White paper
March 2009

Information Management software

Information as a service: A smarter way to SOA success

Mark Simmonds
Marketing Manager IM System z
SOA and Data Governance.
IBM Software Group

Contents

Executive summary

This white paper discusses the importance of information as a valuable organizational asset and describes the IBM approach to incorporating service-oriented architecture (SOA) on the IBM System z® platform as a means of preparing and packaging data for continual reuse as an enterprise-wide service.

Introduction

Today's business landscape contains many types of marketplace changes and uncertainty. For many companies and industries, these changes are transformative in nature. Marketplace uncertainty goes beyond current headlines and volatile financial markets. It also includes industry consolidation, aggressive global competitors, complex regulatory requirements and other market changes that require immediate responses from organizations. Business leaders in the global marketplace are focused on creating new competitive advantages and breakthroughs.

Organizations must be able to leverage information in order to foster innovation in their businesses. But business leaders face many challenges when they try to leverage enterprise data as a strategic asset. It is not surprising that as globalization and competition have increased, new problems with data volumes and quality have emerged.

As organizations try to view their customers and business in a unified way, they are quickly discovering that the individual silos of information that have been created or acquired over the years are now a major obstacle in optimizing business processes and relationships.

A recent survey by Accenture[1] of 162 CIOs worldwide showed that:

- **75 percent believed that they can get a competitive advantage by better using and managing their data.**
- **78 percent were actively looking for ways to improve how they used and managed their data.**
- **Only 15 percent thought they were doing that well and comprehensively.**

There is a clear gap between aspiration and execution. Corporations have endless data, but making sense of it and turning it into a trusted strategic asset remains a challenge for many businesses.

Gaining agility

One of the keys to being agile in the marketplace is having the right data readily available. It must be accurate, complete and visible. Making decisions or feeding applications and processes with bad data can expose organizations to severe business risk and financial ruin.

Organizations tend to be swamped with data but remain information-poor because they have many data stores spread across multiple departments, on multiple platforms and in different structures. Due to this siloed approach to data management, many organizations are not aware of the information they possess and are unable to leverage it across the corporation.

Getting smart with an information agenda

Information is a strategic corporate asset with significant value to organizations. Even with successfully implemented information projects across numerous departments, many businesses still have difficulty understanding their risk postures. Projects are implemented without taking into account the strategic view of how data can be used at the enterprise level. IBM has helped countless customers overcome these hurdles by moving them from an application agenda to an information agenda, which places information at the heart of the business.

Adopting a service-oriented architecture

Over the past five years, SOA has seen rapid adoption with significant success. The SOA approach is to view business functions as a set of discrete but linked, interchangeable services that can be reused. SOA equips organizations with a well-defined interface to services, which can be catalogued and managed over the course of their lifecycles. Business units that need to build solutions have access to these common services, which enable them to rapidly build composite applications and new processes. These applications and processes can then be registered as services for others to use.

The challenges of successfully implementing SOA lie in identifying the right information, ensuring that it is of the right quality, and ensuring that it is used consistently. Figure 1 depicts how multiple processes can result in multiple points of access to information, resulting in:

- *Inconsistent views of data across processes (one process gets account data from different places than another process).*
- *Inconsistent application of rules (calculations are performed differently from process to process).*
- *Multiple points of maintenance for the same logic, which is complex, time-consuming and expensive.*

