Information Management software

IBM

Data Warehousing with the DB2 Family for System z Customers

Leveraging Platform Strengths

December 31, 2008

Authors:
Gary Crupi, Executive IT Specialist, IBM
Jonathan Sloan, Executive IT Specialist, IBM
Jay Allen, Consulting IT Specialist, IBM

Audience:
Chief Information Officers
Chief Technology Officers
Chief Architects
Directors of Enterprise Infrastructure
Business Intelligence and Data Warehouse Solution Architects
Business Analysts
Database Administrators

Executive Summary

As businesses evolve to take advantage of new market opportunities, they are driving changes throughout information technology, and that includes driving investments in data warehousing and business intelligence initiatives. IBM has responded to these new demands with business intelligence solutions based on System z® and with more investment in the development and evolution of System z alternatives that complement existing System x™ and System p® offerings.

The IBM DB2® family of data servers play an important role in past, present and future demands because of their industry-leading performance, security, availability and scalability – and because they can capitalize on the strengths of each unique operating environment. Ranging from Linux®, UNIX® and Microsoft Windows™, to the enterprise class z/OS® platform, DB2 offers wide-ranging solutions that can deliver the capabilities businesses demand, for a manageable, efficient environment. DB2 is the information backbone in an enterprise that addresses critical business requirements for daily operations, business intelligence and data warehousing solutions.

This white paper will discuss the unique support for data warehousing and business intelligence offered by the IBM portfolio of platform solutions. In particular, it will show how an investment in System z, integrated with the DB2 family of solutions, can be used to meet an organization's requirements for data warehousing and business intelligence solutions.

IBM leads the business intelligence and data warehousing marketplace with a family of solutions designed to address the many requirements of enterprise customers. From appliance-like offerings to business intelligence solutions that integrate into the operational IT fabric of the business, DB2 was designed specifically to offer choices for each business's unique requirements.

Managing decision support systems

As businesses try to integrate more of the business insight they gather into operational processes, requirements are forcing a shift away from separate, departmental systems with varying degrees of availability. These modern processes depend on well-integrated components that are optimized for real-time, and response-time-oriented solutions that deliver near continuous availability. In response, IBM has made several major announcements about enhancements to its System z based solutions[1], including the System z10™ Enterprise Class and Business Class Server announcements[2]. These announcements, in addition to presenting new InfoSphere™ data warehousing solutions that will be reviewed later, showcase IBM's ability to deliver a range of solutions with unique capabilities to meet evolving business requirements.

Many businesses already leverage an existing System z infrastructure to meet their business intelligence and data warehousing requirements. But System z businesses running DB2 for

[1] www.ibm.com/common/ssi/rep_ca/1/897/ENUS206-181/ENUS206-181.PDF
[2] www.ibm.com/common/ssi/rep_ca/1/897/ENUS208-041/ENUS208041.PDF

z/OS have a generous range of options to consider when designing their data warehouse solution. Options include System z centric, hybrid or distributed. Some businesses have used the DB2 family to create hybrid solutions that specifically match platform strengths to business requirements.

Over the past decade, in response to customer demand, IBM emphasized the promotion of data warehouse solutions on the distributed platforms. In addition, the introduction of the "data warehouse appliance" has created a new tier of data warehouse vendors who promise lower costs through commoditized hardware. The introduction of the data warehouse appliance came at a time when the architectures required to support initiatives like real-time data delivery and operational business intelligence introduced many new complexities.

As this new breed of business intelligence applications begins to take hold, businesses are realizing that the management of decision support systems is critical. The introduction of the data warehousing appliance helped meet the need for a solution with a faster time to market. That alone, though, failed to address some of the complexities of the workloads running on the appliance. Now that the appliance has been more widely used, businesses are beginning to see the trade-offs between less expensive hardware with higher administration costs and reduced availability. While IBM has recognized this and offers distributed platform solutions, there is an alternative for businesses invested in the System z mainframe platform.

The integration issues of databases have been addressed, but their administration has not. When a business houses the majority of its operational data on the mainframe, and it's considering a business intelligence application, issues related to data latency, integration and lower administration requirements are all easily addressed by integrating the operational business intelligence workload on the System z platform. This is one reason why for certain business intelligence workloads, the System z platform should be considered for data warehousing.

Data warehousing and business intelligence market changes, improvements in technology, changes in environmental factors and the cost of ownership have made System z a highly desirable platform for data warehousing and business intelligence. System z customers have the ability to leverage the entire DB2 family of products. The unique cross-environment options mean that you can take advantage of existing or preferred technology investments. Because IBM platform solutions are designed for the specific platform upon which the database is running, businesses can align the advantages of the platform with the requirements of the business application.

