IBM Software
Thought Leadership White Paper

August 2010

Delivering information you can trust

The benefits of quality data and IBM System z

Executive summary

Would you decide to have an operation or life-saving medical procedure based on impartial or incomplete data? Imagine if you did—and later found out that you were incorrectly diagnosed and the treatment was not necessary. You would have put your life at risk because of bad data.

In the business world, data quality problems may not be a matter of life and death, but they still pose a serious threat. More and more companies and government organizations are learning that poor-quality data is damaging their ability to support business processes, comply with regulations and make accurate decisions.

Even though most business leaders understand the need for high-quality data, they are often not sure how to achieve it. And yet, it has become clear that an investment in the infrastructure must be made to ensure measurably acceptable data quality.

This is where IBM® InfoSphere® Information Server for IBM System z® can make a big difference. InfoSphere Information Server is the foundation of many successful data quality initiatives, helping organizations derive more value from the complex, heterogeneous information spread across their systems. Meanwhile, System z provides a resilient, reliable, high-performance platform for mission-critical data; it's estimated that 95 percent of Fortune 1000 companies store business data on System z.[1]

Data quality can determine business success or failure

Having a clear understanding of customers, partners and suppliers can mean the difference between growing a business and failing to compete. Critical initiatives for information governance, compliance and master data integration simply will not succeed unless the quality of the data in systems is clearly understood and actively managed.

Put another way, bad data is like a virus. A virus is small, yet if left undetected or misdiagnosed, it can spread and become more aggressive, eventually even crippling its host. Imagine that one of your core information systems contained some unchecked or inaccurate data. The data might start out in a part of your strategic master data management (MDM) system used by departmental systems, data warehouses, business intelligence (BI) systems, subsidiaries, trading partners or regulatory reporting staff to make essential business decisions. That bad data will also be used by workers across the information supply chain—and they will consume, process and then spread the "infected" information to others. Along the way, it will skew metrics, reduce report accuracy and ultimately affect business decisions.

The effects of poor data quality include failed business processes, lower productivity and wasted materials. Lost, inaccurate or incomplete information also generates higher costs and extra work, such as hunting down information or additional reconciliation.

IBM System z is an excellent platform for co-locating business-critical applications, processes, transactions, BI systems and data warehouses, and can deliver many performance, security and operational advantages. However, if left unaddressed, poor quality data will undermine these benefits of the System z platform.

Information governance

Most organizations have not yet evolved their processes, policies and infrastructure to be able to help ensure high data quality levels. As a result, organizations are beginning to adopt information governance, a quality-control discipline for adding new rigor to the process of defining common terminology and managing, using, improving and protecting information.

Effective information governance can enhance the quality, availability and integrity of a company's data by fostering cross-organizational collaboration and structured

policy-making. It balances factional silos with organizational interest, directly impacting the four factors that an organization cares about most:

- Increasing revenue
- Lowering costs
- Reducing risks
- Increasing data confidence

Information governance enables an organization to monitor its information supply chain as an end-to-end system, helping to ensure that information is consistently defined and well understood, of high quality, managed throughout its life cycle and protected and secured wherever it lies. With information governance, organizations achieve many goals, from improving decision making to simplifying and strengthening regulatory compliance.

A forum for information governance

Now more than ever, the challenge to protect and manage data has become a universal concern for organizations. To help better understand the emerging space, IBM created a leadership forum in November 2004 for chief data officers and security, risk, compliance and privacy officers concerned with information governance issues.

Since then, the IBM Information Governance Council has steadily grown to comprise nearly 55 leading companies, universities and IBM Business Partners, including large financial institutions, telecommunications organizations, retailers and even government agencies. The Council designed an information governance framework to help businesses understand the supporting and core disciplines and enablers of information governance. It also produced a maturity model to help assess information governance within an organization.

For more information on the IBM Information Governance Council, please visit: ibm.com/software/data/system-z/data-governance/workshops.html

IBM InfoSphere Information Server for System z

The success of an information governance program and supporting data quality initiatives hinges upon a robust data integration technology infrastructure. Enter InfoSphere Information Server for System z, a software platform from IBM that provides breakthrough productivity and performance for understanding, cleansing, transforming and moving information consistently and securely throughout the enterprise.