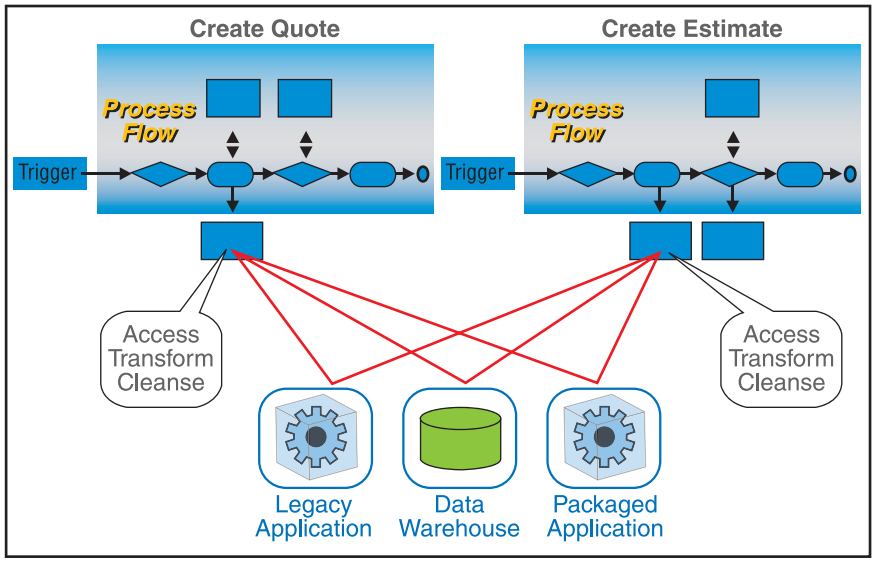

Figure 1: Tight coupling of data to workflow

> *"Effective SOAs include robust data services within an enterprise information structure."*
> — Gartner[2]

SOA centralizes and standardizes the approach to data integration for applications and processes. It packages information as a service to business processes so that consistent, manageable information is made available in a standardized way (see figure 2). It is important to emphasize granularity of services regardless of what kind of services they are. It's about balance. Having data services too coarse may result in less flexibility and reuse. If they are too fine grained, there may be performance impacts, increased complexity and management issues as well as denying the business the return on investment they seek.

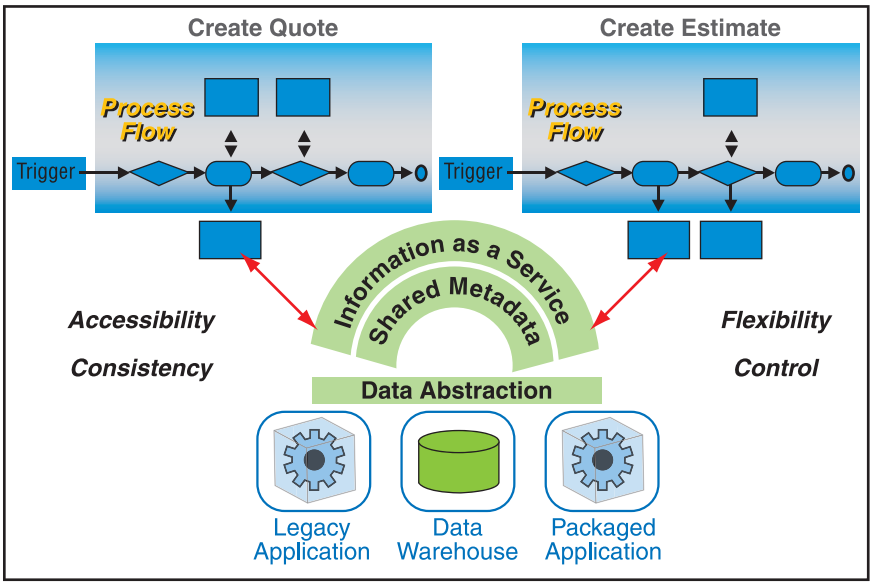

Figure 2: Information virtualization improving flexibility

The SOA approach provides:

- *Consistent definition and packaging of data from process to process.*
- *Consistent rules applied to the data.*
- *Improved data quality.*
- *Centralized control and maintenance.*
- *Metadata relationships to document information sources.*

Figure 3 shows applications and data sources represented as services that can be consumed by other services and orchestrated into processes. This approach to SOA provides a "baggage free" layer of applications and data services that supports dynamic business processes. Advantages include faster time to market and the flexibility to enable a competitive advantage.