Market Perspective

Changes in the market and the opportunities it offers are driving an evolution in business that in turn is driving changes throughout the information technology marketplace. This includes new investments in data warehousing and business intelligence initiatives. IBM has responded to these new demands with business intelligence solutions based on System z, and more investment in the development and evolution of System z alternatives that complement System x and System p solutions used by small-sized to medium-sized enterprises.

New market factors are driving business interest in incorporating System z into the data warehousing and business intelligence architecture. These include:

Data warehouses are becoming critical to business processes.
The lines between operational and warehouse applications are becoming blurred as operational systems evolve to integrate business insight into infrastructures. The demanding requirements for operational systems are now bleeding over into key business intelligence and data warehousing solutions, altering the way organizations view business intelligence systems availability. This introduces platform selection factors such as availability, latency, transaction scalability and response times into the decision-making process. (Refer to the section, “Making Warehouse Workloads More Operational in Nature” for more information.) IBM coined the term “dynamic warehousing” to reflect the simultaneous ingestion and expression of information for consumers of warehouse information.

Maintaining quality of service without introducing risk.
When data is being moved across multiple tiers, it can introduce many issues and risks to the solution, including delayed access. By deploying a data warehouse on the same platform as the operational systems, you can reduce and often eliminate the data issues that arise. Maintaining proximity to the originating data takes advantage of the scalability and quality of service advantages that the System z platform offers, making DB2 for z/OS the right choice for many applications.

New releases of DB2 family members have focused on data warehousing and business intelligence.
DB2 V8 and DB2 9 for z/OS introduced significant data warehousing and business intelligence functionality that enables businesses to leverage existing investments in System z. The introduction of these new releases is both a response to customer demand and impetus for further growth. Of course, innovative new features to support data warehousing are not limited to DB2 for z/OS. IBM continues to drive innovation across the entire DB2 family.

Customers are demanding lower total cost of ownership (TCO).
IBM has introduced a number of offerings which help lower the TCO for its mainframe solutions. On System z9® and z10 servers, the Integrated Facility for Linux (IFL) specialty engine helps extend the underlying qualities of service of System z to new workloads at an attractive price. In addition, the System z Integrated Information Processor (zIIP), helps

lower the cost of executing or expanding workloads. This group of specialty engines can run independently, or complement each other to create a customized environment to meet business needs. For DB2 workloads, the zIIP redirects DB2 based processing from the System z general processors to help make System z based data warehousing solutions more affordable. This allows a System z customer to reduce costs or increase capacity without impacting associated software costs. IBM recently performed terabyte-scale data warehousing workload tests using data and queries that are recognizable as industry standards. These showed zIIP redirect percentages ranging from 44 percent to 79 percent which indicates a significant cost savings. IFLs also help reduce the cost of hardware and software for the overall data warehousing solution. In February 2008, IBM announced specialty licensing on DB2 for z/OS in support of new warehousing workloads. Called "DB2 Value Unit Edition for z/OS," it is licensed like DB2 on distributed platforms with a one time charge (OTC) rather than a monthly license charge (MLC).

Security exposures continue to plague businesses that are subject to regulatory penalties, financial exposure and class-action lawsuits.
IT costs are not the only costs associated with delivering and maintaining information systems. The System z platform leads the industry with "baked-in" functionality[3], including hardware-assisted encryption, to prevent information exposure. This is a major benefit for financial services organizations, transportation, healthcare, government agencies and in global organizations that conduct business in countries with stringent confidentiality laws. Keeping data on an integrated System z environment, without the exposure of data traveling over networks, simplifies security and reduces risks.

In fact, System z is the only computing system to earn Common Criteria's highest security classification, EAL5. It reached this level through advanced virtualization technology, a hardware environment with end-to-end data protection, a centralized security server that integrates security management and enforcement across all resources, and integrated cryptographic facilities. DB2 for z/OS Version 8 is evaluated under the Common Criteria for Controlled Access Protection Profile (CAPP) and Labeled Security Protection Profile (LSPP) with a conformance claim of EAL3+. z/OS V1.7 with the RACF® optional feature has achieved EAL4+ for CAPP and LSPP. DB2 9 adds several security features that help streamline regulatory compliance and create more effective security policies. These features include network trusted context, database roles, improved auditing, Secure Sockets Layer (SSL) and encryption enhancements[4].

Changes to the cost comparison picture: labor, electricity (power/cooling) and real estate.
Fully 40 percent of the total cost of ownership of applications is due to labor, and labor scales by necessity with the number of distributed servers. Conversely, on System z, labor costs do not increase with incremental workload increases. In recent years, System z data center staffing levels have remained largely the same or decreased, while MIPs consumption has increased significantly.