InfoSphere Information Server for System z helps you leverage information across all of its sources. The software delivers the functions required to integrate, validate, enrich and deliver trusted information for your key business initiatives. It enables you to:

- *Define* a common business language for your information
- *Understand* all sources of information within the business, analyzing its usage, quality and relationships
- *Cleanse and standardize* information to assure its quality and consistency
- *Transform* information to provide enriched and tailored information
- *Federate and deliver* information to make it transparently accessible to people, processes and applications

Those functions are based on a parallel processing infrastructure that provides leverage and automation across the platform (see Figure 1). InfoSphere Information Server for System z also provides connectivity to almost any data or content source, and the ability to deliver information through a variety of mechanisms.

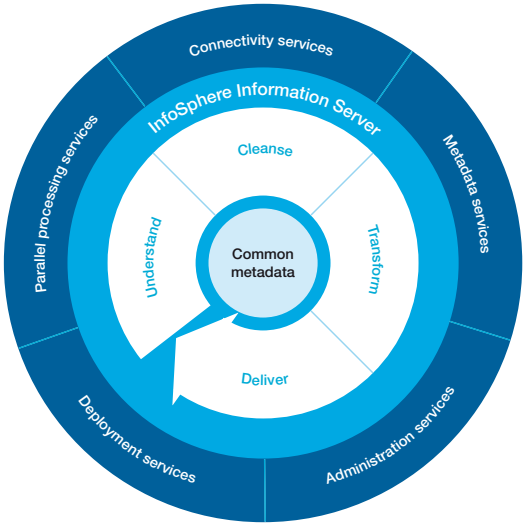

Figure 1: InfoSphere Information Server, built on a foundation of parallel processing and other services, enables critical functions for creating trusted data.

A project life cycle approach

InfoSphere Information Server for System z employs a unified metadata management foundation that enables the seamless sharing of knowledge throughout a project life cycle—and a detailed understanding of what information means, where it comes from and how it relates to information in other systems.

The common metadata infrastructure facilitates a shared understanding across business and technical domains, helping to reduce the time between specification and build. The infrastructure also provides a persistent record of understanding that can dramatically reduce downstream project delivery times and help improve overall insight and confidence in information.

All functions of InfoSphere Information Server for System z share a metadata model that spans design and operational metadata, making it easy for different roles and functions to collaborate seamlessly. The platform provides comprehensive reporting on data movement, data lineage, business meaning and the impact of changes and dependencies across InfoSphere Information Server for System z modules and third-party tools.

InfoSphere Information Server for System z provides:

- Access to a broad range of information sources
- Extensive integration functionality, including federation; extract, transform, load (ETL); in-line transformation; replication and event publishing
- Flexible integration capabilities, including support for Service Oriented Architectures (SOAs), event-driven processing, scheduled batch processing and standard APIs like SQL and Java™

Proven technologies for ensuring data quality

InfoSphere Information Server for System z achieves new levels of information integration, speed and flexibility by providing data quality and transformation capabilities to boost trust in information, automate partitioning and process pipelining for scalability and rapidly deploy services to enhance value. The following components provide additional functionality for InfoSphere Information Server for System z.

IBM InfoSphere Discovery

Before you can implement an information governance program or information-centric project, you must know what data you have, where it is located and how it relates between systems. For most organizations, the data discovery process is manual, requiring months of human involvement to discover business objects, sensitive data, cross-source data relationships and transformation logic. The result is a time-consuming and

error-prone process that slows time to value, establishes doubt about the accuracy of the data within the new system and creates the possibility that the new system will never become operational.

IBM InfoSphere Discovery provides a full range of capabilities to automate the data discovery process. It addresses single-source profiling, cross-source data overlap analysis, matching key discovery, prototyping and testing for data consolidation and automated transformation discovery. InfoSphere Discovery also uses heuristics and sophisticated algorithms that automate analysis to help companies realize 10 times more time and cost savings compared to performing the same tasks manually using a profiling solution.[2]

InfoSphere Discovery capabilities include:

Data profiling: InfoSphere Discovery provides advanced data profiling with results that are "fit for purpose." This includes column analysis, automated primary-foreign key discovery and simultaneous cross-source column overlap analysis of multiple data sources. These sources can be as simple as text files on a PC or as complex as virtual storage access method (VSAM) on System z—or both at the same time.

Unified Schema Builder: Unified Schema Builder takes the output of overlap analysis and uses it as input into a process for helping a data analyst determine the rules by which data will be consolidated for data migration, MDM or a data warehouse, to name a few examples. The Unified Schema Builder component delivers automation software with an embedded workflow to help you complete your consolidation project on time and within budget.

