HURWITZ & ASSOCIATES

Information as a Service:

IBM Information Server

Judith Hurwitz, CEO

Marcia Kaufman, Partner

© Copyright 2006, Hurwitz & Associates

Introduction – Leveraging Information in the Age of a Service Oriented Architecture

The information an organization collects and manages is as important to a company's future growth and viability as its own products and services. If a company can provide a strong and well constructed information platform, it can better understand its customers and help foster innovation in order to surpass its competition. In many companies, the data needed to reach this milestone is locked in disconnected silos in different departments across the enterprise. Each data source has its own definitions depending on the context within a specific business unit. Today, business managers are asking questions in new ways that can only be answered by breaking the chains that have kept information tightly controlled and coupled within disconnected business applications. Satisfying customer needs requires a cross-enterprise view that can only be achieved with more effective business and IT collaboration and the use of new information technology. Successful organizations are addressing their most important business requirements by unlocking the information that has been kept behind closed doors.

The advent of a Service Oriented Architecture – which fosters the ability to easily connect and reuse information and software assets – has empowered IT management to develop a new approach to turn data into a much more powerful corporate asset. To achieve this goal requires a life cycle approach to the management of information that ensures that the business can trust where the data comes from, understands who is responsible for it, and ensures that it can be reused throughout the organization.

Turning data into a competitive weapon requires a comprehensive approach predicated on the ability to move away from disconnected data sources to information integration. Traditional approaches to information integration were dominated by independent products that were architected to provide point to point connections between different applications and multiple data stores. Typically, each situation required extensive custom programming. This approach works when there are a small number of data stores used in a controlled manner. However, for organizations to achieve a more accurate and actionable understanding of their customers, processes, and business opportunities requires a flexible, repeatable approach to managing

Turning data into a competitive weapon requires a comprehensive approach predicated on the ability to move away from disconnected data sources to information integration.

information. This goal is achieved through providing consistent interfaces that support the ability to change the business context without massive recoding or programming. To achieve this level of information management requires a movement to Information as a Service. The implementation of Information as a Service is called the Information Server. This paper will define the Information Server and its use of Information as a Service in the context for how it benefits customers.

Why Information Management Is Changing

In traditional approaches to information management, customer or product data are tightly connected to applications for a specific line of business or division. While this data may be well understood and controlled within a specific department, when these data sets are integrated with other departments, they are not necessarily consistent. The definition of a customer may vary widely based on the different business processes required by the marketing, sales, finance, and other departments within an organization. Because business units have different uses for the same data elements, they often create their own versions of data sources. Therefore, rather than leveraging the same set of consistent definitions, various functional areas have created their own definitions leading to a proliferation of inconsistent data sources. For example, the marketing team creates a database to support marketing efforts aimed at selling more products to high-value customers. To accomplish this goal, the team needed to define the customer as the aggregate corporate entity, the sales team creates an independent database where each customer is identified based on the individual unit that pays each invoice.

The resulting inconsistencies in definitions of critical data such as "customer" or "product" have caused information managers at many companies to feel that they have lost control over the integrity of their information. The multiple connection points required to establish a consistent view across the business costs too much to maintain and is too complex to ensure accuracy. The business suffers when managers do not have sufficient trust in the quality of their company's data and they are unable to make well informed decisions.

The business suffers when managers do not have sufficient trust in the quality of their company's data and they are unable to make well informed decisions.

Information as a Service: The IBM Information Server

When two companies merge, inconsistencies in data and data definitions often become more obvious and more troubling to the business. The management team often discovers that integrating data across the data stores of the newly merged companies results in misinterpretations of the combined data base. For example, if the definition of a customer varies between the companies, it becomes very hard to determine, without a lot of manual intervention, which customers are likely to generate the most revenue or which customers may present some type of risk to the business. The lack of a single view of the customer across the various lines of business can lead to missed opportunities for the business and dissatisfied customers.

There are many situations where organizations have to manipulate data in order to provide a consistent view of customers, products, or processes. It is always a complex endeavor with many independent steps involved. While there are often many manual steps required to eliminate inconsistencies in data, much of the data integration process used by businesses to link data sources together has been automated through the use of software programs that come under the heading of Extract-Transform-Load (ETL) tools. These tools are used to locate and access data from a particular data store (data extraction), change the structure or format of the data so it can be used by the business application (data transformation), and send the data to the business application (data load).

