Your SOA Could Be DOA Without MDM

A service-oriented architecture (SOA) can deliver such benefits as faster application development and the ability to build powerful composite applications. But to fully harness the power of SOA, an organization must effectively manage its information assets. Services whose data sources rely on redundant, inconsistent, inaccurate, or outdated data will not deliver the expected benefits or yield the desired results. That's why organizations are increasingly implementing a comprehensive master data management (MDM) solution as part of their enterprise SOA strategy. Doing so can eliminate issues with data inconsistency, align enterprise data assets, help disseminate accurate information within and outside the enterprise, and help ensure that SOA initiatives deliver all of their potential benefits.

This paper discusses the role of MDM, or enterprise information management (EIM), in supporting and enhancing SOA initiatives.

TIBCO®
The Power of Now®

"The harmonization of product and customer data provides a cornerstone on the road to SOA. With true master data, web services and the related business processes will become more accurate, timely, and efficient, leading to improved ROI on existing investments as well as improved business intelligence."

Forrester Research, From Trends 2006: Master Data Management; March 6, 2006

Introduction

SOA delivers value in two key ways. First, it reduces cost by offering a smarter, better way of developing applications. IT organizations can develop applications using building blocks of reusable, standards-based services that can be combined and orchestrated to produce functionality faster and cheaper than before. And the applications developed are easier to maintain, because developers can change components of the application without overhauling the entire application.

The second promise of SOA relates to composite applications. Composite applications allow developers to leverage pieces of existing functionality and data that may be locked within independent applications and combine them to rapidly produce new business services or applications – for example, a cross-sell/up-sell application. A bank, for example, may have data on a specific customer in the retail banking branch, in the home mortgage branch, and in the student loan branch. With customer information siloed by product it becomes hard if not impossible to up-sell or cross-sell intelligently. A composite application solves this challenge by cutting across system and geographical boundaries.

But, along with the promise, SOA exposes all manner of additional and buried data problems, especially when it comes to correlating data across systems.

Origins of Data Inconsistency

Systems store data according to their own purposes. Each system has its own taxonomy, format, and context. For example, a billing system will store such information as product price and cost of goods sold but will not store information necessary for order fulfillment. Out of 100 to 200 possible attributes, a given system might only track 40 or 50. A lot of information is stored outside of systems, in spreadsheets and paper. Furthermore each system has its own set of rules for entering new information and a lot of information is entered and re-entered manually, a process that is prone to errors.

This fragmented, system-centric view of information, combined with ad hoc processes for updating information, leads to redundancies and inconsistencies – particularly in large, complex organizations with multiple geographies, business units, and IT systems. The impact of data inconsistency is felt throughout the organization, resulting in: operational disruptions in critical business processes, such as customer service or order fulfillment processes; a lack of visibility into business activities; inaccurate information being fed to decision makers; and a waste of resources as employees spend time resolving errors resulting from bad data.

Consequences of Data Inconsistency in an SOA

Until data is reconciled syntactically and semantically, organizations can't reap the full benefits of SOA.

The problem of data inconsistency isn't new, but in an SOA environment it is more evident and problematic. In a closed environment, it may not matter that SAP data is inconsistent with Oracle, Siebel, or legacy system data. With composite applications, which cut across systems, inconsistencies become problematic. For example, to take advantage of up-sell opportunities in the retail banking example, it's necessary that a customer's name be represented consistently across systems or it will be difficult to identify that the accounts belong to the same person. Similarly, if a company is ordering the same product from an SAP system and an Oracle system and it's not recognized as the same product, the company may miss out on strategic sourcing benefits. While rationalizing vendor information for strategic sourcing is a big problem for many large organizations independent of SOA, such problems are now often tackled with an SOA strategy. To receive the true benefits promised by those strategies, addressing data inconsistency becomes paramount.

Until data is reconciled syntactically and semantically, organizations can't reap the full benefits of SOA. Syntactic reconciliation, which refers to bridging the gap between the different data formats that individual systems use, has been solved by providing transformation maps from the data format of one system to another. This has been the domain of integration providers, and the technology to address this is quite widespread and well understood within IT organizations. However, enterprise application integration (EAI) technologies do not solve the semantic integration challenges. EAI moves data where it needs to go in the format it needs to go in but does not have knowledge of whether that data is accurate or consistent. Nor does it know how to resolve conflicts in the data. EAI is charged with keeping systems in sync, but once they are out of sync, because of externally imposed changes, EAI has no way of preventing or resolving the problem. EAI provides essential functionality but inherently presumes that semantic consistency is already in place, which is often not the case.

