Thornthwaite, a consultant with The Kimball Group, a data warehousing education and advisory organization. "Whether you're dealing with things like growing data volume and the need for deeper data analysis or wondering how to handle unstructured data, you need to turn change into an opportunity," he explains.

1. Exploding data growth

Data is expanding in at least two ways. The amount of information stored inside warehouses is snowballing as content accumulates over time. A 2008 study by a major market research firm revealed that enterprise data requirements are growing at an annual rate of 60 percent.

Meanwhile, as more enterprise processes are instrumented and recorded, warehouse managers face a growing avalanche of data that must be organized and analyzed for possible warehouse use.

Data growth requires enterprises to create data warehouses that can be expanded quickly and efficiently, says Greg Lotko, vice president of warehouse solutions for IBM Information Management. "Look for an offering where modular building blocks allow enterprises to start with a warehouse of a certain size and then, as it grows, click in new modules of hardware and software together," he says.

But data warehouses can't be scaled upward infinitely. To prevent useless data from burdening systems, enterprises must also pay attention to the age and overall quality of their archived data, says George Goodall, an analyst at Info-Tech Research Group.

"Once information gets locked up in a database, organizations are very reticent to get rid of it," Goodall explains, noting that enterprises tend to err on the side of caution. Many opt to keep everything forever, either worried that the information may be needed to fulfill some type of regulatory mandate or simply assuming that at least some of the stuff may have future value. "Enterprises have to start paying attention to the effective life span of data," Goodall says. Information lifecycle management tools that help administrators rate and organize data can make this job easier.

Bank of America's Randolph feels that gaining the upper hand on mounting data is primarily a matter of creating strict-yet manageable—data retention guidelines. "Define retention periods and then stick to them," he says. "If exceptions are requested, make people justify why they need to go around whatever your standard for retention is—then really focus on keeping data only for the period of time that it's needed." Don't assume, for instance, that a compliance mandate requires permanent storage of a certain type of file or record—check the facts to learn what information is really needed and for how long.

Randolph says data modeling is the best way to manage the flow of information into a data warehouse. "You really have to focus on making sure that you're only bringing in data that adds value, as opposed to just saying, 'Hey, here's all this data, let's throw it in the warehouse and we'll figure out what to do with it later,'" he says. "It's simply a matter of thinking out and planning each data source."

[↑ Back to top](#)

2. Picky end users

As data warehouses move deeper into the enterprise mainstream, enduser needs and expectations are driving demand for greater accuracy and more refined conclusions delivered in real time. "In just about anything in life, people always want more than they currently have," Goodall observes.

These increasing demands place new burdens on data warehouses and the people who manage them. Randolph says that carefully designed and configured data analysis tools can help managers satisfy increasingly picky end users without driving costs through the roof or sending performance levels crashing into the basement. "It's a mixture of building tools so that they have good response, and quicker response, but also being smarter on the front end where you're only populating the stuff that's really needed," he notes. Managers can, for example, provide end users with standardized analysis models that will help them achieve their desired goals quickly and easily.

Finding, creating, and fine-tuning data analysis tools to meet end users' growing expectations is becoming a major challenge for data warehouse managers, but so is tempering overly optimistic end-user expectations, says John Hagerty, a data warehouse analyst at AMR Research. "It's very important for IT, in combination with very visible business champions, to in essence paint the picture for people of what's really possible," he says. A few minutes spent with an end user, showing him or her how to effectively use a set of data analysis tools to perform various tasks, is often enough to diffuse complaints that the technology is slow, cumbersome, or ineffective.

Hagerty also suggests that managers regularly assess their tools to see if they are keeping pace with both system capabilities and end-user demands. "It's a continuing process," he adds. "You need to keep evaluating in order to ensure optimum performance." impose the lowest infrastructure burden. "It's really a combination of having a strong gatekeeper, having an underlying infrastructure that adequately supports the data warehouse, and using a strong set of analysis tools," he says.

[↑ Back to top](#)

Table of contents

- 1. Exploding data growth
- 2. Picky end users
- 3. The balancing act
- 4. The out-of-the-box warehouse
- 5. Structuring unstructured data
- Tying it together
- About the author
- Comments

Tags

Search all tags

Popular article tags |

My article tags

More Less

.net technology 9_7
 about_the_product access
 administration and availability
 business cloud **cognos**
 content_management
 customization **data**
 data_access data_integrity
 data_mining data_studio
 data_warehousing database
db2 db2_for_i5/os
db2_for_z/os db2_luw db2luw
 dba eclipse ecm **filenet**
 for_glossary **green** green_it ibm
 ids information_mgmt **informix**
infosphere java_technology
linux lock locking luw
 management **optim** optimization
 p8 performance purequery
 purexml **security** smarter_planet
 sql tools_and_utilities tutorial
 udb unix web_applications
 web_services **windows** xml

[View as cloud | list](#)

3. The balancing act

Many data warehouses are at risk of becoming victims of their own success. As more departments and business partners learn how to exploit the technology to their own benefit, an unprecedented number of new requesters and sources threaten to slow performance to a crawl. For data warehouse managers, the challenge lies in maintaining access and stability in the face of growing system loads-without sacrificing speed and security.

Randolph says that the key to maintaining a successful balance between stability and speed is to use security and access control tools that don't adversely impact system performance. He suggests carefully scrutinizing specifications to find the products and services that impose the lowest infrastructure burden. "It's really a combination of having a strong gatekeeper, having an underlying infrastructure that adequately supports the data warehouse, and using a strong set of analysis tools," he says.