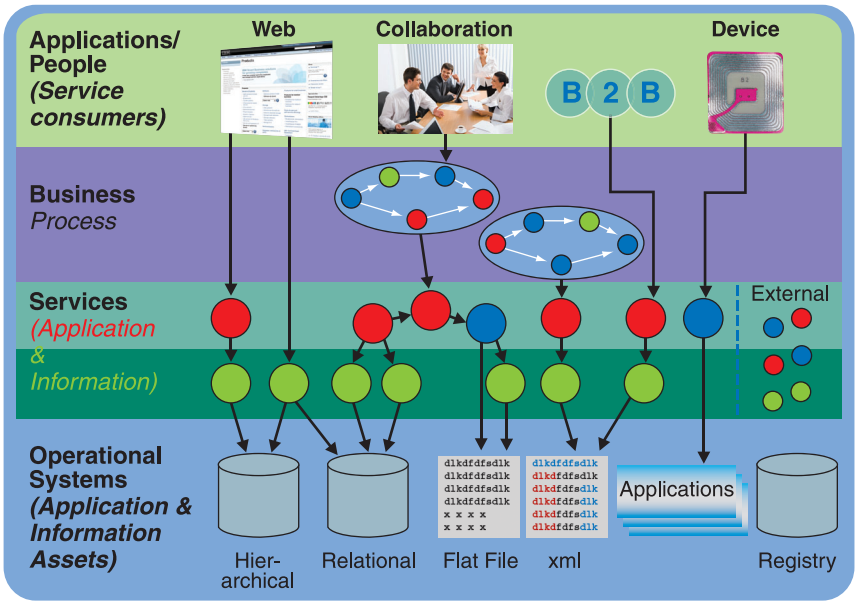

Figure 3: Applications and data sources represented as services

Data is highly distributed, stored, represented and managed in numerous ways across business units. It is important for organizations to determine what data they have and how it is used. They can then catalog key data entities as part of the data architecture and map the physical data model into a logical model.

Such a model enables organizations to discover relationships between data and transform that data before it has an impact on applications. Organizations need to publish their data and the functions performed on them as a set of data services. Doing so makes it possible to package data quality and integrity checks as services that can be invoked in real time.

Leveraging the capabilities of IBM System z as a platform for SOA

IBM System z is at the heart of 95 percent of Fortune 1000 companies,[3] providing unparalleled levels of security, fault tolerance, high availability and transactional integrity. IBM has made significant investments in the architecture and operating system of this business platform to make it ready for SOA.

IBM created the SOA reference architecture, a vendor-neutral way of viewing and planning the set of services that go into building an SOA. The information services and access services components are vital parts of the SOA architecture.

The information services component allows access to and management of data in a unified manner. It provides:

- *A layer of abstraction between processes or applications and the information they use.*
- *Capabilities required to retrieve, combine and modify data from various data sources.*

The information services component provides a uniform way of representing, accessing, maintaining, managing, analyzing and integrating content from heterogeneous sources. It provides a single view of business-critical data for customers, products, employees and partners, delivered in context.

Services (i.e., software components) deployed in the information services component of SOA can be used by services that are deployed in the same component or in others. For example, a Web service that exposes a federation query may be combined with data from heterogeneous sources. This service can be consumed by a business application service or a business process service.

The access services component enables interaction with legacy and pre-packaged applications such as ERP and CRM applications. It provides:

- *A layer of abstraction between processes, portals or applications and legacy or pre-packaged applications.*
- *Capabilities required to access legacy and pre-packaged applications.*

Information sources and the way they are consumed change every day. Therefore, to make business flexible, it makes sense to deliver information as a service and on demand. Virtualized access to information separates information from processes and applications, and makes it easier to change both. Information as a service integrates information to provide a business context to raw information. Services that expose insightful relationships in information serve to help organizations make better decisions.

Preparing organizational data for SOA

To prepare for SOA, organizations must:

- *Discover, understand and catalog key enterprise data entities, including meanings, relationships and lineage.*
- *Map raw, diverse and heterogeneous data into a reusable meta-model.*
- *Cleanse and transform data into trusted content.*
- *Abstract and federate data and functions to simplify access.*
- *Publish data and access to it as a set of information services.*

IBM InfoSphere™ Information Server, which is available on Linux®, UNIX® and Microsoft® Windows® platforms and System z, enables organizations to perform integration functions, including understanding, cleansing, transforming and delivering the information (see figure 4). Underlying these functions is a common metadata and parallel processing infrastructure that enables automation across the platform.

