

Data Warehousing and Business Intelligence for IBM System z
June 2008



Information Management software

Data Warehousing and Business Intelligence for IBM System z

*Mike Biere
IBM Corporation*

Contents	
2	<i>Executive summary</i>
3	<i>Building a dynamic server environment</i>
3	Business considerations
4	Total cost-of-ownership considerations
6	<i>The versatile mainframe platform</i>
7	Delivering new value to the enterprise
9	Ensuring useful, reliable information through sound data management
9	The role of the mainframe in distributed data warehousing
11	<i>Operational business intelligence: Leveraging the mainframe for expanded business insight</i>
14	<i>IBM: Offering an end-to-end data warehousing and business intelligence solution for System z</i>
15	<i>Conclusion</i>
16	<i>References</i>

Executive summary

Companies are faced with massive amounts of data that, when managed strategically, can be transformed into a *wealth* of information for daily, forward-thinking business decisions. Since the late 1980s, organizations have increasingly depended on data warehouses for collecting, storing and staging their corporate data in order to provide a solid, trusted source for decision making. Importantly, the underlying infrastructure of the data warehouse delivers the data needed for today's critical business intelligence strategies, enabling business insight for decision making through reporting, analysis, visualization, statistical data mining and other key functions.

Today, operational business intelligence is emerging as a valuable adjunct to the typical data warehouse infrastructure. The mainframe plays a central role in operational business intelligence scenarios, since most operational data is captured and housed on the mainframe. While the majority of data warehousing efforts are geared toward making management more effective, access to operational data and delivery of business intelligence processes to individuals such as customer service representatives is increasingly important to organizations. This new trend extends the value of data as a corporate asset, while extending the reach and value of business intelligence tools to a new and larger population. The growing focus on operational business intelligence coupled with business demands including power and cooling constraints, consolidation objectives, budget limitations, risk management initiatives and virtualization goals are generating a new interest in the mainframe as a strong data warehousing and business intelligence option.

An important enabler for today's organizations finding tangible value in the mainframe platform is technology companies dedicated to developing business intelligence and data warehousing solutions specifically for the mainframe, development efforts that have been somewhat lacking in the past.

IBM offers not only a flexible, scalable, cost-effective mainframe platform – the IBM System z™ – but also delivers a portfolio of innovative business intelligence and data warehousing solutions for the mainframe. The emerging trends for business intelligence tools to access data on demand and the surge in software and hardware enhancements for IBM System z and IBM z/OS® are enticing many enterprises to re-examine their current data warehouse and business intelligence strategies, and how they can maximize the value of the IBM System z for strategic advantage.

Building a dynamic server environment

Companies want easy, ongoing access to corporate data so they can make accurate and strategic business decisions at any moment of any day. They count on cross-organizational systems and technologies to deliver timely, meaningful information. More and more customers are investing in business intelligence tools as part of their core business strategy, since timely access to data and having a tightly integrated set of data warehouse and business intelligence processes are imperative to maintaining competitive advantage. These tools provide historical, current and predictive views of business operations – such as sales, production and finance through reporting, analysis, visualization, statistical data mining and other business intelligence functions. The data warehouse is the underpinning infrastructure that enables business intelligence, or what the end user sees “on the glass.”

When selecting the best platform for data warehousing and business intelligence, companies must consider the platform’s ability to deliver the best value to the enterprise and its end users. They need useful, enterprise-wide data that is stored securely and readily available to help increase productivity, competitiveness and customer satisfaction.

Business considerations

Various industry trends and business considerations can drive enterprises to examine and sometimes rethink their data warehouse and business intelligence strategies. Some important considerations include:

- *Security issues and regulatory compliance*
- *Costs of maintaining a server farm*
- *Green issues such as power consumption restrictions*
- *Server consolidations for better centralized control*
- *Dynamic data warehousing*
- *Data warehouse building and update cycles*
- *Standardization of business intelligence tools*
- *Operational business intelligence*
- *Real-time information needs*
- *Enterprise portals served from a mainframe hub*
- *New software and hardware offerings for the mainframe*
- *Heterogeneous data access (cross-DBMS and platforms)*
- *New data formats such as XML (Extensible Markup Language)*

Total cost-of-ownership considerations

When choosing a business intelligence and data warehousing platform and assessing associated costs, many organizations focus mainly on hardware and software costs. Many additional costs, however, should be considered. For example, every server in an enterprise must be managed, which involves labor costs. Each server will also require systems administration, monitoring, backup and various associated IT processes. In addition, some companies also face power issues, when increasing their current energy consumption is not an option.