[3] www.ibm.com/software/tivoli/features/ccr2/ccr2-2007-04/core-strengths.html?s_cmp=rss
[4] www.ibm.com/systems/z/advantages/security/certification.html

In addition, the System z platform consumes less electricity compared to an equivalent amount of distributed server processing. For example, the System z10 Enterprise Class reduces energy usage by greater than 80 percent and saves floor space by greater than 85 percent when used to consolidate x86 servers[5]. Energy consumption and floor space are playing an increasing role in platform choice. Power consumption and availability is a primary concern among data center operators. Many times, they also deal with the proliferation of distributed servers, which can cause floor space and real estate concerns. The consolidation of workloads onto System z can alleviate these issues.

System z support for workload consolidation is driving applications to this platform. Many businesses want to consolidate workloads on a hardware platform that can support virtualization. The current industry trend in virtualization follows a 40-year heritage of virtualization within the System z platform. Virtualization allows different workloads and different environments to coexist on System z, while on distributed platforms a separate platform is at times necessary. (See "A Fresh Look at the Mainframe: Mainframe Total Cost of Ownership Issues" [6].)

Historically Common Solution Approaches to Data Warehousing

For the last two decades, server sprawl was not a big issue. Electricity was cheap, labor was not a factor and organizations focused solely on project-specific return on investment (ROI), and largely ignored all other factors. Big departmental budgets allowed organizations to make project-optimal decisions, but organizationally sub-optimal decisions. During these years, matching the platform to business requirements was not a consideration since the easy answer was to buy new, cheap processing. Many times this was done whether or not there was available capacity, or the existing mainframe system could support the decision process. It had become a "rote" decision.

These seemingly "cheap" warehouse systems evolved into extensive, and very expensive, devices to support relatively simple decisions. Many of these business decisions required nothing more than index look-ups and simple aggregates; however, the upgrades to the latest and greatest models and vendor support drove the cost of these systems skyward. These "commodity" type decisions could easily be handled relatively more cheaply by leveraging existing investment in System z technology. (Of course, some customer requirements are still best handled by platforms other than System z.)

IBM has been able to leverage, and often lead, the innovation in computing for the benefit of its data warehousing customers. Improvements in microprocessor technology by IBM and its hardware partners, in conjunction with pioneering research in parallel processing, have laid the foundation for the IBM InfoSphere Balanced Warehouse based on DB2 for LUW. By tackling the greatest data warehousing problems with a seemingly simple approach of "divide and conquer," IBM has harnessed the power of parallel processing for its customers and enabled them to realize the business value hidden in their enterprise data.

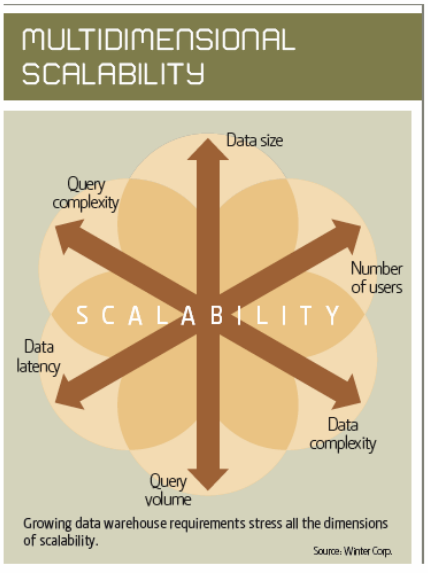
[5] www.ibm.com/systems/resources/systems_z_news_announcement_pdf_ZSO03018.pdf
[6] www.ibm.com/software/htp/tpf/tpfug/tgs07/tgs07e.pdf

Increases in microprocessor speed and the increasing maturity of Linux have helped to drive the cost of these innovations down, making data warehousing technology affordable for even IBM's smallest customers. IBM is the only vendor that gives choices to match platform to business requirements. Customers should understand their unique requirements before deciding which platform is the best to satisfy their business needs.

Scalability

Customers have expressed the need for highly scalable solutions that can meet their increasing volumes of data and user communities. DB2 for Linux, UNIX and Windows has been a leader in data warehousing benchmarks since the mid-1990s. The Transaction Processing Council's TPC-H benchmark measures a solution's ability to support a standard data warehousing workload and can be executed at a number of database sizes. Historically, IBM has executed the benchmark across all scale factors published by the Transaction Processing Council and held the leadership position at all database sizes simultaneously. While IBM no longer executes the benchmark at lower scale factors because they don't represent the size of today's enterprise data warehouse, with DB2 9.5 for LUW, IBM owns the lead at the 10TB scale factor and has done so longer than all other vendors combined.

Of course, benchmarks and query workloads are only a portion of the scalability factor. The Winter Corporation documents several different "dimensions" of scalability[7]. The DB2 family of products is built on platforms that have different characteristics and hence support scalability along different dimensions. DB2 for z/OS supports complexity along a number of the dimensions as does DB2 for LUW. Business requirements drive the type of scalability and therefore the selection of platform.