Under the broad umbrella of delivering trusted information to the business, there are many requirements that go beyond the capabilities of ETL technology. There are many "home-grown" and packaged software applications available to meet these requirements. The various applications include software for data matching, data cleansing, delivering value to the business from unstructured data, and governing and securing the data. Combining all of these functions together has been a cumbersome and complex task for many IT departments. One of the major challenges to this information integration approach is that the integration process has not been optimized to achieve data quality. One department might assume that its data is already accurate. However, when that data is used by another department or even within the same department for another purpose, data may not be as clean as the user assumed. Ironically, it is usually the business expert who understands the intricacies and the specifications of a specific data entity. That business expert has often not been adequately involved in

The lack of a single view of the customer across the various lines of business can lead to missed opportunities for the business and dissatisfied customers.

the information integration process because of the technical complexity of getting the job done.

A New Approach to Managing Information: The Information Server

In order to ensure that businesses can trust their information when analyzed across the enterprise, an information management strategy that enables a much higher level of IT and business collaboration is required. The IT organization needs to efficiently provide the business with the right data definitions in consultation with the business user. This consistency and collaboration can be optimized by implementing an Information Server – an integrated environment that provides a platform to bring the various types of data services together with a consistent set of interfaces. One of the benefits of implementing an Information Server is that there is a consistent approach when a user moves from one function to the next. These functions when combined together include the checks and balances to ensure that data is managed in a consistent and trustworthy manner. This integrated approach will dramatically reduce the level of complexity and costs involved in information management, allowing the organization to move away from its focus on programming to create the linkages from one function to the next. Instead, the standardized programming interfaces of this abstracted and integrated Information Server will enable the organization to focus its resources on the business requirements for consistent data.

There is a strong correlation between the Information Server and the Application Server. In the past decade, the Application Server that was developed as a way to allow data to be accurately and consistently passed back and forth between the data store, the application, and the web browser to the end user without explicit custom coding. Prior to the development of the application server, an organization would have to write custom code to make each and every connection work efficiently. Once this process was abstracted into the Application Server, it was possible to repeatedly perform this information passing process without coding for each instance. In addition, it enabled faster loading time at the browser because the new environment was developed and honed to be efficient. Just as the Application Server created a common language and approach to moving

This integrated approach will dramatically reduce the level of complexity and costs involved in information management...

information between applications and the web environment, the Information Server enables this consistent movement between data stores and business services – codified business processes.

Fast Forward to the SOA Environment

Service Oriented Architecture (SOA) has paved the way for businesses to begin looking at the role of data stores differently than in the past, when they were tightly integrated with departmental or even corporate applications. SOA is defined as a software architecture for building applications that implement business processes or services using a set of loosely coupled black-box components orchestrated to deliver a well-defined level of service. This approach to building IT systems is much more flexible than prior architectural methods enabling businesses to leverage existing application assets and combine them in new ways to support the business as it adapts to change. The two pillars of SOA – flexibility and reuse – have provided the catalyst for change in the approach to managing the information integration process.

Within SOA, business services are designed to be reused in many different situations. Likewise, the information used by these services has to have the same level of flexibility that allows for consistency and reuse. In order to be assured of this flexibility, a developer needs to be able to quickly access the data source at runtime and ensure that the data been profiled, cleaned, moved, transformed, federated, and secured according to the changing needs of the business. Needs change. When these various functions are developed into a series of reusable components, these functions can be called upon quickly to "service the data" and can, therefore, be referred to as data services. The business requires that IT provide it with some key benefits: consistency, reliability, predictability, and reusability of the information. The business users demand more standards for the critical information that defines their organization – especially outside of the departments that in the past had total control of data within packaged applications.

A lot of what is being done with SOA for data management is not new. The difference between SOA based data management and traditional management is related to the degree of automation that is possible. In the past, many of the steps required to achieve quality information required a high level

The two pillars of SOA — flexibility and reuse — have provided the catalyst for change in the approach to managing the information integration process.

of manual intervention. Except in cases of unlimited resources, this has meant that some of the steps could never be completed accurately, if at all. With a SOA-based approach to data, including such principles as the loose coupling of data to applications, the level of consistency, quality and reuse is magnified. For example, the marketing department will be able to leverage the data created by three different product lines in order to understand which customers are responsible for most revenue over time. This would not be possible if each product area created its own unique definitions of their customer data. If the definitions and other information about the data (metadata) are consistent, then the data can be maintained in multiple data stores. It is not necessary (or even wise) to create one all encompassing and centralized data store to control and synchronize all the various data definitions. At most large organizations, this approach would be impossible to manage, and too costly to implement.

Hurwitz & Associates calls the application of the SOA approach to the management of data assets "Information as a Service." Information as a Service loosens the tight connections between data and applications so that data can be controlled and shared across the enterprise. This approach allows businesses to reach a consistent view of enterprise-wide information that has been previously very hard to achieve. The foundation for information as a service is the metadata repository.

The Role of the Metadata Repository

To meet the requirements for information as a service requires a common language and common understanding of data at all levels of the enterprise. One way to achieve this common understanding is by developing a metadata repository. Metadata is defined as data about your company's data. The definitions, mappings, and other characteristics used to describe how to find, access, and use the data are all called metadata. In order to establish common knowledge and a shared understanding of these definitions and other characteristics of the data, the metadata is collected and maintained in a unified way in a container called a metadata repository.