The total cost of ownership for any application platform is based on many factors. Hidden costs can greatly impact the total cost of platform ownership and are often overlooked. Consider, for example, a system's value to an enterprise when costly downtime is a regular occurrence. Security is also a growing economical concern today, with security breaches at times resulting in significant monetary fines. The following industry quotes substantiate some of the concerns enterprises face – including management, security and availability – when considering their platform options:

- ***Management and administration***
 - “The costs of supporting and managing these complex environments and infrastructures have soared, and now far outweigh the customer's expenditure on new systems themselves.” – Software Strategies, November 2005

- **Security breaches**

- There has been “a 43-percent rise in costs compared to 2005 as affected companies scrambled to notify customers, bring in investigators, invest in new security technology and respond to lawsuits.”
- “The total average cost of a data breach grew to \$197 per compromised record, an increase of 8 percent since 2006 and 43 percent compared to 2005.”
- “The average total cost per reporting company was more than \$6.3 million per breach and ranged from \$225,000 to almost \$35 million.” – *Data breach costs soar*, Bill Brenner, Senior News Writer, SearchSecurity.com, 29 November 2007, based on *2007 Annual Study: Cost of a Data Breach*, The Ponemon Institute, Elk Rapids, Michigan

- **Downtime**

- “Cost of downtime can vary by industry and can range from hundreds of thousands to millions of dollars per hour.” – Robert Francis Group, 2005

In assessing the potential costs of employing business intelligence and data warehousing on the various platforms available, companies should align their platform choices with their plans for business growth and expansion. They should ultimately select and build their data warehousing architectures based on how they use information – or more importantly, how they can optimize their corporate data day to day to empower them in their highly competitive marketplace today and in the future.

In a distributed server environment, adding capacity requires adding servers, thus costs go up linearly with expanding workloads. In a mainframe environment, an enterprise will run a “mixed” workload of production, data warehousing and more. The mainframe is optimized for heavy use and companies can add incremental capacity without adding resources to manage the environment. In the end, companies need to select the platform that will deliver the highest return on investment possible, given costs specific to their environment. The mainframe often proves to be the most flexible, cost-effective option available.

The versatile mainframe platform

The mainframe is important to businesses that require reliability, security, performance, low total cost of ownership, resiliency, scalability and more, and is arguably the most successful business-computing platform in history. For decades, mainframes have hosted most of the world's business data and run most of its core transactions, and is used for a variety of purposes such as server consolidation and data centralization, transaction processing, data hosting, Web-enabled applications, functioning as back-end servers, and other purposes. For companies leveraging the mainframe as an integral part of their enterprise data warehousing and business intelligence infrastructure, they will use one of several implementation strategies:

- **Self-contained**—*All of the data and relevant functions reside on the mainframe for data that originates there.*
- **Source of data and additional data provider**—*Mainframe data is treated as an additional source for business intelligence processes as well as for feeding the data warehouse.*
- **Back-end data provider**—*A back-end repository provides new or updated information, but all data warehousing and business intelligence functions are performed on the distributed platform.*

Given the prevalence of the distributed platform in the 1990s and into the past decade, along with the solutions to support it, some companies wonder if the data warehouse can be built and maintained on the mainframe today with the same support and infrastructure as on a distributed platform. Until recently, only some components of the data warehousing software infrastructure were available on the mainframe. Therefore, companies previously opting for a mainframe data warehouse had to customize their own solutions. Today, customer demand and a surging interest in the mainframe for data warehousing and business intelligence have resulted in new, significant end-to-end offerings for running data warehousing and business intelligence on the mainframe platform.