[8]

[7] www.intelligententerprise.com/showArticle.jhtml;jsessionid=VPCPPZQ00LKOOQSNDLPSKH0CJUNN2JVN?articleID=193105574

[8] Graphic used with permission provided by Richard Winter, Winter Corp.

Matching Requirements to Platform Strengths

The market factors outlined above are driving the incorporation of new and enhanced capabilities for data warehouse and business intelligence systems. These trends include the driving of information-based decision-making down the organization, continual cost pressure, the drive for differentiation from competitors and the need to embed complex analytics within operational systems. As information moves down the organization, the availability of the data to operational processes makes today's data warehousing and business intelligence systems even more critical. It is important for organizations to determine their requirements before selecting a data warehousing and business intelligence solution platform. Matching requirements to the platform that best meets them is one step toward a successful solution. The following sections will analyze some key requirements and how various platforms address them.

IBM offers several data warehousing platforms, including System z, System p and System x solutions. Below is a brief description of each of them. You should consult with your local IBM representative to determine which option is best for your specific requirements.

Data Warehousing with DB2 for z/OS on System z10 – This solution is built on IBM's industry leading reference platform, System z. Customers who leverage the availability, scalability, performance and security of DB2 for z/OS for operational systems can receive the same qualities of service for their data warehousing and business intelligence workloads. Mixing transactional and decision support data on the same platform offers low latency, consolidation, management, TCO and lower risk benefits. Today, the total solution includes DB2 for z/OS, InfoSphere Information Server for Linux on System z, Cognos® 8 BI for Linux on System z and System z servers.

IBM InfoSphere Balanced Warehouse E-Class on System p[9] – This is a complete data warehousing solution comprising pretested, scalable and fully-integrated system components of InfoSphere Warehouse powered by DB2, IBM System p servers and IBM System Storage™. The platform provides robust, market-leading functionality and represents the highest performing and most scalable offering from IBM for Linux, UNIX or Windows deployments.

IBM InfoSphere Balanced Warehouse C-Class and D-Class on System x[10] – These solutions offer similar functionality to the InfoSphere Balanced Warehouse E-Class offerings, but are delivered on the more economical System x infrastructure. IBM InfoSphere C-Class offerings are single server solutions designed to provide a quick return on investment for departmental or small company data warehouses and data marts, and are available on both Windows and Linux. The IBM InfoSphere D-Class offerings provide a high performance, scalable data warehouse platform for businesses that have embraced Linux.

[9] wwwibm.com/software/data/infosphere/balanced-warehouse/e-class.html
[10] www.ibm.com/software/data/infosphere/balanced-warehouse/c-class.html
www.ibm.com/software/data/infosphere/balanced-warehouse/d-class.html

IBM DB2 for Linux on System z with the Data Partitioning Feature

– IBM offers a solution for Linux on System z. With DB2 9.5, the "Data Partitioning Feature (DPF)" works with DB2 for Linux on System z for those customers that have a specific need: To capitalize on consolidation opportunities and use other options for deploying a data warehouse with market-leading business intelligence functionality.

Factors to Consider for Platform Choices

Availability

The downtime cost of an application varies significantly by the application and the business, and these costs should be considered when selecting a database platform. Unplanned and planned outages should be addressed to provide the best availability. Both impact the business and need to be managed based on established service level agreements (SLAs). DB2 addresses both types of outages with a rich portfolio of availability features and functions.

The System z platform has as its foundation, continuous data availability[11] for DB2. The design point for the platform is a "Mean Time Between Failure" (MTBF) of greater than 30 years. DB2 for z/OS tightly integrates with the System z hardware and z/OS operating system to focus on providing five nines (99.999%) availability, masking outages and performing fast, granular data recovery. This is accomplished using industry leading DB2 for z/OS data sharing[12] technology.

System z and DB2 for z/OS data sharing offers continuous availability that is unmatched by any other DBMS vendor. For multiple DB2 applications that could be running on different servers or DB2 subsystems, data sharing on a System z Parallel Sysplex® gives concurrent read-and-write access to the same data. The DB2 subsystems that access the shared data belong to a data sharing group. Each DB2 subsystem in the group is considered a member of the group and share the same DB2 catalog and directory in addition to application data. The Parallel Sysplex leverages specialized hardware and software to manage high-speed caching, locking and list processing needed by the members of the group.