If, despite a manager's best efforts, a data warehouse is beginning to buckle under end-user pressure, it may be time to consider a new approach. "What we're telling our customer base is, spin off a logical datamart inside the data warehouse with Cubing Services," says Bill Wong, program director of data warehousing solutions, strategy, and market offerings at the IBM Toronto Laboratory.

Using IBM Cubing Services, organizations can create, edit, import, export, and deploy cube models over the relational warehouse schema. Cubing Services also provide optimization techniques to improve the performance of online analytical processing (OLAP) queries. "It's helping a lot of companies save on the real estate, the administration of extra servers, power, and things like that," Wong says.

[↑ Back to top](#)

4. The out-of-the-box warehouse

Like bespoke suits and hand-rolled cigars, the custom warehouse is becoming the exception rather than the rule. Today, a growing number of enterprises are turning to warehouse appliances and industryspecific data models that enable a data warehouse to be created in days or hours as opposed to weeks or months.

Goodall says that the "out-of-the-box" approach is highly appealing to organizations that want to build a data warehouse quickly, with less effort, and at a potentially lower cost. "These offerings have abstracted away a lot of the infrastructural complexity that one gets into with building a data warehouse," he explains. "They make a lot of the infrastructure side of things much easier as well; they make it very easy to scale up the scope, the complexity, and the size of the data warehouse."

As Goodall sees it, the signal challenge to prefabricated appliances and data models is that the one-size-fitsall approach should really be labeled "one size fits most." That's because product developers aim for the "average enterprise," not the organization that needs a data warehouse that reflects its exceptional or unique way of doing business. "If you're a leader, and you have gone out of your way to do something different from your competitors, then those industry-standard models can be a bit of a liability," Goodall observes.

On the other hand, despite its inherent limitations, prefabricated technology is certainly a time-saver that will help almost any enterprise get a running start on building its data warehouse. The infrastructure can then be further configured and tweaked to bring it in line with its adopter's specific and custom requirements. and Web pages. "Users get to see and track opinions, attitudes, sentiments, and other concepts that aren't easily represented in traditional data fields," he says.

If, despite a manager's best efforts, a data warehouse is beginning to buckle under end-user pressure, it may be time to consider a new approach. "What we're telling our customer base is, spin off a logical datamart inside the data warehouse with Cubing Services," says Bill Wong, program director of data warehousing solutions, strategy, and market offerings at the IBM Toronto Laboratory.

Using IBM Cubing Services, organizations can create, edit, import, export, and deploy cube models over the relational warehouse schema. Cubing Services also provide optimization techniques to improve the performance of online analytical processing (OLAP) queries. "It's helping a lot of companies save on the real estate, the administration of extra servers, power, and things like that," Wong says.

[↑ Back to top](#)

5. Structuring unstructured data

As data warehouse technology matures and grows more sophisticated, an increasing number of enterprises would like to use their systems to tap into the hidden knowledge that's locked inside unstructured data.

Unstructured data-information that doesn't fit a standard data model-can arrive from many sources, including online surveys, Web forums, and e-mail. "Unstructured data means all the stuff that comes in on the questionnaires or document scans that you can now leverage directly and pair with traditional structured data," says IBM's Lotko. "Then you can derive new insights that you wouldn't have been able to create previously because you didn't have access to the information." Free-form text fields within customer relationship management (CRM) applications, for instance, can give enterprise decision makers the information they need to identify ongoing dissatisfaction trends as well as recurring issues that may be causing the problems.

AMR Research's Hagerty notes that an emerging family of business intelligence (BI) products and services are beginning to give data warehouse end users the ability to peer into and derive meaning from data contained in e-mail, call-center notes, chat transcripts, and Web pages. "Users get to see and track opinions, attitudes, sentiments, and other concepts that aren't easily represented in traditional data fields," he says.

Hagerty sees a bright future for unstructured data. "Once the technology catches up to the promise, unstructured data will become as ubiquitous as traditional BI or analytic technology," he predicts. But embracing unstructured data will require data warehouse managers to undergo a mind change: "One of the things a lot of data warehousing professionals have drilled into them is that things have to sit in rows and columns," he says. "Unstructured data will require these people to look at data in an entirely new light, understanding that text and even media can impart at least as much intelligence as numbers."

[↑ Back to top](#)

Tying it together

Recognizing emerging trends, while important, isn't enough to ensure a data warehouse's long-term viability, says IBM's Wong. He notes that it's equally important to act upon changes as they appear, perhaps by adding new solutions or by adapting established practices to new paradigms. "Warehouses that are not responsive or flexible- they'll die," he says.

Randolph agrees with the need for flexible and responsive systems. "To accomplish this, you've got to stay on top of things, become knowledgeable, and be open to considering new technologies and approaches," he says. "Then, you shouldn't be afraid to make changes, not for the sake of change itself, but always to keep your data warehouse on the leading edge."

About the author

John Edwards (jedwards@gojohnedwards.com) is a technology writer located near Phoenix, Arizona.

Comments

0 comments | [Add comment](#)

[Report inappropriate content](#)

Be the first to add a comment

[Add comment](#)

[↑ Back to top](#)

[Trademarks](#) | [My developerWorks terms and conditions](#)

[About IBM](#) [Privacy](#) [Contact](#) [Terms of use](#) [IBM Feeds](#)