Transformation Analyzer: This component is designed to automate discovery of complex cross-source transformations and business rules by analyzing data values and patterns across two data sources. Transformation Analyzer is used when you know that two data sources are related, but you also know that the relationship can't be described by simple overlaps in data values and requires figuring out how data is transformed between the two sources. Data migration, application retirement, data warehousing and MDM almost always require the mapping and discovery of complex transformation logic between two or more data sources. Transformation Analyzer helps accelerate this process by automating much of the analysis involved and replacing tedious manual work.

The InfoSphere Discovery analysis process establishes an understanding of your data sources and how they relate to each other, generating actionable output that can be immediately consumed by a wide range of information projects, including archiving, test data management, data privacy, data integration, MDM and data consolidation.

IBM InfoSphere Information Analyzer

IBM InfoSphere Information Analyzer, a product module of InfoSphere Information Server, delivers data profiling and rules analysis functions within the context of a complete information integration platform, enabling more accessibility and consistency throughout the enterprise. Active metadata across InfoSphere Information Server simplifies the collection and management of metadata across the entire integration spectrum. Within the InfoSphere Information Analyzer module, profiling results are stored in the common metadata repository. By using the data profiling capabilities of InfoSphere Information Analyzer in the early phases of your data integration projects, you can:

- Expedite delivery of data-driven projects
- Utilize business-driven Rules Analysis with reusable construction and application across multiple data sources to offer quick time to value
- Help minimize costs and resources of critical data integration projects
- Eliminate the risk and impact of proliferating incorrect and inaccurate data
- Help ensure the timely delivery of trusted information

IBM InfoSphere Business Glossary for System z

Difficulties in understanding and interpreting data, determining what data is important and then managing that information creates roadblocks as business and technical users attempt to collaborate for effective information integration. The problem of business definition inconsistency across enterprise environments is often attributed to the absence of an enterprise-wide data dictionary and stewardship program.

InfoSphere Business Glossary for System z helps you create, manage and share an enterprise-wide controlled vocabulary that acts as the common language between business and IT. This is a critical step in better aligning technology with business goals. In addition to a controlled vocabulary, the InfoSphere Business Glossary hierarchy and classification systems provide additional business context.

Actively connected to InfoSphere Information Server metadata services, InfoSphere Business Glossary enables data stewards to link business terms to technical artifacts shared between IBM InfoSphere Data Architect, InfoSphere Information Server or a third-party data integration solution. The result is a common set of semantic tags used by data modelers, data analysts, business analysts, governance stewards, data architects, developers and end users. To help ensure high quality and tight security, only authorized data stewards can use the administrative functions within InfoSphere Business Glossary to create and manage the glossary.

The solution also serves as a history of records to help ensure compliance with regulatory rules, such as the Sarbanes-Oxley Act and Basel II. Business terminology is always subject to change: What defines a "high-value customer" today may be different tomorrow as business requirements evolve. Being able to see the history of what changed, why it changed and who changed it is as important as the change itself. Such a history is critical to data governance protocols, as it increases the trust and understanding of your information.

InfoSphere Business Glossary has a web-based interface designed to enable data stewards to administer the contents of the common glossary. Plus, InfoSphere Business Glossary Packs for different industry verticals help you accelerate the implementation and deployment of your business glossary. Based on knowledge gained during work with more than 400 IBM clients and 10 years of experience in key industries such as banking, financial markets, retail, telecommunications, healthcare and insurance, those packs allow you to quickly deploy, promote and adopt InfoSphere Business Glossary—and therefore get a fast return on your investment.

InfoSphere QualityStage

IBM InfoSphere QualityStage™, part of the InfoSphere Information Server suite, enables enterprises to create and maintain an accurate view of master data entities, such as customers, vendors, locations and products. Core capabilities include data investigation, standardization, address validation, probabilistic matching and survivorship. InfoSphere QualityStage may be deployed in transactional, operational, or analytic applications, and in batch and real-time environments.

InfoSphere QualityStage enables a comprehensive process to manage and maintain data quality. The product's core capabilities focus on:

- **Investigation:** Understand the nature and extent of data anomalies and enable more effective cleansing and matching
- **Standardization:** Create a standardized view of customer, partner or product data. This capability also enables global address cleansing, validation and certification (for significant postal discounts in select localities) and geolocation
- **Probabilistic matching:** Provides an industry-leading matching engine to help ensure the best match results possible; built on a platform enabled for high connectivity and scalability
- **Survivorship:** Helps ensure the optimum consolidation, householding or linked view of record information; enables consolidated and accurate view of customers, partners, products and more

InfoSphere Metadata Workbench

InfoSphere Information Server is designed to be a complete platform for integrating and enriching information across disparate source systems. By leveraging an active and shared metadata repository layer, InfoSphere Information Server can support a full range of integration activities and user roles with collaboration and reuse principles. These artifacts include technical metadata about the various sources of information, business metadata that describes the business meaning and usage of information and operational metadata that describes what happens within the integration process.