Delivering new value to the enterprise

The benefits of implementing a data warehousing and business intelligence solution on the mainframe will vary by company and the characteristics of a particular enterprise. Yet according to business intelligence and data warehousing experts such as Claudia Imhoff, Ph.D., the mainframe may be the optimal platform for data warehousing and business intelligence solutions for many organizations, while other companies can benefit from leveraging their existing mainframes as a more integral component to the enterprise data warehousing architecture. Imhoff suggests key mainframe benefits to consider, including:

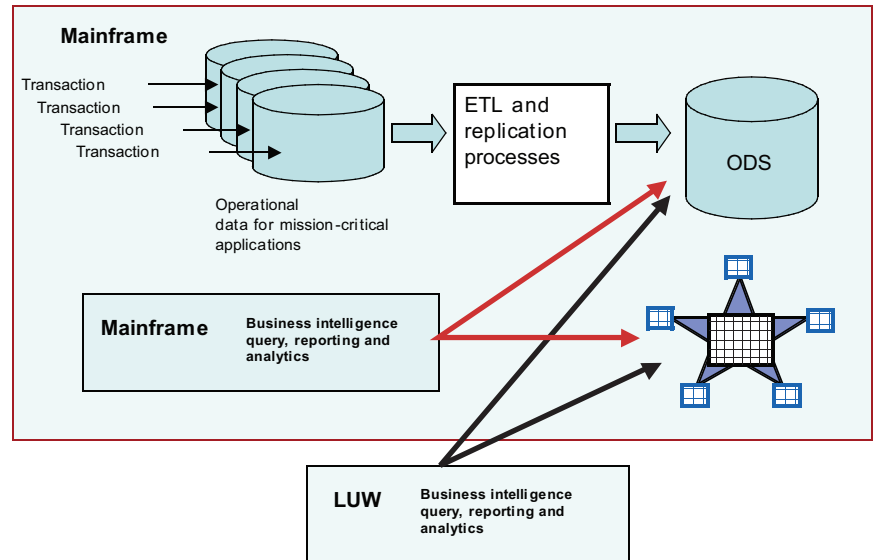
- *Location of captured data*
- *Security*
- *Availability*
- *Scalability*
- *Mixed workload requirements*
- *Near real-time operation*
- *Green initiatives*
- *Server consolidation*
- *Total cost of ownership*

Mainframes are designed to run at maximum capacity to derive benefits and value for every available cycle. Mainframes are optimized for mixed workloads and for automated systems management with facilities such as Workload Manager (WLM), as well as new hardware features, such as IFLs, zIIP and zAAP engines, for offloading some of the processing costs. In a mainframe environment, everyday production systems run alongside the data warehousing and business intelligence workloads. IT stays involved in any data warehousing and business intelligence solution on the mainframe because of the critical nature of the other co-resident functions.

The mainframe offers distinct operational efficiencies as well. For instance, the majority of today’s online transaction processing (OLTP) data is captured on the mainframe, which provides rapid turnaround time from data capture to operational business intelligence use. By placing the business intelligence infrastructure closest to the source, real-time availability of data, and therefore real-time business insight, can happen.

Figure 1 depicts a simplified version of a data warehousing and business intelligence solution on the mainframe. The extract, transform, and load (ETL) and data warehousing infrastructure is housed in its entirety within a mainframe environment, though some may be native IBM z/OS and some on Linux®. The business intelligence tools may be resident on the mainframe or on the distributed platform or both.

Figure 1: A Mainframe-Centric Data Warehouse Architecture



In the mainframe-centric data warehouse environment, the typical end user will either access data via a browser (thin client) or a workstation (rich client), and usually is not concerned with where the data is hosted as long as it is available, accurate and beneficial. Timeliness of updates and new information are critical to a well-disciplined data warehousing and business intelligence environment.

Ensuring useful, reliable information through sound data management

In any data warehousing and business intelligence environment, companies must consider many important data-related issues. Some key considerations include:

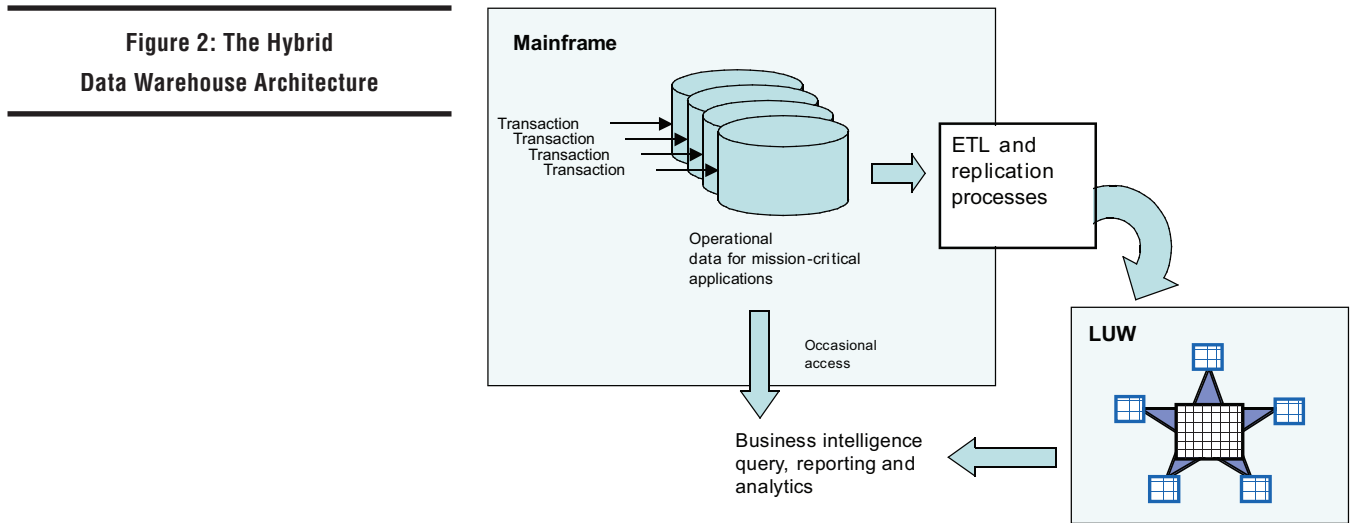
- *Extract, transform, load (ETL) requirements*
- *Intended business intelligence use: query, reporting, data mining, OLAP, operational business intelligence*
- *Periodicity of data capture and replication and how to synchronize it*
- *Monitoring and managing the data warehouse efficiency*
- *Security, access, backup and recovery demands*
- *Scalability over time*

To meet these types of demands, a steady stream of ETL and business intelligence products have emerged within the last 15 years that incorporate the many data warehousing processes as a set of integrated solutions. Quite often, the data needed to feed the data warehouse comes from mission-critical business applications and non-relational sources such as IMS™ or VSAM. In such cases, the data is held in a format that is not suitable for data warehousing and business intelligence, and may be stored in compressed and unusual formats. However, ETL processing is required regardless of the intended data warehousing platform.

The role of the mainframe in distributed data warehousing

Many companies have a mainframe infrastructure where the bulk of operational data is collected within applications, used in building the data warehouses and leveraged to run the business. A large percentage of such installations provide data warehousing and business intelligence solutions on distributed platforms. The mainframe data is either accessed remotely or is replicated and transformed into a form that can be optimized for business intelligence analysis. In this case, the mainframe functions only as a data provider.

In a distributed data warehousing environment, the data must not only have ETL functions applied to it, but the data must also be replicated to the new platform. These processes may be implemented in a variety of ways. One of the common data warehousing architectures, the hybrid, is shown in Figure 2.



Implementing the hybrid solution shown in Figure 2, which features a mixed workload carried by both mainframe and distributed platforms, involves considerations and costs that specifically apply to the distributed platform. When looking at the hybrid as a platform option, a number of factors need to be weighed, including:

- *Additional hardware requirements, such as processors, disks and telecommunications*
- *ETL and replication software*
- *Business intelligence tools*
- *Administration and personnel costs*
- *Coordination and timing of updates and synchronicity*
- *Performance monitoring*

As an enterprise succeeds in a hybrid environment, they typically grow by adding servers and interconnecting them to accommodate increased usage, resulting in a server farm to support their data warehousing and business intelligence architecture. For some companies, distributed and hybrid platforms support their current and emerging business requirements. For others, taking advantage of existing mainframe assets or investing in new ones is the answer, especially with the growing availability today of sophisticated business intelligence and data warehousing solutions for the mainframe.