MDM serves SOA projects by making master data available on-demand to any consuming application.

Reconciling Semantic Data Inconsistency with Master Data Management

Master data management (MDM) software enables organizations to align enterprise master data assets (product, customer, vendor, etc.) across multiple systems and departments and with trading partners. It also enables organizations to support the necessary processes, policies, and procedures to ensure that information stays accurate and consistent as new information is added and updated. Further, it ensures that clean master data is disseminated back to the transactional systems and decision makers that rely on it for day-to-day operations. By doing so, an MDM solution ensures that data within the organization is consistent across systems and geographies, which is critical to successful SOA initiatives.

In addition to driving consistency, an MDM solution also serves SOA projects by making master data available on-demand to any consuming application, such as a composite application, via standards such as web services. An MDM solution allows organizations to create a master information services platform – a set of services that can be invoked in a standard way by consuming applications and processes. These services can be granular data services, such as 'Query Record' and 'Delete Record', or can include business logic such as 'Add Customer' and 'Update Address'. By delivering such a platform of data services, developers of business services and composite applications within an SOA can focus on delivering high-value business functionality rather than worrying about how to access and update the underlying data. As developers of services that affect one part of the organization start to access data in another part of the organization they are going to require assurances, in the form of service level agreements (SLAs), that the underlying master data will be accurate, consistent, and available as needed in the format they need it and in a secure and reliable manner. This level of trust and ability to provide internal SLAs is critical in achieving the type of benefits promised by SOA.

To build these master data services, three things become increasingly important: data governance, high performance, and metadata management.

DATA GOVERNANCE

Just as there is a need for SOA governance, to ensure proper levels of security and performance for services, there is a need for information governance to resolve such issues as: Who can view the data? Who can edit the data? If there is a conflict between customer data in Siebel and SAP, which system wins that conflict when choosing the

Organizations can clean up their data and spend a fortune on consultants, but without effective governance, they won't achieve lasting data consistency.

customer's correct address? If a product weight changes, who should be alerted? What is the process for adding a new customer or new account? What are the required attributes? What are the optional attributes? All of these rules – and many more – comprise data governance. Organizations can clean up their data, remove inconsistencies, and spend a fortune on consultants, but without effective governance, they won't achieve lasting data consistency. The consequences of inadequate governance can often be quite expensive for organizations. If a bank were to receive different information about a security from two feeds such as Reuters and Bloomberg and did not know how to properly rationalize the inconsistency, the result would be a failed trade and an expensive manual settlement process. For other companies it translates to poor customer targeting and service, missed or wrong shipments, and unhappy vendors and customers.

Given how important data governance is to running a business, it cannot be an IT mandate. It must be incorporated into day-to-day business processes and become the responsibility of the natural business users and owners of that information. For example, a person in Logistics is most concerned with product dimension data. IT should give that person the tools to effectively maintain that information. Effective governance ensures that clean data stays clean for a long-lasting, high-ROI MDM solution.

Data governance also supports compliance with contractual agreements and with regulations such as Sarbanes-Oxley and Basil II. As information in one part of the organization is leveraged by services built by another part of the organization, it is important for detailed audit trails to be maintained, recording what information was accessed or modified by whom and when. This gives data owners the assurance that data is being handled by authorized users, according to an approved process that complies with regulatory mandates.

HIGH PERFORMANCE

For a master data solution to be effective in providing transactional services and composite applications with the master data they depend on, it needs to be efficient, reliable, and scaleable. There are many ways to achieve high performance including loading up on hardware, replicating the master data system in different geographies to minimize long network hits, and using advanced techniques such as distributed caching and multi-threaded processing.

Additional hardware is an expensive option and typically the last resort. Maintaining multiple master data nodes can be beneficial as long as effective synchronization between those nodes can be seamlessly managed. The last thing an organization needs is conflicting master data repositories. The third option is the most technically sophisticated.

An MDM solution is well served when data that needs to be repeatedly accessed by SOA applications and services is pre-loaded into a cache. This cache should be distributed across nodes and make optimal use of available hardware. Other techniques such as breaking down large master data record sets that need to be imported, validated, processed, and published, into smaller batches that can be processed in parallel by multiple threads also considerably enhances the availability of rich, accurate, consistent data to the rest of the IT ecosystem.