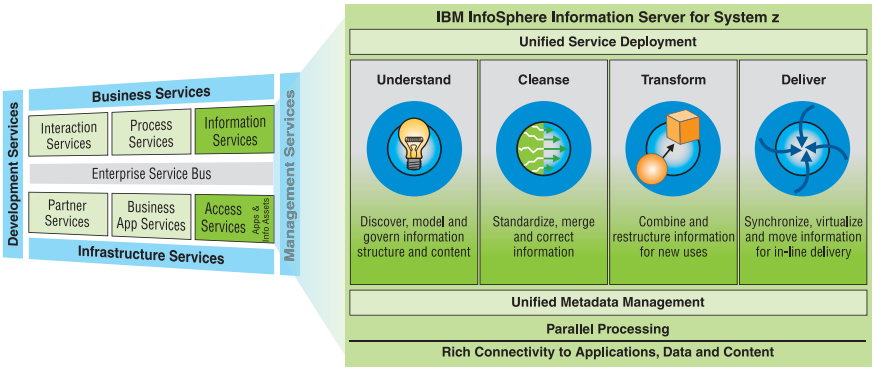

Figure 4: IBM InfoSphere Information Server for System z

The IBM InfoSphere Information Server provides:

- *Access to the broadest range of information sources.*
- *The broadest range of integration functionality, including federation, in-line transformation, replication and event publishing.*
- *The most flexibility in the way these functions are used, including support for SOAs, event-driven processing, scheduled batch processing and standard APIs.*

InfoSphere Information Server integrates information from heterogeneous systems. Its unique, metadata-driven design can help to align business goals and IT activities, providing a consistent understanding of what things mean, capturing business specifications and using them to automate development tasks, and providing deeper insight into data by tracking its lineage.

InfoSphere Information Server has native connectivity to the broadest range of data sources, including databases, applications and files. Its parallel processing foundation enables it to scale to meet the needs of the environment. It improves productivity of development teams working on data integration projects, and improves collaboration across development teams and between developers and business users who set requirements. It also maximizes reuse within and across project teams.

Understanding data assets

Understanding data requires consistent business and technical term definition, data assessment and profiling, and data modeling that is aligned with service and process models. In many cases, business and technical communities use different terms and have different understandings of data meaning.

Data users are often uncertain about data quality. Users need to know whether the organization has consistent keys across systems, how the data is connected, and the data types that are defined. If there is a conceptual data model that applies to various applications, it is most often isolated from the service model. Services are often specified in isolation, and the input and output of the services are specified implicitly in message models that are not consistent with the conceptual data model. Consequently, transformation may be required to map data in the format of the message model to the format of the underlying systems.

In order to address the problem of inconsistent business and technical terms, it is necessary to establish a business glossary that defines terms and shares them across business and IT. InfoSphere Business Glossary provides a Web-based tool for authoring, managing and sharing business metadata. It was designed for business users and subject-matter experts to define data stewards and record business terminology definitions and taxonomies. Its "technical view," created for IT use, is focused on the structure and location of information, while the "business view," created for business operations users, is focused on the usage and characteristics of information, and the rules that govern it.

Once a common understanding of key terms is established, business users must also gain a better understanding of data to determine if it satisfies integrity constraints. The InfoSphere Information Analyzer component can help organizations accomplish this goal. They must perform data assessment or profiling to discover inconsistencies, inaccuracies or anomalies in the data before using the services. Doing so enables organizations to ascertain which cleansing or transformation rules are required to produce and provide trusted information.

Architecting the data model

In the modeling process, the architect defines the future state of the data. InfoSphere Data Architect, formerly IBM Rational® Data Architect, helps define and build the logical data model to enforce common rules and nomenclature across the enterprise. InfoSphere Data Architect not only provides strong logical and physical data modeling capabilities, but it also provides facilities to map across models and enable automatic discovery of relationships.

Cleansing data for better use

Data cleansing improves the quality of data within and across databases but can also be leveraged as a service by applications and activities in business processes. Based on the results of data assessments, organizations may find that they do not have data standards in place or that they have data inconsistencies and redundancies. In such cases, these organizations must execute strategies to resolve their data quality issues.

To cleanse data, IT specifies how to identify duplicates, resolve conflicts, and enrich the information. Those rules are executed against data that is extracted from various sources and then applied to consolidated warehouses or master data management systems.