By placing the definitions of business data and rules for mapping data to their actual physical location in the system in a metadata repository,

Information as a Service loosens the tight connections between data and applications so that data can be controlled and shared across the enterprise.

the business is able to achieve a much higher level of control and over the consistency of the data. It is this consistency and reliability that will enable data that is linking from different sources to come together correctly. The semantics and rules that apply to all your company's data can be organized, tracked, and managed through the metadata repository. The abstraction of rules, definitions, and other instructions from the data stores provides your business with a way to achieve consistent data while still maintaining your extensive investments in data management assets. The metadata repository ensures that the data is of the right structure and quality before it's consumed by a business service.

A more complete technical term for the metadata repository is a metadata abstraction layer. This abstraction layer includes the rules, definitions, and mapping instructions about the data that are either replicated or separated from the data stores. The process of abstracting the data rules and definitions adds flexibility to the data infrastructure, which provides programmers with a way to loosen the tight connections between data stores and specific business applications. In addition, this provides technique for IT and business management to establish a data stewardship standards that increases the collaboration between the organizations. This repository resides in a technical layer between the actual data stores and the business services. The metadata repository is often referred to as a metadata layer because of its strategic position in the information infrastructure.

The metadata repository ensures that the data is of the right structure and quality before it's consumed by a business service.

The Business Requirements for Providing Information as a Service

The priorities of the business can take center stage when some of the restrictions and constraints from the more rigid IT systems of the past are removed. Information as a service enables the business to optimize the delivery of required information based on the following four main requirements:

- **Enable quality data.** The IT organization needs to be able to certify the quality of the data. The information that comes from many different sources must be trusted to be accurate and consistent no matter how the data ends up being used. You need to be able to get the quality data you need when you need it.

- **Enable consistent data.** The meaning of data needs to be unambiguous so it can be interpreted and processed appropriately both by businesspeople and by machines. The definitions of data should be based on the way the business as a whole needs to understand sales, customers, products, and profit. Businesses use various software tools to ensure that when multiple sources are integrated, the data from those sources will work well together. These tools are used to profile, ensure quality, transform, audit, and otherwise create trusted data. All of these data processes can be exposed as data services so they can be shared and used across all of a company's various data sources.

- **Enable independent data.** If the data is loosely coupled from its original sources, those data elements can be more easily brought together in different ways to meet many different business needs. The data can be retained in the existing sources, but the loosening of lineages provides the flexibility needed to deploy a federated approach to bring them together.

- **Find the proper balance between control and flexibility.** It is critical to allow departments to have control over the data that matters most to them. At the same time, there should be enough flexibility to allow for a single, consistent view of customers or products at the enterprise level.

The IBM Information Server

One approach to delivering information as a service has been introduced by IBM as the IBM Information Server, a unified platform that aggregates technologies to enable an efficient and consistent information integration process. It is based on SOA principles so that each component is designed both to work on its own, if necessary, and at the same time to work as part of a loosely coupled environment. The goal of the Information Sever is to enable customers to move efficiently from understanding to cleansing to transforming the data sources within the organization and ultimately to delivering consistent and trusted information to the business users.

The priority for delivering trusted information to the business is to begin with a higher level of business/IT collaboration.

The priority for delivering trusted information to the business is to begin with a higher level of business/IT collaboration. The only way to ensure consistency and accuracy of data is to bring various business and IT

constituents together to foster a shared understanding of business terms. By aggregating the common functions required for information integration into a unified architecture with role-based (such as the data owner or the business user with no authorization to make changes) user interfaces, the IBM Information Server is designed to help the business achieve this goal. The products can be used separately, but the combined capabilities embed data quality tools in the information integration process. IBM's goal for customers who deploy the Information Server is to bring together the components of an information management strategy so that the overall process of integrating data can be reused in many different business situations rather than applying these efforts to a single project, and having to recreate those same processes when another project comes along.

IBM Information Server includes reusable services for understanding, cleansing, transforming, and delivering trusted information across the enterprise. The capacities include:

- **Improve data quality.** Understand the meaning, context, and lineage of data to ensure a unified, consistent, and trusted view of all the business. Reducing the risk of passing error prone or misleading data from one data source to another returns enormous value to the business. Dealing with more trustworthy data means that projects can be streamlined and productivity improves. In the Information Server, the data cleansing capabilities are coordinated and use the same user interface as the data transformation capabilities, making the data quality software easier to use and more accessible.