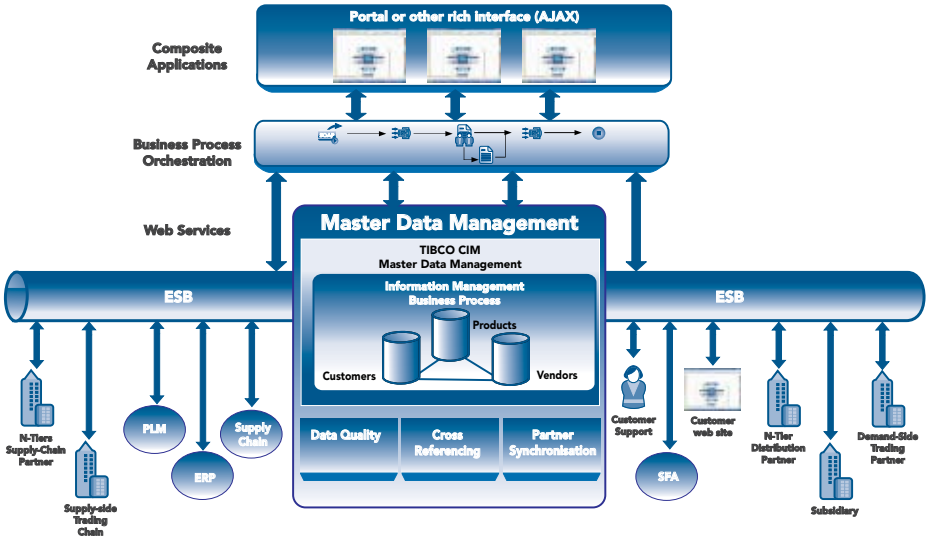

Figure 1. Composite applications in an SOA leveraging a master data services layer.

METADATA MANAGEMENT

Managing metadata is an important part of both SOA and MDM initiatives. Metadata describes the data model and hence forms the DNA of information architecture. Metadata also stores information about other aspects of the data such as mappings, validation rules, business processes, and so on. Metadata defines not only the structure of the data but also serves as a "cookbook" for processing that information as it moves across the organization. Therefore, it is important to have a systematic way of managing metadata and treating metadata as an important asset. This includes specifying roles-based access to metadata and a workflow with approvals for modifying metadata. When metadata changes, it is helpful to be able to publish reports on impact analysis, that is, where that metadata is being used. It should be stored and managed in accordance with standards so organizations can easily share it internally and even with trading partners if necessary. Finally, it is important to properly version metadata for compliance and auditing purposes, especially when financial data is involved. While metadata management, in general, is a large task for IT organizations – including managing the metadata for both master data

and transactional data – metadata that represents the models and governing rules for master data specifically plays an important role in planning and implementing an effective MDM strategy.

Enabling Data Services

Metadata also plays a crucial role in enabling data services. Master data services that provide standards-based access to underlying master data stored centrally or distributed across the organization should be built on top of metadata or be "metadata-driven" rather than hard-coded. This allows the services to adapt to a change in the underlying data structures.

It would defeat the purpose of moving to an SOA if developers had to understand the structure, validation rules, and policies that govern the underlying customer data.

These data services allow developers of business-level services to easily access the master data they need so they can focus on building the value-added functionality their service provides. For example, a developer building a cross-sell composite application might need to build a service called 'Get Customer Credit'. Ideally that developer would have access to master data services such as 'Get Customer Name' or 'Get Customer ID' as building blocks, greatly facilitating productivity. It would significantly slow application development and defeat the purpose of moving to an SOA in the first place if that developer had to re-invent the wheel to understand the structure, validation rules, and policies that govern the underlying customer data every time a new service was rolled out.

Equally important to increasing productivity, data services protect higher-level services such as 'Get Customer Credit' from breaking when the underlying data structure changes, similar to the abstraction that an adaptor provides when interfacing with an application. Any time a system needs the customer's internal credit score, the same 'Get Customer Credit' master data service can be called, which creates consistency. Any change in functionality of that service can easily be implemented in one centralized place and, because the interface remains the same, other services that rely on 'Get Customer Credit' will not break. This is exactly the flexibility and rapid development that justifies moving to an SOA. Thus, a collection of master data services allows organizations to fully leverage the benefits promised by an SOA.

