Using IBM System z as a Hub for Business Analytics

An IBM Redbooks® Point-of-View publication by
IBM Systems and Technology Group

By **Alex Louwe Kooijmans**, Senior Solution Architect, **Eric Michiels**, Executive IT Architect, **Eric Schnoor**, Client Technical Architect, and **Theresa Tai**, Executive IT Specialist

Highlights

Decentralization of data management causes the following issues when implementing business analytics solutions:

- Less reliability of the decision making process
- Increased component costs with a decentralized infrastructure
- Greater risks because of security threats

Addressing the challenge of siloed data when implementing analytics solutions

The phenomenal growth of global data volume in recent years originates from many sources, including automated business processes, web applications, and user-created content, as well as digital data that is generated through the instrumentation of the physical infrastructure. Sources vary from relational tables, to documents, log files, sensor readings, blogs, tweets, digital photos, webcam videos, and voice recordings. Taking advantage of this big data presents new opportunities for today's organizations to accomplish tasks more intelligently and productively.

The problem is that the necessary information is rarely available to the organization precisely when and where it is needed because the data is growing faster than IT processes can extract and process it. Over time, subsets of the data have ended up in *siloed* warehouses and applications, where it is then replicated and duplicated into multiple forms throughout the organization. Traditionally, many organizations have delivered against their strategic objectives for business analytics by implementing this discipline on a department-by-department basis. Each department or business unit works within its own budget, with the mind-set to satisfy its own limited scope of goals. There is little focus on commonality, reuse, and sharing of business and delivery patterns.

This decentralized data infrastructure poses critical challenges to those who provide and consume business analytics information. Implementing business intelligence and analytics solutions within siloed organizational business units has accelerated the sprawl of data marts, data warehouses, and even operational data stores. With the decentralization of the data also comes an explosion of processes to take care of extracting, transforming, moving, and reloading data. In most cases, these processes are coordinated by a strategic architectural design and therefore result in a fragmented and siloed approach to business intelligence.

These isolated departmental projects quickly turn into a resource-consuming vortex of hidden costs that require multiple support and maintenance capabilities and that demand extended administration resources. These investments are typically dedicated to individual projects that lead to limited visibility about "who is accessing what." Most importantly from a business analytics viewpoint, these individual projects often fail to provide a complete picture of the organization and the challenge to be addressed and, as such, can produce contradictory or errant data query results.

The solution needed

Today's organizations have more data on hand for making critical business decisions than ever before. To extract business value from such a huge volume of data, businesses need new systems with high capacity that integrate traditional transaction processing capabilities with deep analytics. These new systems allow businesses to take actions from real-time transactions and to become more responsive to trends as they develop. Transaction processing systems are critical both for traditional back-end reconciliations and for real-time transactions.

In real-time analytics, the following aspects are key:

- Real-time, online, and updated data is used.
- The analytics operation is executed as part of a transaction, not a delayed batch process.
- The business process is designed to apply the results of the embedded real-time analytics functions.

As the need for "real-time" analytics increases, where scoring algorithms are executed as part of real-time transactions, your infrastructure must include both online transaction processing (OLTP) and business analytics capabilities. Therefore, the key to using this data efficiently and effectively is managing the data in an organized, simplified way and in a centralized, streamlined hardware environment.

If your organization is currently implementing a siloed data management infrastructure, a *hub-and-spoke architecture*, such as that offered by IBM® System z®, can provide a more reliable and secure mainframe-based data repository (Figure 1).

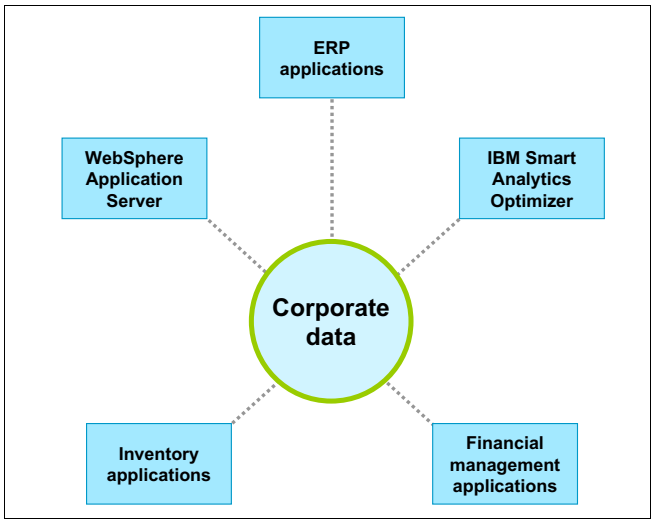

Figure 1 System z hub-and-spoke architecture

This type of architecture can provide a cost-effective solution by using a centralized hub that provides the following advantages:

- Improved reliability of decision making processes
- Reduced cost by reducing physical infrastructure components
- Minimized risks of security threats
- Enhanced information governance

Improved reliability of decision making processes

Collocation of operational data stores, data marts, data warehouses, and business intelligence (BI) tools generally reduces the time that is needed for extract, load, cleanse, transform, and replication processes. In addition, no external network is involved and remote file sharing is not needed. In the case of transferring data between multiple system images on System z, IBM HiperSockets™[1] can be used. Therefore, decision makers have access to the latest information, which can help to make decisions more accurate and reliable.

Reduced cost

One of the ways to reduce cost is to reduce physical infrastructure components, which avoids duplication of data stores, unnecessary file transfers across the network, and hidden costs in the business that result from delayed and inaccurate information. Also, resource use is optimized in a centralized environment because you can dynamically assign more resources to resource-intensive extract, transform, and load (ETL), online analytical processing (OLAP), mining, and report generation processes.

For example, a System z server typically has time intervals with relatively lower utilization rates available during the daylight processing, and ETL processing can consume this white space. Running queries on System z can be cost-effective and efficient when used in combination with the IBM System z Integrated Information Processor (zIIP) and IBM DB2® Analytics Accelerator. zIIP is a specialized processing unit that is fully transparent to applications and is used for specific tasks related to accessing the database. DB2 Analytics Accelerator is a specialized hardware accelerator that executes qualified queries on behalf of DB2 on IBM

[1] In a controlled lab environment, using HiperSockets provided a speed of 10 GB per second, and in a specific client situation, a speed of 4 GB per second was achieved.

z/OS®. It reduces the cost of running these queries substantially.

Minimized risks because of security threats

Having data on fewer physical servers means fewer components, such as network connections, ports, switches, and storage devices, that need to be secured. Also, having too many data stores in different physical locations creates a significant challenge to keep all the physical locations synchronized. Data that is not synchronized and duplicate sets of data that are not identical can cause issues and create a risk in providing accurate reports, which can result in noncompliance.

Enhanced information governance

Today's complex business analytics infrastructure must improve the level of trust that users have in information while also ensuring the consistency of data and establishing safeguards over that information. *Information governance* is about data and information quality, complete lifecycle management, risk management, regulatory compliance, security and privacy, and an enforced process and standards to ensure a single trusted data source. When information is trusted, businesses can optimize outcomes and competitive advantages.

The processes and software components that make up the *information governance ecosystem*, illustrated in Figure 2, are typically mapped to the physical infrastructure, which includes servers, storage, and specialized computing engines.

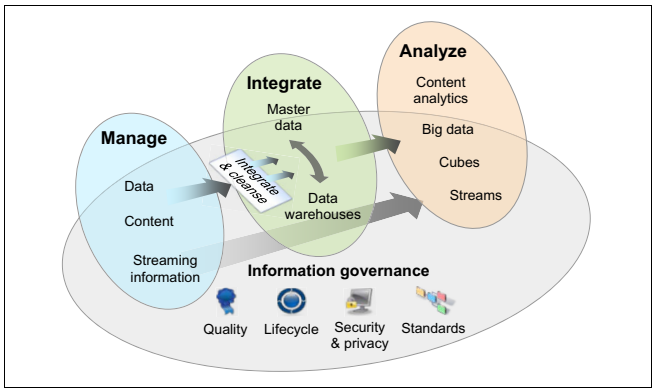

Figure 2 Information governance ecosystem

Often, organizations assume that these processes and software components should be implemented on different platforms and that moving data from one platform to another is the best method for maintaining the ecosystem. But is this really the case? Or should data, in various shapes, be kept in one place as much as possible?

