the key to managing data is managing metadata

Enterprise Information Integration

"Enterprise Information Integration, seamless integration of disparate data sources on an enterprise scale, provides the strategic advantage organizations require."

Enterprise Information Integration System

MetaMatrix presents the first solution to the EII problem. The MetaMatrix product suite, with its flagship MetaMatrix Server™ and MetaBase™ repository, provide a platform through which an enterprise can manage and provide uniform access to disparate information systems for timely application development in a hyper-competitive inter-networked world.
Enterprises of all descriptions have long sought to integrate and manage all of their information assets within a single system. The goal remains elusive because enterprises create and store their assets in a myriad of disparate systems—relational databases, mainframes, different operating systems, free text, hierarchical repositories, and so on. Using new and existing data assets in an efficient, integrated, and interchangeable manner has become the key to surviving, and thriving, in the new economy. Enterprise Information Integration, seamless integration of disparate data sources on an enterprise scale, provides the strategic advantage organizations require.

Consider a typical business scenario in the new global economy. An enterprise trying to integrate its supply chain across its international divisions needs to unify disparate data sources inherited from numerous mergers and acquisitions. Valuable information is locked in isolated systems—Oracle™ databases in Europe, DB2™ in Asia, a proprietary message-based system in North America, and XML files in South America. How can the enterprise pull all this information together and integrate it to provide a business advantage?

The answer: Enterprise Information Integration (EII). EII provides access to information, regardless of source or storage format. EII uses existing descriptions of the physical data resources’ characteristics to construct a virtual representation that hides the differences of the underlying data sources. The enterprise can deploy business-critical applications more quickly and cost-effectively. How does EII accomplish this magic? Through metadata.

### Metadata’s Role

Metadata means “data about data.” What kind of information is in the system? How is it named? How is it formatted? How is it indexed? How is it organized? However, this simplistic view usually refers to technical data definitions and, while critical, must include an understanding of metadata as related to the end-user of information. From a knowledge worker standpoint, metadata represents a non-tangible information context mechanism that, implemented correctly, provides the user a common and integrated view of an organization’s information infrastructure.

When an enterprise establishes this information context, its users reap many benefits:

- Information needed to understand, access, and use an organization’s information resources.
- Enterprise-wide standardization and maintenance of the organization’s business definitions and business rules for data.
- Dynamic information dictionary, a common definition of data structures and the data content represented by component data elements.
- Enterprise-wide decision support.
- Data lifecycle management. Metadata captures the history of data over time.
- Information exploration and extrapolation.
- Information exchange among applications, both vertically and horizontally, within an organization’s information architecture.
- An information baseline to help guide the data extract/transform/load (ETL) processes.
- Allows the reusability of data without the risk of reinterpretation.
Metadata and Metamodels

Many people working on the problem of Enterprise Information Integration (EII) discover that the solution must center on metadata management technology. Ongoing efforts of standards bodies address the issues of metadata management, if not complete information integration. MetaMatrix aims at both problems. MetaMatrix provides flexible, scalable, efficient, and fault-tolerant data access to disparate data sources through the use of metadata.

The only solution to this problem is the construction of metadata-based models for each physical source and the integration of these models into metamodels. The MetaMatrix Server’s sophisticated query technology executes requests against a runtime metamodel to retrieve data from its native sources.

Metadata is at the heart of data management technologies used for decades. File systems provide metadata information about the organization and contents of files on computer systems. Relational database management systems use metadata, called schemas, to describe the table structures which store much of the today’s critical enterprise data. IMS databases contain data in hierarchical records whose structure is defined by metadata. Enterprises have used metadata to manage their data resources for a long time.

To support portals for B2B and B2C interactions, enterprises must integrate the information in these disparate systems and make the information readily accessible. But how? The enterprise uses a model of how each system stores its data. This model, the metamodel, contains a model of the metadata in each system. A metamodel models metadata; in other words, it is a higher-level model that describes the different types of models that describe how each system stores data. This hierarchy of models and their relationship to the underlying data is illustrated in Figure 1. EII uses metadata as all other data access technologies do. The key difference in EII, its uniqueness, lies in modeling the different types of metadata – the use of meta-metadata. Using metadata in a more abstract form lets an enterprise create relationships between information in different information systems. Using EII, an enterprise can relate a customer record in a relational database management system to a record in a mainframe IMS system despite the inherent differences in these two systems.