- **Simplify data transformation.** Understand, integrate, and deliver information from multiple sources efficiently and accurately. The information integration process becomes much more adaptable and efficient if the transformation process is easier to work with. The IBM Information Server simplifies this process by included many transformations functions in a pre-built ready to use format.

- **Provide a complete view of structured and unstructured information.** Access and integrate data in real-time leveraging metadata and federated design schemes. Because the same logic is used at the core of all of the data services, accurate information can be more easily delivered whenever it is needed.

Dealing with more trustworthy data means that projects can be streamlined and productivity improves.

The main components of the Information Server are:

- **WebSphere Information Analyzer.** This component analyzes the structure and determines the relationship between data across fields and sources. The business-side subject matter experts and data analysts can use this product to ensure that integration of data sources is based on a thorough understanding of the available data.

- **WebSphere Business Glossary.** This is where the business metadata –such as how the data are used and the rules for governing that data – are recorded to provide a business context for the data. Business subject matter experts and other business users may deploy this component to establish an understanding of how the data elements are defined by the business.

- **WebSphere QualityStage.** Based on certain rules that can be adjusted by the business, this product is used to clean up the duplicate records and enable organizations to create a single view of their customers or products across various data stores.

- **WebSphere DataStage.** This product is designed to carry out the transformation of information across data stores. It is the core for the primary ETL (extract, transform, and load) functions of the integration process.

- **WebSphere Federation Server.** This technology helps hide the complexity of the various data stores from the end-user. It provides a way to access data across the organization regardless of its location, format, and application.

- **WebSphere Information Services Directory.** This module provides a unified mechanism for publishing and managing shared services across data quality, data transformation, and federation functions, allowing information specialists to easily deploy services for any information integration task and consistently manage them.

- **WebSphere Metadata Server.** This is the metadata repository for IBM Information Server. The metadata services infrastructure of IBM Information Server is designed to allow metadata to be more easily managed, accessed by those who need it, and shared across heterogeneous technologies through a SOA.

One of the framework's most important accomplishments is that there is shared metadata across the platform.

One of the framework's most important accomplishments is that there is shared metadata across the platform. In addition, the platform includes a role-based user interface. The combination of these two factors helps increase the level of collaboration across business and IT.

Conclusion

One of the key requirements for delivering information as a service is that the meaning of data needs to be unambiguous. The data needs to be understood, interpreted and processed appropriately by business managers with different needs and responsibilities for the business. In order for the IT managers to make effective use of technology to move, understand, analyze, and report on this data, the business and IT must all speak the same language. This is a practical and simple concept, yet in reality it is so hard to accomplish because of the complexity and disconnected structure of enterprise IT systems. The framework of the IBM Information Server is designed to overcome many of these challenges and encourage effective communication between the business and IT. It is designed to provide a way for all the key players in the information management process such as the data steward, data architect, business analyst, and developer to put a stake in the ground and work together to create trusted data that can then be leveraged across the business.

The Information Server is designed to deliver trusted information as a way of driving business growth and innovation. By providing an abstracted model of the technical infrastructure, the Information Server reduces the cost of leveraging information across the organization. It enables the company to be able to achieve a single view of the customer without having to resort to expensive custom coding. The end result is that the Information Server improves the productivity of the business by improving the ability to understand the context of data to the business. Overall, the business is better able to understand its customers and products, which translates into a more predictable way to manage growth and mitigate risks. The Information Server helps businesses audit the quality of their data and therefore improve the business climate.

The end result is that the Information Server improves the productivity of the business by improving the ability to understand the context of data to the business.

How Do You Evaluate Your Requirements for Information Management?

Hurwitz & Associates believes there are ten key questions that a customer should be asking if they are going to be able to leverage their information as a trusted asset and deliver that information as a service to business users.

1. Can silos of data be joined together quickly to address a business need?

2. Is the data consistent and accurate from one business entity to the next?

3. Are the data sources in the right context for the business problem being solved?

4. Is the data itself accurate? Are there duplicates?

5. Are there business rules that determine how data can be used based on corporate governance requirements?

6. Can the right information be delivered to all business users across the organization as needed in a repeatable and consistent way?

7. Do you have a way for business users to verify the accuracy of their data?

8. Can you track the origins of the data you are working with?

9. Are your data sources tied to a specific application?

10. Can you use your data to improve your customer relationships and the value of your products?

About Hurwitz & Associates

Hurwitz & Associates is a consulting, research and analyst firm that focuses on the customer benefits derived when advanced and emerging software technologies are implemented to solve pragmatic business problems. The firm's research concentrates on understanding the business value of software technologies, such as Service-Oriented Architecture and Web services, and how they are successfully implemented within highly distributed computing environments. Additional information on Hurwitz & Associates can be found at www.hurwitz.com.

Judith Hurwitz is CEO of Hurwitz & Associates. Marcia Kaufman is a Partner of Hurwitz & Associates.