Unplanned and planned outages are masked by a data sharing group. In the rare case that a member of the group fails, other members can continue to process the workload that is routed to it. Data that was locked by the failing member can be quickly released utilizing "Restart Light" to restart the failing member on a surviving server, releasing any retained locks, and shutting back down to avoid over-allocating resources. Of course, the probability of a hardware error is significantly reduced on the System z platform. This is due in part to hardware redundancy and the z/OS operating system "Quality of Service"(QoS) layer. DB2 for z/OS integrates with the QoS layer and naturally benefits from this increased level of reliability and is the standard in reliability across all platforms. Data sharing also allows

[11] www.ibm.com/systems/z/advantages/resiliency/index.html
[12] www.redbooks.ibm.com/abstracts/sg247322.html?Open

new releases and maintenance to be rolled into the environment incrementally across the DB2 members allowing for continuous availability unmatched by any other platform.

All DB2 subsystems, whether data sharing or not, can take advantage of the duplexed control data sets and transaction logs, system parameters that can be configured dynamically, online schema changes, and online utilities. System parameters, better known as DSNZPARMs, are an intricate part of configuring the function and behavior of the DB2 subsystem. The ability to dynamically configure these settings means that you can rapidly change to meet operating and business requirements without an outage. Today, all configuration parameters that make sense to change dynamically have the capability. Online schema changes let you alter the data model when the business needs to without disruption. In the past, changing the length of a data field, for example, would have required an outage. With online schema changes, the change is made without disruption and without the awareness of applications. In addition, DB2 utilities such as backup and reorganization can occur concurrently with read and write access to the data by applications.

DB2 for LUW also addresses availability requirements across the topologies of multiple hardware and operating systems. A highly available solution for data warehousing starts with a highly available database. DB2 9.5 for LUW makes it possible for customers to run nearly continuously: All utilities run online, in conjunction with other applications; most configuration changes can be made to DB2 instances and databases without requiring the database to be restarted; and, schema changes can typically be made to DB2 databases without requiring an interruption in service.

Building on the high-availability characteristics driven deep into DB2, the IBM InfoSphere Balanced Warehouse is architected with high availability in mind. The hardware components that comprise the IBM InfoSphere Balanced Warehouse are configured to eliminate nearly all single points of failure. This includes redundant power and cooling for servers and disk subsystems, battery backups for the disk subsystem’s cache, the use of RAID technology to prevent service interruptions caused by disk failures and dual networks to ensure connectivity across the Balanced Warehouse surviving a failure. Finally, customers can have an IBM InfoSphere Balanced Warehouse configured to provide high availability that can survive the failure of an entire server. Using Tivoli® System Automation to automate failover processing, IBM InfoSphere Balanced Warehouses can be configured with either a failover module (D Class), or to utilize mutual takeover (E Class).

In summary, industry leading high availability is offered across the DB2 family. DB2 for z/OS offers unparalleled capabilities and is considered the gold standard in the industry.

Disaster Recovery

DB2 for z/OS customers with existing disaster recovery processes, procedures and infrastructure can easily provide the same level of service for their data warehouse that their operational systems receive. The options that are generally used include:

- Geographically Dispersed Parallel Sysplex™ (GDPS®) to extend the data sharing group over long distances
- Various disk and tape based solutions – Remote Copy, Peer-to-Peer Remote Copy (synchronous PPRC), z/OS Global Mirror, Hiperswap, Remote tape library
- BACKUP and RESTORE utilities using FlashCopy to provide very fast, non-disruptive backups and very fast recovery for many terabytes
- Image copies, archive log(s) and boot strap data set (BSDS) copies and transport to an off-site location
- Replication techniques
- Disk volume dumps and transport to an off-site location
- DB2 Tracker log shipping

System p and System x customers routinely leverage a number of techniques to provide disaster recovery and continuous availability with the IBM InfoSphere Balanced Warehouse. Each customer's choice is influenced by their own recovery time and recovery point objectives, objectives that often vary by business application. Similar methods as used with DB2 for z/OS can be used in the distributed environment.

- Backup and log shipping
- Backup and log mirroring
- Replication techniques
- Disk-based replication

Mixed Workloads and Workload Management

DB2 and z/OS are uniquely qualified to service mixed workloads as customers increasingly provide business intelligence information in their operational applications. The System z Workload Manager (WLM) manages application workloads to service level agreements, response time and velocity goals. It ensures that long-running applications do not monopolize the server, applications receive consistent service, response time goals are met, and resources are allocated per business goals. It uses a number of WLM services, enclaves, application environments, execution delay monitoring services, along with a combination of response time and velocity goals. With WLM, a particular element of work can be given an initial priority based on business needs. More importantly, over time, the priority of a given element of work can be "period aged" to change the level of service it receives based on changing business priorities detailed in the WLM policy. DB2 for z/OS works with WLM[13] to ensure that the policy is enforced. By managing the system in this manner, it ensures that the system is fully utilized to service a wide range of workloads and business goals. This allows operational and business intelligence applications to co-exist in the same DB2 system.