In the SOA context, the same cleansing rules can be applied to input from a service request. The standardized data is generated as the output of a service response. The same cleansing rules used for database records can now be used across various other data sets and applications.

Transforming data

Transforming data involves creating a standard set of services for data access and creation to be shared across all applications within the enterprise. These services reduce the burden on individual development projects by shielding developers from the complexities of data access and harmonization. They also enable changes to be made on the data side, without requiring that each program be changed individually. Furthermore, they standardize the way data is accessed or created, resulting in more accurate, complete and consistent data.

Data often resides in disparate data stores. Sometimes complete information is not available due to a lack of integration. Sometimes the integration is implemented manually within the application, which causes slow response times when processing large data volumes and when performing complex transformations to merge the data into a consistent format.

SOA from IBM serves to expose selected sets of trusted information as a service. In many cases, a warehouse or integrated data store is the only place an organization has established trusted information. If so, this data can be leveraged in the SOA context where needed. However, organizations may want to consider consolidating data if they have services that require information from disparate systems.

Many companies prepare the analysis of account information as part of a business process. Data is not populated in a warehouse at arbitrary times, but only as needed. The InfoSphere DataStage® component of IBM InfoSphere Information Server can facilitate this information preparation by invoking the population as part of the business process.

Taking advantage of data federation services

Data federation services also address customer challenges by helping to create a layer of standardized data access and representation. This layer allows application developers to focus more on application logic and reuse well-defined data entities rather than having to stitch pieces of data together and worry about data semantics. With federation services, an enterprise gains a 360-degree view of their data assets, enabling them to provide their clients with up-to-the-minute information.

InfoSphere Classic Federation Server for z/OS® is the component of InfoSphere Information Server that provides data federation. It works by:

- *Abstracting access to data stored in various databases, including IBM Information Management System (IMS), CA-IDMS, CA-Datacom, Software AG Adabas, and IBM DB2® UDB for IBM z/OS databases.*
- *Logically integrating structured and unstructured data so that data appears in one relational database accessible through SQL statements and stored procedures calls.*

Data federation services are designed to avoid unnecessarily copying information when integrating it. In addition to the need to have real-time access to current information from various sources, there is also a demand for an architecture that enables stability of its interface when more sources are being added over time as required by mergers and acquisitions.

Data federation servers provide data virtualization, which makes it possible for an application to access distributed data through the data federation server. This approach enables organizations to implement real-time access to distributed data through services, improving time to market and performance.

Transforming and federating on demand

Large-volume data migrations and transformations can be initiated through events or process flows. The same transformation, quality and enrichment rules can be applied consistently across both real-time and batch data. Complex data-level transformations can be handled simply through the InfoSphere DataStage component of IBM InfoSphere Information Server.

Transformations may involve processing multiple rows of data or multiple sources of data. They may also necessitate the introduction of quality rules or metadata. IBM provides integration of operational data stores with historical marts and enterprise data warehouses. It also provides consistency across batch and real-time environments.

IBM adds the following unique capabilities:

- *Pre-packaged complex transformations*
- *Data quality rules*
- *Lifecycle metadata management*
- *Automated source system analysis and mapping (profiling)*
- *Large volume transformation and load capabilities*

Packaging data and information as services

InfoSphere Information Services Director (see figure 5) packages DB2 queries, stored procedures, or information integration logic as services using a graphical eclipse-based tool. The tool provides multiple bindings for these packages, including EJB, JMS, or Web services. InfoSphere Information Services Director can deploy services to be managed through the application server—IBM WebSphere® Application Server—which comes embedded with the product. In addition, InfoSphere Information Services Director provides load balancing and fault tolerance for service requests.

It also provides an easy-to-use deployment and management framework and visual SQL design through an SQL builder that is an integral part of the tool. The service metadata is stored and managed in InfoSphere's metadata server.