An information governance solution, such as that provided by System z and IBM InfoSphere® Information Governance, can provide a consolidated approach to data. These types of solutions provide information integrity, availability, and quality for a secure and scalable decision support system. For example, IBM InfoSphere Information Governance solutions, which consist of InfoSphere Server, Master Data Management, IBM DataStage®, IBM QualityStage®, IBM Optim™, and Guardium®, provide a foundation for business analytics and optimization. These offerings include information integration, master data management, warehousing, information quality, and governance.

Choosing System z as your data hub

Many large corporations currently depend on System z as their core transaction processing platform. In addition, many of these companies rely on System z to manage their corporate operational data as well, simply because no other computing device is as safe for accommodating corporate data as System z.

System z offers the following advantages as a data hub:

- Security

  Access to data is protected by a fine-grained multi-layered security architecture that cannot be compromised. Companies using System z typically adhere to a strict governance around application deployment and access to their IT systems.

- Integrity

  Data integrity is guaranteed by sophisticated transactional mechanisms when updating databases and automatic recovery mechanisms in case updates cannot be completed successfully.

- Availability

  Data on System z is highly available. Databases and files can be implemented for multiple system images that are dispersed over multiple physical systems in multiple data centers, while keeping one logical copy of the data.

In environments with access to a System z infrastructure, approximately 70% of the data that feeds data warehousing and business analytics solutions originates on the System z platform.

- Performance

  Database systems on System z, in most cases DB2 or IBM Information Management System (IMS™) Database Manager (IMS DB), and file systems, such as VSAM, are designed for high performance.

- Disaster recovery

  Most companies using System z have implemented a bulletproof solution for backup and recovery and have well-tested procedures in place to switch over to a failover site and resume operations after a fallback.

In addition to these benefits, when new applications and databases are added to an existing System z environment, you do not need to reconfigure the system. The new applications and databases benefit automatically from these strengths.

System z offers a fit-for-purpose infrastructure for analytics

Fit-for-purpose means deploying and running applications on a platform that offers the best performance possible at the best price. When using a layered architecture, organizations can realize the best value proposition in analytics using functional components that run on System z in the following ways:

- Accelerate existing and new queries by an order of magnitude without changing the application
- Incorporate scoring algorithms inside DB2 on z/OS to execute real-time analytics in a fast and reliable manner based on integration with IBM SPSS®
- Extract, transform, and load (ETL) processes to benefit from high-speed and secure communication technology when executed in an end-to-end fashion on System z

Query acceleration

IBM DB2 Analytics Accelerator is a high-performance analytics accelerator for System z that is designed to deliver faster complex business analysis transparently to all users. By combining System z quality of service and IBM Netezza hardware-accelerated analytics, DB2 Analytics Accelerator speeds up complex queries and delivers unprecedented response times, up to 1,000 times faster, in a highly secure and available environment. DB2 Analytics Accelerator is a workload-optimized, add-on appliance that enables companies to integrate business insights into operational processes to drive winning strategies.

Your organization can realize immediate value through this easy-to-deploy appliance. With queries running in minutes rather than hours, IBM DB2 Analytics Accelerator enables a “train of thought” analysis, where people are no longer constrained by the speed of the system.

IBM DB2 Analytics Accelerator transparently accelerates the most complex analyses from a wide variety of applications and tools when using dynamic SQL, such as IBM Cognos®, IBM DB2 Query Management Facility™, and SAS servers.

By opening opportunities for new approaches and applications, IBM DB2 Analytics Accelerator can help businesses become information-led organizations. IBM DB2 Analytics Accelerator speeds up delivery of business-changing analytics to decision makers in an environment of industry-leading security and availability, taking the risk out of business analytics.

Real-time analytics

Real-time analytics allow organizations to directly incorporate the latest, relevant data into the decision making process in real time. The real-time scoring capability in DB2 for z/OS, combined with System z quality of service, offers improved accuracy, speed, and performance while also reducing overall cost and complexity. The streams-based models are developed and deployed by using IBM SPSS Modeler Client Version 15.0, which does not require any specific z/OS skills to use.