[13] www.redbooks.ibm.com/abstracts/redp3927.html?Open

The introduction of the DB2 Workload Manager in DB2 9.5 for LUW builds on the stellar reputation of DB2 to support mixed workloads on the same platform. As data warehouses become more operational in nature, both delivering information to and accepting information from operational systems, the need to balance the differing needs of analytical and operational applications is critical to success. Leveraging the lessons learned from the DB2 Governor and DB2 Query Patroller, and the z/OS WLM, the DB2 Workload Manager allows customers to move confidently toward dynamic warehousing.

Real-Time Operational Data Store

The changing nature of today's business requires quick adaptation. To enable flexibility to meet requirements and ensure customer satisfaction, many of IBM's customers have implemented Operational Data Stores (ODS). These real-time or near-real-time ODS environments enable an extension of operational systems, access to operational data via low latency refreshes of transactions, and separation from the operational systems.

ODSs let businesses add new functionality that was not considered during original system design. For example, it has become commonplace to add e-mail addresses or multiple phone numbers within an ODS instead of extending an existing system. With real-time or near real-time access to data, customers, customer service representatives and applications can have query access to the data while insulating the operational systems from online query workload. An ODS can also provide an initial staging area for integrating multi-source data and separating this information source from true operational databases. Lastly, with the operational source system and the operational data store on the same platform, there is less risk associated with data latency and security associated with transporting data across a network.

Incremental Growth

Each member of the DB2 family has an approach to incremental growth. System z allows you to dedicate processing and customize your environment to specify the amount of resources dedicated to a particular workload. This allows businesses with the available capacity to dedicate processing to new workloads and add resources incrementally, as required. With System z, you do not have to purchase an entirely new system in order to dedicate processing or risk over-purchasing for what your needs might be a year from now. This incremental hardware upgrade growth differentiates System z from other platforms. Capacity on demand, when spare capacity is available, lets you add processors or "books" as new CPU capacity is required and may be done while the system continues to run – with no disruption to the existing system. In addition, System Assist Processors (SAPs) can be added in a similar manner to help address issues such as I/O contention.

By contrast, the IBM InfoSphere Balanced Warehouse offers an alternative growth model and is designed to allow customers to deliver incremental, measured growth for their organizations. The modular construction of the Balanced Warehouse means that when a business outgrows its existing implementation, it can simply add capacity where it makes sense. If the size of a customer's database starts to outstrip capacity, additional data modules can be added to provide additional capacity. If a customer's concurrent workload begins to tax the current implementation, more user modules can be added to provide

additional resources. If the addition of new applications stresses the Balanced Warehouse, new application modules can be added to host those applications. This building block approach to growth ensures simplicity and preserves the balance required to prevent resource bottlenecks.

Market Hardened Solution

The DB2 family has been subjected to years of market-based testing that has resulted in a set of proven solutions. DB2 for z/OS recently passed its 25th year in production. It was originally designed for decision support and Online Transaction Processing (OLTP) and has been used for these workloads since its inception. With its 25-year history, DB2 for z/OS is a market-tested and hardened environment that has delivered a favorable cost of ownership to thousands of IBM customers. Furthermore, the IBM InfoSphere Balanced Warehouse builds on this with more than a dozen years of research, development and deployment of parallel database technology.

More than 3,000 companies are already using IBM DB2 for their warehousing needs, on distributed systems and on System z. These include:

- 11 of the top 12 banks
- 7 of the top 8 auto manufacturers
- 5 of the top 6 insurance companies
- 4 of the top 6 general merchandisers
- 4 of the top 5 specialty retailers
- 3 of the top 4 food and drug stores

Low Risk / Prescriptive Implementation

All IBM data warehouse and business intelligence solutions offer a low risk to implementation. Most IBM customers view DB2 for z/OS as a well-known platform that provides continual and ongoing value and return on investment. Businesses place their trust in its reliability, security and manageability and have extensive skills on the platform that are ready to leverage. For these businesses, data warehouse workloads are just a new workload to manage among a varying set of workloads and requirements.

The Balanced Warehouse solution was developed to implement best practices gathered from IBM's data warehouse benchmarking team, IBM Global Services data warehouse practitioners and IBM's most successful data warehousing customers. Rather than requiring customers to assemble the hardware and software components to support a data warehouse workload, the IBM InfoSphere Balanced Warehouse is preconfigured and pre-tested to support a data warehouse workload with predictable performance and exceptional platform stability. IBM has ensured that the Balanced Warehouse offerings include the proper ratio of processors, memory and storage to eliminate bottlenecks for data warehouse workloads.

