**Data-centric architectures**

Enterprises are increasingly focused on unifying their enterprisewide data and designing architectures to maximize the usefulness and accessibility of that data

[](https://www.infoworld.com/author/Doug-Dineley/)By [Doug Dineley](https://www.infoworld.com/author/Doug-Dineley/)

Executive Editor, InfoWorld | MAR 11, 2005

Enterprises have always been concerned with data quality and integration. But the interest in improving data and content management is clearly on the rise, as companies are increasingly focusing on unifying their enterprisewide data and on designing architectures to maximize the usefulness and accessibility of that data.

The reasons are at least twofold. First, the costs of error-ridden, inconsistent, and obsolete data are high, in terms of slowing business processes and hindering automation. Second, business leaders are keen to take more information into account -- either structured or unstructured, from both transactional and content systems -- when making decisions, and too much information remains locked away in silos.

[look at the end of that list and we can see that a number](https://www.infoworld.com/article/2671590/link)

For many large companies, a data-centric architecture starts with rationalizing the “master data” -- the identities and attributes of customers, products, employees, and other core reference data -- at the heart of the business. In a global enterprise, customer or product data is typically spread across dozens, even hundreds, of implementations of CRM, ERP, and other systems, often from different vendors.

Each set of data is typically tailored to a specific business need -- engineering, sales, or marketing -- and location. The result, from the top-down view, is a sea of fragmented data that leads inevitably to faulty BI.

The emerging class of master data management solutions from Oracle, SAP, Siebel, and other enterprise application vendors attempts to bring order to this chaos. Oracle’s Enterprise Data Hubs, for example, combine a publish-and-subscribe mechanism, process automation based on configurable rules, and a knowledge base that helps data managers reconcile differences among source systems. Some solutions, such as Siebel’s, throw in business analytics capabilities. But all master data management solutions aim to create a canonical master data set that gets pushed to all kinds of data repositories -- mainframes, transactional systems, data warehouses -- throughout the organization.

The goal is not merely to synchronize data across systems but to improve data quality and to deliver as a service accurate, consistent data to transactional and operational systems. “It isn’t simply a matter of connecting the plumbing between many different data sources,” says Robert Shimp, vice president of technology marketing at Oracle. “There’s a quality function that has to be applied, to clean, dedupe, and reconcile all of this information. You don’t just need data; you need services-based information.”

In addition to mastering the master data, enterprises are also beginning to bridge the gaps between structured and unstructured data sources, as new technologies and techniques -- especially XML, SOAs, and enterprise search -- are making it easier and less expensive to do so. IBM’s WebSphere Information Integrator, for example, can combine SQL-, object-, and content-oriented access methods -- as well as enterprise search techniques -- to perform queries across relational databases, XML stores, mainframes, file servers, content management systems, even e-mail systems.

According to Eric Sall, IBM Software Group’s program director of information integration, the benefits go beyond the obvious operational advantages, such as a user of a CRM application being able to view an open trouble ticket in the customer service system. The pervasive, on-the-fly querying capabilities of enterprise search also extend the capabilities of traditional BI to include real-time data not yet loaded into the data warehouse.

Looking ahead, Oracle’s Shimp thinks this universal approach to searching and reporting will eventually put the data warehouse to pasture. The key enablers here are databases that can store relational data and native XML together.

“Traditionally, people have had to load and unload data, cleanse it and reformat it, do all kinds of complex gyrations, add all kinds of banks of servers for separate OLAP or data mining applications,” Shimp says. “That’s all going away. We’re simplifying down to just a core database that can handle all of this directly inside the database engine.”

It will take some time before we reap the full benefits of services-based information and universal data access. OASIS and other standards groups continue their work to establish the core identities and semantics within vertical industries and across them so that companies can more effectively share information through XML. Meanwhile, the walls between database silos, application silos, and organizational silos are coming down.

As IBM’s Sall puts it, “You can’t be compliant in a silo. You have to be able to look across silos to have any prayer of being compliant as an organization. Same thing with business intelligence. You don’t want to be intelligent about a silo, and not about the silo next to it. These are the reasons why this kind of more holistic or enterprise view of information is beginning to be such a big issue with the industry.”