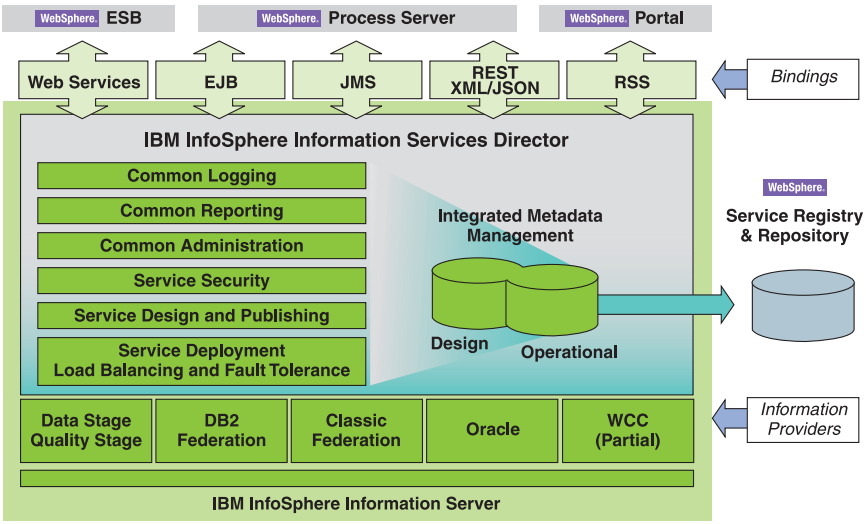

Figure 5: IBM InfoSphere Information Services Director

Storing critical data with IBM SOA for System z

Critical data is often held in data servers and processed by transaction managers. IBM DB2 for zOS and IMS are two such data servers that store, manage and deliver high end-online transaction processing systems.

There are multiple advantages to using IBM System z with IBM SOA.

- *The economics of virtualization and consolidation drive simplified information infrastructures.*
  - *Enables increased focus on business processes*
  - *Reduces system and information-related costs*
  - *Enhances growth with higher levels of business flexibility and responsiveness*

- *IBM System z Integrated Information Processors (zIIPs) and IBM System z Application Assist Processors (zAAPs) integrate information and transaction processing across the enterprise, freeing up processing on central processors and reducing overall processing costs for ERP, data warehousing and systems management workloads.*
- *Tightly integrated environments optimize XML and SOA data, serving information-intensive workloads.*
  - *High performance processing of IBM DB2 9 pureXML® through z/OS XML System Services*
  - *Faster XML parsing with IBM zIIP and IBM System z Application Assist Processor (zAAP) support and the integration and transparent running of IMS Java™ workloads on zAAP processors help reduce total cost of ownership*
  - *The natural fit of IMS for XML data storage, with both being hierarchical, offers universal information interchange with the most efficient use of resources and top overall performance*
- *System z optimizes communications information workloads between virtual environments. For example, IBM HiperSockets™ virtual network technology reduces access time to critical information across virtual environments.*

Improving application performance, security and manageability

IBM Data Studio Developer is an integrated database development environment that speeds application design, development and deployment while increasing data access efficiency and performance. Using Data Studio Developer enables organizations to develop and test routines, generate and deploy data-centric Web services, create and run SQL and XQuery queries, and develop and optimize Java applications.

Together with IBM Data Studio pureQuery Runtime, these offerings improve application performance, security and manageability. Integrated with the Rational Software Delivery Platform, Data Studio Developer complements and extends the capabilities of Rational Application Developer and Rational Developer for System z. Data Studio Developer enables stored procedures, SQL statement or XQuery statements to be transformed into a Web service for access to core data held in DB2 on System z or Informix®.

Accessing data sources

In addition to providing access to relational data in DB2 family servers, DB2 Connect™ provides access to non-DB2 mainframe data. This access is provided by using the stored procedure function of DB2 for IBM z/OS, and includes data sources such as VSAM files, IBM CICS® and IMS transactions and IBM WebSphere MQ queues. Access to these data sources is transactional, and data integrity can be validated through the z/OS RRS facility. For example, a DB2 stored procedure can update DB2 data and execute CICS and IMS transactions and verify that if any of these operations failed, the entire transaction would be rolled back. This is known as a "unit of work."

DB2 for z/OS with pureXML enables unified access and management of traditional enterprise data with XML data. It combines the strength of traditional database technology that provides performance integrity, protection and scale, with the flexibility of XML when persisting and retrieving data. As a result, organizations can manage both types of data in one uniform repository and provide unified access to and management of that data.