The Role of Events

Events are constantly being generated by systems and people throughout the enterprise and across the value chain. Examples include accepting a customer order, receiving a shipment, or issuing an invoice. Complex event processing (CEP) software can help organizations capture, correlate, and analyze events in real time against established patterns to take advantage of opportunities or automate corrective actions before problems snowball. An example might be discovering that fraud has likely occurred at a credit card company, based on a series or pattern of suspect events. Another example would be a series of delayed shipments to a large customer, a possible precursor to canceling their business. MDM can give context to events by providing the master data that describes the entities that are interacting during the event, such as the product or customer information that is being acted upon. A CEP engine could pull information from an MDM application via data services but this might not yield the high access speeds required to keep up with the thousands of events constantly being generated at large organizations. A high-speed cache embedded within the MDM systems, distributed across multiple nodes serving up contextual master information including relationship information in real time, would allow a CEP engine to effectively analyze and apply rules against different event sequences constantly and rapidly occurring. If architected correctly, the combination of MDM, SOA, and CEP can yield a powerful infrastructure on top of which customers can build solutions to better visualize their business activities, understand opportunities and threats, and act accordingly.

THE CASE FOR MDM: RISK MANAGEMENT AT AN ENERGY TRADING COMPANY

A multi-national energy trading company needed to get better insights into its risk exposure. During day-to-day operation, traders need to know the risk exposure they face when doing business with a particular counterparty in order to fully evaluate and execute the trade. This is a difficult challenge as the company has counterparty data spread out over a highly dispersed IT operation, with over 60 instances of SAP and over 250 custom applications. The company operates in a dynamic market with numerous joint ventures and partnerships.

Initially, the IT department decided to invest in an SOA infrastructure to solve the data fragmentation problem. They planned to build composite services that would access all their different trading systems – across geographies and products – and aggregate the information to create a counterparty-centric view, which would then allow them to run their risk algorithms. But as they got into the project, they realized there was an issue with the underlying data. Until they reconciled counterparty data so that it was represented in a common way across systems, they would not be able to aggregate it and turn it into meaningful information.

In addition, the IT organization had embarked on a global business process management (BPM) project to streamline trade discovery, execution, and settlement processes. Each of these processes had sub-flows that needed to access, create, and modify other types of master data (counterparties, vessels, pipelines, vessel owners, etc.). Information spread across a multitude of systems with no consistent source of master data was a serious hindrance to the success of the BPM project and the organization's ability to settle trades without significant and costly manual intervention.

After identifying the root problem hindering their ability to better assess counterparty risk and efficiently process and settle trades was MDM-related, the company chose TIBCO's MDM solution to first reconcile discrepancies and create a consistent view of counterparty data across the organization. Master data entities, such as counterparty, always have meaningful and often complex relationships with other types of master data, such as the relationships between a product and a vendor or a product and a store location. Thus the solution required the creation and automation of numerous complex business processes that not only had to understand counterparty data but also information about wells, pipelines, and other assets – each master data type with its own lifecycle and governance processes. Many of these business processes are long-running and can last from three to six months. During this time it was imperative that changes to the data or the data model (metadata) be properly managed, versioned, and tracked so as to reduce operational risk.

By providing a comprehensive MDM solution, TIBCO enabled the company to streamline trade processing and gain the kind of insight into risk exposure that will enable traders to make decisions based on more accurate, complete information.

SOA + MDM – The Integrated Enterprise

For many organizations, the ultimate goal is to create an integrated enterprise – what TIBCO calls the real-time enterprise. SOA connects people, processes, and information by integrating systems and providing a platform to develop new functionality while getting the most out of existing investments. MDM provides the information management component – ensuring that critical information assets, the lifeblood of the organization and foundation for other solutions, is aligned internally and across the value chain and delivered to people and systems alike in real time.

The combination of SOA and MDM allows organizations to be connected in real time at nearly every level, from processing day-to-day activities to making strategic decisions. Some organizations consider MDM to be a subset of the overall SOA strategy rather than a separate discipline, further reinforcing the role of MDM within an SOA. Regardless, SOA augmented with MDM provides IT organizations the tools to boost productivity and cut application development costs. More importantly, it gives them the agility to rapidly roll out new solutions that take advantage of accurate up-to-the-minute data about the company's products, customers, and vendors. It ensures that information is being properly managed and utilized throughout the organization to help the company achieve its business objectives.

For more information, visit www.tibco.com.

TIBCO®
The Power of Now®

Global Headquarters
3303 Hillview Avenue
Palo Alto, CA 94304

Tel: +1 650-846-1000
Toll Free: 1 800-420-8450
Fax: +1 650-846-1005

www.tibco.com