IBM InfoSphere Metadata Workbench provides a powerful metadata management interface that supports not only InfoSphere Information Server metadata but also other key metadata that play critical roles in data integration processes. A centralized and holistic view across the entire landscape of data integration processes, with visibility into data transformations that operate inside and outside of InfoSphere Information Server, arms businesses with critical information that can lead to better decisions.

InfoSphere Metadata Workbench highlights include:

- Web-based navigation of key information assets through an interactive and powerful interface provides an easy way for users to access critical information.
- Visual cross-tool and cross-platform data lineage enables an understanding of the complete information lineage, including where data came from and what happened to it as it moved across data integration processes, with extended visibility into enterprise data flows outside of InfoSphere Information Server.
- Visual cross-tool impact analysis allows complete understanding of the impact of a change before the change is made, even when the impact extends beyond a single tool.
- Reporting on information assets, through simple and advanced search with save, repeat and publish capabilities, helps business and IT users to quickly understand complex environments.
- Automated linkages to InfoSphere Information Server metadata services help organizations reduce their overall IT costs and accelerate productivity.
- Collaboration and shared metadata with InfoSphere Business Glossary promotes data stewardship, business and IT alignment and better understanding of information assets.
- Various access levels for different user types provide the flexibility to customize requirement management and enforce information security.

System z: business-critical resilience for information governance

Given the far-reaching effects of an information governance initiative, it's critical that the host platform be utterly reliable and always available. Consider the effect of data not being available for a period of time. A report, "Business resilience: Ensuring continuity in a volatile environment," by the Economist Intelligence Unit in 2007, stated that according to the U.S. National Archives and Records Administration, 25 percent of the companies that experienced an IT outage of two to six days went bankrupt immediately.[3] A disruption in business continuity equates to significant risk with direct costs (loss of immediate business) and indirect costs (long-term damage and credibility to the brand). The damage can extend well beyond the financial realm into key areas, such as customer loyalty, market competitiveness and regulatory compliance.

While some solutions offer weeks or months of uptime, the System z platform offers uptime in terms of years. It provides disaster recovery configurations and is designed to deliver 99.999 percent application availability when implementing IBM Parallel Sysplex® technology to create dynamically balanced clusters with near-linear scalability. This configuration helps avoid the downside of planned downtime, equipment failure or the complete loss of a data center. It also employs some of the most advanced security technologies in the

industry—helping organizations meet rigid regulatory requirements that include encryption solutions, access control management, and extensive auditing features.

As a result, System z is the hub for many organizations' information governance solutions; the world's top banks, insurers and retailers to rely on IBM to help deliver continuity and to secure sensitive business transactions.

Make your information work harder for you

InfoSphere Information Server for System z is a fully integrated software platform that profiles, cleanses, transforms and delivers information from mainframe and distributed data sources alike to drive greater insight for the business without added IBM z/OS® operational costs. It can help you derive more value from the complex, heterogeneous information spread across your systems—and support your information governance initiative.

With breakthrough productivity and performance for cleansing, transforming and moving this information consistently and securely throughout your enterprise, InfoSphere Information Server for System z lets you access and use information in new ways to drive innovation, help increase operational efficiency and help lower risk.

For more information

To learn more about information quality and the IBM System z platform as part of your information governance strategy, please contact your IBM representative or IBM Business Partner, or visit:

- **ibm.com**/software/data/integration/info_server_system_z
- **ibm.com**/software/data/db2imstools/solutions/data-governance.html

IBM Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
July 2010

[1] Moutsos, Kim. "IMS at 40: Stronger than Ever," IBM Database, October 2008. www.dbmag.intelligententerprise.com/story/showArticle.jhtml?articleID=211300235

[2] Time and cost savings based on reports from IBM client engagements.

[3] The Economist Intelligence Unit. "Business resilience: Ensuring continuity in a volatile environment," February 2007. http://graphics.eiu.com/files/ad_pdfs/eiu_Bus_Resilience_wp.pdf

Please Recycle

IMW14326-USEN-00