For example, a stream can be developed for credit card fraud detection that predicts the probability of a credit card transaction being fraudulent. This stream then references historical data and predicts a probability score that can be used to decide whether to accept, reject, or keep a transaction on hold. After development, the stream is published to the scoring adapter for IBM DB2 on z/OS. Then, the stream is invoked by using an SQL call that, in turn, returns the predicted score that you can use to make further business decisions directly as part of the transaction or

later. This SQL call can be issued, for example, from an IBM CICS® or IMS transaction or a Java application running in an IBM WebSphere® Application Server environment.

DB2 for z/OS, combined with SPSS Modeler, provides an easy to use development and deployment environment that allows you to perform immediate analytics at the point of customer impact.

On-platform ETL

IBM zEnterprise® incorporates multiple servers and databases into one physical architecture. Servers are connected to each other through a high-speed, dedicated private network and managed from a single point of control. Thus, ETL processes can be better streamlined and take significantly less time.

By hosting source files and databases on zEnterprise zBX Blades, on Linux on System z, or on z/OS, while all being managed by a zEnterprise management software layer, the information in the data warehouse can be actualized quicker and more frequently. Extraction and transformation of data can take place wherever it makes the most sense from a performance and throughput point of view. That is why organizations combine intelligent ETL with extract, load, and transform (ELT) approaches. In addition, transmission to the data warehouse takes place over a super-fast internal network.

Data is often sensitive and confidential. With zEnterprise, the risk of security breaches is reduced because data travels over the internal private and secure network.

By collocating operational data, data warehouses, data marts, ETL, and ELT software on the same platform, end-to-end processing time to refresh the data used in decision making is significantly reduced, while overall risk is mitigated, and costs can be reduced.

What's next: How IBM can help

IBM has a broad range of expertise to help you throughout the entire process of decision making, design, and implementation for your analytics needs. For example, IBM can organize and perform the following activities for you:

- Assess the effectiveness of your current infrastructure for analytics, now and in the future
- Conduct a "Fit for Purpose" study to determine the best computing platform for the various components in your information management architecture
- Work with you on an approach for transformation to new technology or a new architecture
- Organize solution design workshops, benchmarks, proofs of technologies, and proofs of concepts (POC) around analytics on System z

For more information about using System z for your analytics needs, contact your local IBM representative.

Resources for more information

For more information about the concepts highlighted in the paper, see the following resources:

- Business analytics on System z

  http://www-01.ibm.com/software/os/systemz/badw/
- IBM System z solutions

  http://www-03.ibm.com/systems/z/solutions/
- IBM DB2 Analytics Accelerator

  http://www-01.ibm.com/software/data/db2/zos/analytics-accelerator/
- IBM Systems Magazine article

  http://www.ibmsystemsmag.com/mainframe/Business-Strategy/Green-IT/power_crunch/
- Clabby Analytics on System z

  ftp://submit.boulder.ibm.com/sales/ssi/ecm/en/zsl03159usen/ZSL03159USEN.PDF
- IBM System z Integrated Information Processor

  http://www-03.ibm.com/systems/z/hardware/features/ziip/index.html
- IBM System z Academic Initiative

  http://www-03.ibm.com/systems/z/education/academic/schools_ap.html

- IBM Technical Conferences

  http://www-304.ibm.com/jct03001c/services/learning/ites.wss/zz/en?pageType=page&c=a0002173

- Enterprise Executive

  http://enterprisesystemsmedia.com/magazines/enterprise-executive

- International DB2 User Group

  http://www.idug.org/

- IBM System z Education and Training

  http://www-03.ibm.com/systems/z/education/index.html

- Big Data University

  http://www.bigdatauniversity.com/

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:
INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

This document, REDP-5062-00, was created or updated on October 29, 2013.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (or), indicating US 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 `http://www.ibm.com/legal/copytrade.shtml`

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:
CICS®
Cognos®
DataStage®
DB2®
Guardium®
HiperSockets™
IBM®
IMS™
InfoSphere®
Optim™
QualityStage®
Query Management Facility™
Redbooks®
Redbooks (logo)
SPSS®
System z®
WebSphere®
z/OS®
zEnterprise®

The following terms are trademarks of other companies:

Netezza, and N logo are trademarks or registered trademarks of IBM International Group B.V., an IBM Company.

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

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

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