Desirable Total Costs

There are several ways to measure system cost, including Total Cost of Acquisition (TCA) and Total Cost of Ownership (TCO). Several studies over the years have demonstrated that

System z has a favorable TCO. Typically, an organization that looks at TCO considers software, hardware, labor, electricity for power and cooling, real estate, upgrade, and other software ancillary costs. For many, this is enough to help decide to implement a data warehousing workload on System z. (See "A Fresh Look at the Mainframe; Mainframe Total Cost of Ownership Issues".[14])

Other organizations focus on the cost of acquisition or TCA, which is the cost of hardware and software. When viewing acquisition with a focus on TCA, distributed systems typically appear much less expensive.

The goal should not necessarily be about "which is cheapest," but how the organization measures costs and how costs are allocated over the entire organization. For example, for many years IT costs were allocated out to individual departments based on mainframe computing usage. These costs were often overburdened with IT overhead that distributed system calculations did not have to bear. A more modern and accurate perspective would change the way IT costs are modeled so that resource costs are allocated fairly among mainframe and distributed systems. This way, neither system is unfairly priced internally.

Financial Implications of Trade-In Value

Hardware is typically refreshed about every three years because there is so much change within the hardware industry in three years that upgrades are necessary. This is true either because of significant improvements in CPU and storage technology, or because the risk of failure increases significantly. System z offers protection to the issue of obsolescence. This can significantly impact the evaluation of systems based on Net Present Value costs. When you upgrade to the next generation of mainframe, the specialty processors (including zIIPs) are typically upgraded free of charge. Businesses with increasing workloads typically receive a credit for the existing MIPs investment when they upgrade to a newer generation mainframe. A full trade-in value is applied to the upgrade and increase of MIPS. The long-term implications of this on TCO can be very important. It also indicates that by looking at the life cycle of a warehouse application, you can see that TCA reflects only an initial point-in-time purchase, but not the cost of the application over its life.

Performance Implications of Collocation

The System z ability to tightly integrate with legacy file systems, applications and subsystems such as CICS®, IMS™, and VSAM, in addition to DB2 data, is a key advantage to data warehousing and business intelligence environments. Collocation of data, applications and processes provides the best performance relative to architectures that include a physical network. The benefits of collocation include low latency and a simplified, consolidated environment. Data collocation is achieved by keeping the target data systems, like the ODS or data warehouse, as close as possible to the source data. In addition, placing reporting applications, ETL and replication servers on the server hosting the data can offer similar efficiencies when deploying applications on z/OS or Linux on System z. Mixing all of these disparate workloads on a single server is made possible with WLM.

[14] www.ibm.com/software/htp/tpf/tpfug/tgs07/tgs07e.pdf

Physical networks bridge disparate servers and resources effectively, but they are not intended to improve total solution performance. In fact, the Local vs. Remote Database Access: A Performance Test[15] Redbooks® paper showed in a simple test that network latency was the largest contributing factor to producing elapsed times four to five times greater for remote data compared to local data access. Other, more "real-world like" tests have seen a 50-percent reduction in elapsed times for collocated workloads. Of course, specific customer experiences may vary, but eliminating network access almost always decreases elapsed time. As a result, a standard recommendation is to eliminate physical networks wherever possible throughout the data warehousing and business intelligence solution to enhance performance and reduce complexity.

Implementing reporting tools, ETL servers, replication and other processes on System z will eliminate any physical network. When a portion of the solution utilizes Linux on System z, HiperSockets™ can be configured to enhance performance. HiperSockets are implemented via an internal memory-to-memory transfer, which is the fastest type of data movement. There is still some software network protocol overhead, but a significant portion of the networking delay has been reduced.

In a data sharing group, where operational members reside in the group with the data warehouse or ODS, the two databases can access each other's data without leaving the DB2 subsystem. This facilitates "ELT" (movement of data within the subsystem) and application access to both operational and business intelligence content. This method of data access avoids additional network connections between applications or DB2 systems. In addition, operational applications can retrieve both transactional and data warehouse data. This can be accomplished with a single database connection only when the data warehouse and operational data reside in the same data sharing group.

Making Warehouse Workloads More Operational in Nature

Over the last few years, organizations have recognized that business intelligence information integrated into operational systems can drive greater effectiveness, efficiency and profitability. For example, business intelligence content depicting a more complete customer profile is now included in teller systems at several banks. These operational or tactical applications are becoming a new strategic advantage.

This new operational business intelligence environment requires exactly what System z is good at: reliability, availability and scalability. Placing mission-critical business intelligence solutions on the same platform as the operational systems, with the same quality of service, reduces risk and complexity. Data warehousing and business intelligence have become mission-critical and should be managed on a mission-critical platform, System z.