Operational business intelligence: Leveraging the mainframe for expanded business insight

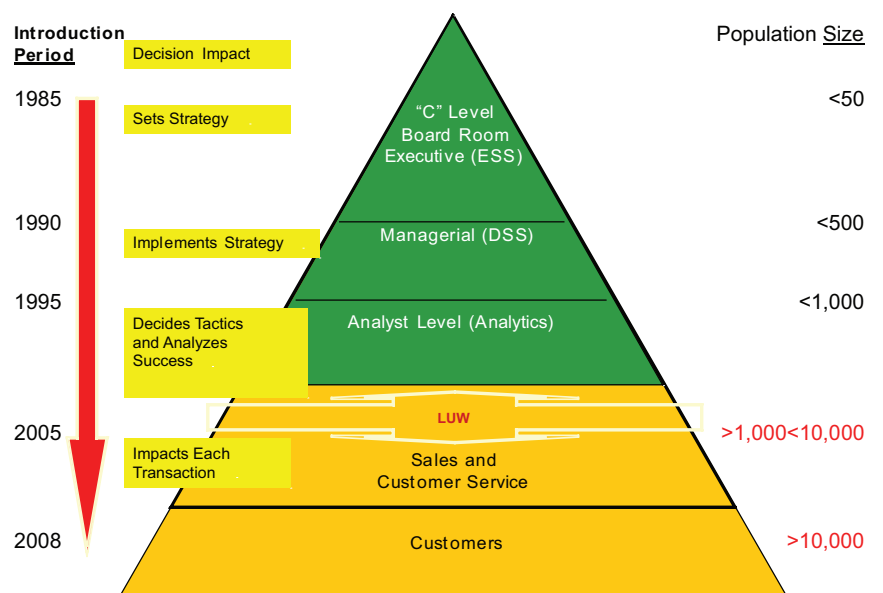
An emerging capability of the mainframe platform is operational business intelligence, also known as operational intelligence. Operational business intelligence complements traditional business intelligence systems by allowing more people to share more information. With operational intelligence strategies, business intelligence results are delivered to non-technical staff as part of a packaged solution. In these scenarios, the data used by the operations staff is more detailed and granular than the data stored in the data warehouse. The data may not need to go through the ETL cycle, but just needs to be available. This is a good example of real-time or near real-time information access.

Since most operational data is captured and housed on the mainframe, more organizations are using this information in its initially captured form or performing some ETL operations on the mainframe to store the data in a format suitable for business intelligence, such as IBM DB2® for z/OS. The operational data may be used to feed a data warehouse, but may also be used directly to support an operational business process such as customer support within an enterprise portal. Operational business intelligence typically involves embedding business intelligence processes such as query, report and dashboard elements within a single application. Operational intelligence data is granular, very current and immediately available since it is not dependent on the data warehousing ETL build/update cycle.

Research suggests that an estimated 80 to 85 percent of enterprise end users are not serviced by their existing data warehousing and business intelligence solutions. Operational business intelligence solutions can alleviate this disparity and may deliver immediate value to an often overlooked user population.

As shown in Figure 3, most current data warehousing and business intelligence solutions target the upper echelons (green) in the enterprise. It may be best to deliver such solutions on distributed platforms as a hybrid, though the majority of the potential users in an enterprise are within the lower (orange) strata.

Figure 3: Operational Intelligence and User Populations



After more than a decade of concentrating on the enterprise’s higher levels, there is now a drive toward operational enablement. Operational intelligence solutions typically involve access to small answer sets from a large database, whereas the more strategic end of the business involves massive amounts of data condensed and aggregated into business groupings.

Strategic business intelligence solutions tend to be more internally focused and provide insights into the past, present and future of the business and its operations. Data at this level is operated on by more technically savvy individuals such as analysts or by top levels in the enterprise looking at the business through dashboards and other executive views. In some ways, these upper tiers have much in common with the lower operational levels: they do not have an in-depth knowledge of the business intelligence tools nor do they have the time or inclination to learn.

Tactical or operational business intelligence is targeted toward those individuals who directly interface with the customer and require immediate data access. Like their executives, they do not possess the skills or time to interact with business intelligence tools. Often, their value is more in their ability to directly interact with the paying customer. The better informed these employees are, the better they can deliver value and increase customer satisfaction. So their direct interaction with operational data on the mainframe is extremely important.

To foster a culture of accountability, where every category of employee can be involved in solving problems, companies can employ operational business intelligence strategies. By doing so, employees are empowered to proactively make strategic, analytical and operational decisions to support corporate and departmental initiatives.