The pure XML support in DB2 offers efficient and versatile capabilities for managing XML data. Unlike XML-only databases, DB2 V9 also provides seamless integration of relational data and XML data within a single database or a single row of a table. XML can be queried in any of the following four ways:

- *Plain SQL (no XQuery involved)*
- *SQL/XML (XQuery embedded in SQL)*
- *XQuery as a stand-alone language (no SQL involved)*
- *XQuery with embedded SQL*

Integrating IBM Information Management System and SOA

IBM Information Management Systems (IMS) is committed to open standards, connectivity and new and emerging technology, with continual SOA enhancements being shipped in each release. It places core business data and transactions at the heart of SOA. It offers message queuing, transaction processing and database support and is fully integrated into application development tools.

IMS is focused on leveraging IMS assets, promoting integration, serving as an integration focal point, and supporting industry standards with rapid enablement for a wide variety of clients. The IMS integrated on demand solutions protect customer investments by enabling access to IMS transactions and data.

IBM has also introduced several solutions focused on new IMS application development. These solutions offer direct IMS database access from a variety of environments. IMS now can fully house and maintain a Java Virtual Machine, which enables IMS to process Java workloads. In addition, IMS offers Java class libraries, which contain a complete API for use by Java developers. IMS also allows use of these same libraries from several different runtime environments. These environments are WebSphere Application Server, DB2 stored procedures on z/OS, and CICS using the JCICS API. SQL, XQuery, and DL/I are also supported.

For all of these solutions, IMS also offers tooling support via the DLIModel utility. This is now an Eclipse-based GUI tool that offers visualization of IMS databases. With respect to new application development, the utility also generates database metadata definitions that are consumed at runtime by the Java libraries. This enables the libraries to convert an SQL query into a native IMS DLI call, for example.

All of these solutions position IMS as a major player in the emerging SOA marketplace.

Summary

Useful, uncontaminated information is vital to all businesses. IBM Information Management solutions for System z put information at the heart of an organization's SOA. It can help ensure that the right data is readily available. Information Management SOA solutions enable scheduled and dynamic cleansing, transformation and federation of data as services. The SOA approach abstracts data as a set of reusable data services, protecting applications and processes from lower level information changes and enabling solutions to be brought to market faster.

IBM enables "information as a service" as part of your SOA solution by providing database management systems and tools that render data servers, their data management functions, queries and access as a set of services.

IBM has been focused on SOA for many years and offers hardware, software, best practices, industry expertise and services to help ensure successful SOA implementations. With more than 7,000 SOA customers, IBM has numerous SOA success stories available for study. IBM has built a strong ecosystem around SOA, with more than 7,400 Business Partners providing value-added SOA services and offerings, more than 100,000 trained SOA professionals, and 13,000 assets in the IBM SOA business catalog.

For more information

To learn more about how IBM can help you establish an information agenda, contact your IBM sales representative or IBM Business Partner, or visit: **ibm.com**/software/data/information-agenda.

For more information on IBM SOA, visit **ibm.com**/soa.

For more information on the IBM DB2 for zOS family, visit **ibm.com**/software/data/db2/zos/family.

For information on the IBM IMS family, visit **ibm.com**/software/data/ims.

For more information on the IBM InfoSphere Server on System z, visit **ibm.com**/software/data/infosphere/info-server-system-z/overview.

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

Produced in the United States of America
March 2009

[1] "Accenture CIO Survey: Information Management Trends in Data Management and Architecture," Accenture, 2008. www.accenture.com/Global/Technology/Information_Mgmt/Information_Mgmt_Services/R_and_I/AccentureArchitecture.htm

[2] "The Emerging Vision for Data Services: Becoming Information-Centric in an SOA World," Mark A. Beyer, David Newman, Daniel Sholler, Ted Friedman. Gartner, April 2006.

[3] "More than 95 percent of the top Fortune 1000 companies use IMS" in "DataBase Magazine October 2008" http://www.ibmdatabasemag.com/story/showArticle.jhtml?articleID=211300235

ZSW03066-USEN-00