[15] www.redbooks.ibm.com/redpieces/abstracts/redp4113.html?Open

Customer Examples

Customers are leveraging the platform strengths of IBM solutions to meet their data warehousing and business intelligence requirements today, and customer references are readily available.[16] Following are some examples of the solutions that customers have implemented.

Mainframe Solution – Data Warehousing with DB2 for z/OS on System z

A business process outsourcing company recently decided to leverage its System z environment for its next generation of business intelligence applications. The company had been running its data warehouse on a distributed platform, but sourcing more than 90-percent of its data from the mainframe. The company had problems meeting service level agreements due to the performance of the extract, transform and load processes. The issues were compounded by customer demand for more frequent refreshes of data and near-real-time updates. The company selected System z and DB2 for z/OS as its new data warehouse platform. The proximity to the originating data allowed for high speed data movement which met the data latency requirements of its customers. In addition, the company's existing skills and infrastructure meant that everything was "ready to go." Finally, the solution meets the customer's demanding requirements and delivers a high level of value, with the ability to grow capacity incrementally without adding additional staff. The company is now testing Cognos 8 BI for Linux on System z.

Hybrid Solution – Data Warehousing with DB2 for z/OS and InfoSphere Balanced Warehouse on System p

An insurance company is leveraging multiple IBM platforms to service its business intelligence requirements. Operational data is sourced from DB2 for z/OS transaction processing environments and delivered to an ODS, also utilizing DB2 for z/OS. The ODS is used to store detailed transaction data and allows reporting without adding locking or performance complexity to the operational system. The data is further summarized and loaded into an IBM InfoSphere Balanced Warehouse on System p. Policy management, risk management, agents, education and property survey applications use the data for direct mailing, daily reports and executive dashboards. In summary, the ODS handles fresh, detailed reports while the data warehouse services data marts, analytics and marketing campaigns.

Distributed Solution – InfoSphere Balanced Warehouse on System p

Through its clinical data warehouse, an IBM InfoSphere Balanced Warehouse customer in the health care industry, helps medical and life sciences organizations gather and analyze patient data to determine which treatments work best, which drugs are most effective and

16 www.ibm.com/software/success/cssdb.nsf/solutionareaL2VW?OpenView&Count=30&RestrictToCategory=dmmain_DataWarehouse&cty=en_us

which hospitals save the most lives and why. The customer needed to deliver a pre-tested, pre-optimized solution that could pull data from a variety of locations – desktop databases, data marts, legacy systems, and practice management, and ERP applications. The customer implemented its analytics solution on the IBM InfoSphere Balanced Warehouse, making it simple for the company to install its clinical data warehouse at client sites. With DB2 9 technology, IBM InfoSphere Balanced Warehouse provides the embedded technologies, such as deep compression, that help this customer exceed the performance of existing data warehouse solutions at client institutions. The advanced analytics software enables healthcare providers to easily perform ad hoc queries and complex analytics across terabytes of detailed clinical data. Integrated administrative tools enable the customer to easily manage the offering.

Closing Thoughts

Business intelligence and data warehousing requirements are evolving in the marketplace. IBM is partnering with its customers to leverage their existing infrastructures and exploit market-leading solutions and technology. As seen by the analysis, many requirements can be met through multiple DB2 solutions. System z customers who leverage the DB2 family as intended are well-positioned to derive the most business value from their investment.

About the Authors

Gary Crupi is a Senior Certified Executive Software IT Specialist at IBM working with DB2 for z/OS and Linux, UNIX and Windows (LUW). He joined IBM in 2001 and specializes in helping customers select the right member of the DB2 family to serve their database solution needs. He is an IBM Certified DBA for DB2 9 z/OS, IBM Certified DBA and Advanced DBA for DB2 9 LUW, and IBM Certified Solution Designer for DB2 business intelligence. Gary currently works for the IBM InfoSphere Warehouse Best Practices team.

Jonathan Sloan is a Senior Certified Executive Software IT Specialist, with two decades of business intelligence experience with multi-dimensional and relational database systems. He joined IBM in 2000 and specializes in helping customers understand the application of IBM Information Management technology to business problems. Jonathan has experience in the healthcare, financial services and consumer packaged goods industries. He is currently the business intelligence SME and SWAT team lead for data warehousing on System z.

Jay Allen is a Consulting Software IT Specialist at IBM, where he has been employed for the last ten years with a focus on distributed systems databases. Jay has assisted customers with the implementation and performance tuning of their DB2 database applications and has participated in a number of the large customer benchmarks of DB2 technology during his time at IBM. Jay has 20 years of relational database experience, beginning as a database administrator with DB2 for MVS™ Version 1.1, working with mainframe DB2 for five years and continuing as a systems engineer for Teradata Corporation, Informix Software and now IBM. Jay is currently a member of the IBM Information Management Architecture team.

IBM Corporation
Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
12-08