The mainframe plays an enormous role in operational business intelligence scenarios. The multi-purpose role of mainframes, and the new wave of server and storage consolidation, virtualization and use of specialty engines affirms why more and more IT organizations are utilizing the mainframe in increasingly creative ways. With its ability to support and consolidate diverse types of workloads and data through data warehousing, and host new applications including important business intelligence tools, the mainframe is a competitive option for many organizations today. The IBM mainframe, including the powerful new capabilities offered by the IBM System z10™, continues to be a reliable, scalable and secure enterprise platform.

IBM: Offering an end-to-end data warehousing and business intelligence solution for System z

IBM now offers a series of innovative products and extensions to existing products that provide the rich ETL, systems management and business intelligence tools necessary to create, maintain and grow a modern infrastructure for data warehousing and business intelligence on IBM System z. These new and enhanced IBM solutions include:

- **IBM DB2 for z/OS VUE (Value Unit Edition)** – A new value point that enables the deployment of eligible net new application workloads on the System z platform that will provide the same robust DB2 for z/OS data server at a one-time charge price.
ibm.com/software/data/db2/zos/edition-vue.html
- **IBM Information Server for Linux on System z** – A data integration software platform that helps organizations profile, cleanse, extract, transform and load trusted information into a data warehouse to derive more value from the complex, heterogeneous information spread across their systems.
ibm.com/software/data/integration/info_server_system_z/
- **IBM Cognos 8 BI for Linux on System z** – A comprehensive System z offering for enterprise business intelligence enabling all user communities to receive relevant information to drive more informed, faster and more aligned business decisions.
<http://cognos.com/products/cognos8businessintelligence/index.html>
- **IBM DataQuant for z/OS** – A powerful business analytics tool delivering a host of business intelligence capabilities such as dashboards and data visualization.
ibm.com/software/data/db2imstools/db2tools/dataquant/
- **IBM DB2 QMF™ (Query Management Facility)** – Transforms business data into a visual information platform for the entire enterprise with visual data on demand.
ibm.com/software/data/qmf/
- **IBM InfoSphere™ Master Data Management Server for Linux on System z** – A strategic architecture to centrally manage customer, account and product master data for use across the enterprise.
ibm.com/software/data/infosphere/mdm_server/
- **IBM System z10 Enterprise Class Processor** – A world-class server helping businesses create a new enterprise data center for data warehousing and business intelligence.
ibm.com/systems/z/news/announcement/20080226_annnc.html

Conclusion

Today, data warehousing and business intelligence are mission-critical functions for organizations of all types and sizes. Though companies previously viewed business intelligence as an initiative for strategic decision making by top management, they are now rethinking their business intelligence strategies. Many companies are deciding to employ a more holistic platform so they can extend business intelligence capabilities to the entire organization. Executives and operational teams alike can excel at making more informed, faster, more aligned business decisions.

When choosing where and how to implement a data warehousing architecture and business intelligence applications, companies need to carefully weigh their options. Most importantly, companies must consider how they can meet the requirements of handling their current, diverse workloads, along with data warehouse and business intelligence processes. They must focus on aligning the relative cost, complexity and capabilities of an advanced computing environment with their future application needs and vision for the company down the road.

Recent advances in mainframe architectures, along with development of the data warehousing and business intelligence software that support them, are empowering organizations with a reliable, secure, scalable, cost-effective architecture for their new data warehousing and business intelligence functionality. IBM System z supports the entire data warehousing cycle, along with business intelligence software to deliver a total solution infrastructure. The disparity between the mainframe and distributed platform solutions is disappearing, opening the door to a wider range of data warehousing and business intelligence options for today's forward-thinking companies.



© Copyright IBM Corporation 2008

IBM Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
06-08
All Rights Reserved

IBM, the IBM logo, ibm.com, DB2, IMS, InfoSphere, QMF, System z, System z10 and z/OS are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at ibm.com/legal/copytrade.shtml

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Windows is a trademark of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other company, product and service names may be trademarks or service marks of others.

¹*Myths about Mainframe Business Intelligence*
(Intelligent Solutions Inc., 2007)

²*Right-Time Business Intelligence: Optimizing the Business Decision Cycle* (Judith R. Davis, 2005)

³*The Datacenter Mainframe: Cost-efficient, Performance Icon* (Ptak, Noel & Associates LLC, 2007)

⁴*The Mainframe: It's Baaack!*
(Robert Frances Group, 2006)

TAKE BACK CONTROL WITH **Information Management**