Metadata management involves storage and access to metadata information in a metadata repository, as well as a means to define and manipulate metadata models through a graphical interface. Figure 2 displays a typical architecture for a metadata repository.

Hundreds of Fortune 1000 companies, working through the Object Management Group, or OMG, seek to define new standards for storage and interchange of metadata information. The OMG promotes a UML-based means of describing and manipulating metadata. The OMG has also proposed a specialized metadata model for data warehousing. Many of the world’s largest modeling companies have accepted the XML Metadata Interchange standard, XMI.
Information Integration through Metadata

EII not only defines how pieces of information relate to one another, but how the enterprise can access the data, and provides a means of accessing that data. This, in addition to metadata management, is a fundamental part of EII: access to data, in a seamless, uniform manner.

MetaMatrix provides products that address the core requirements of an Enterprise Information Integration solution: metadata management and integrated data access.

Figure 3 depicts the MetaMatrix EII solution, which combines an enterprise-class metadata repository, the MetaMatrix MetaBase™, a storage facility for versioned metadata models; the MetaMatrix MetaData Modeler™, a tool for importing and modeling metadata; and the MetaMatrix Server™, a scaleable, fault-tolerant data access server. Together these products form the first EII solution that provides seamless, uniform access to all enterprise data.

The Next Level of Integration: Virtual Metadata

The MetaMatrix solution provides a means to define, manage, and integrate metadata for physical data models throughout the enterprise. An enterprise can model this metadata with the MetaData Modeler and store, analyze, and version its models in the MetaBase metadata repository. How ever, the MetaMatrix software also can create virtual metadata that represents the “integrated information” for some or all of the enterprise.

This virtual metadata defines virtual databases that are transformations of one or more of the actual data sources. MetaMatrix achieves Enterprise Information Integration. Defining this virtual metadata offers substantial benefits:

- The virtual metadata captures in one integrated system the valuable semantic knowledge of the enterprise information, the definition of how the information is integrated. Additionally, the enterprise can export the metadata through the OMG’s XML Metadata Interchange standard to other tools, such as application development systems. For example, a business analyst might want to know how his enterprise’s business applications compute “Price/Earnings.” By asking the metadata management system this question, the analyst can determine how the computation is performed, as well as where the application obtains the data.

- The MetaMatrix Server can use the virtual and literal metadata stored in the MetaBase to provide real-time, uniform access to the enterprise’s integrated information. This integrated access to an entire enterprise’s information dramatically reduces application development time frames.

- The virtual metadata serves as a buffer between client applications and the underlying physical data stores. The enterprise can develop client applications in terms of the information that they require and minimize dependence upon the maintenance and modification of the physical data sources.
Within the MetaMatrix solution, the enterprise defines the structure of this virtual metadata in the same manner as metadata for physical data sources. Rather than reflecting specific data sources, the virtual metadata reflects the business-specific data structures used by the enterprise. The remainder of the virtual metadata includes the mappings and transformations that define how to create the desired information from the information in the enterprises data sources. The enterprise defines this using the MetaMatrix MetaData Modeler.

The enterprise can use and combine various transformations to define the virtual metadata. One example, a mathematical transformation, appears in Figure 4. The transformation creates a virtual column and computes it from two different data elements. So, retrieval of “Price/Earnings” involves retrieval of “Prc” from one data store and “Earn” from the same or a different data store and division of those values.

Another type of transformation is data element decomposition, which takes a single value and decomposes it into one or more data elements. In the example shown in Figure 5, the physical data source has an element that contains both the first and last names according to the form “LastName, FirstName.” The decomposition creates a new data element that contains the first name and another data element with the last name.

A more sophisticated example involves the business problem posed in the first section—the aggregation of information from different enterprise divisions. The enterprise can define a “union group” in the virtual metadata using transformations and can integrate semantically equivalent, though not identical, data from different sources, presenting that information as if from a single source. This is illustrated in Figure 6.

Decision transformations defer the choice of which data element to use until run time. As shown in Figure 7, the price returned will be the value(s) from Daily.Price only if the time is between 9AM and 4PM; otherwise, the value(s) from Close.Price will be returned.

Of course, other transformations are available. During information retrieval, these transformations are transparent to the client application. The MetaMatrix Server performs the transformations. The MetaMatrix Server and products were developed with this one purpose in mind—Enterprise Information Integration.
Conclusion

Enterprise Information Integration depends on sophisticated technology to collect, manage, and model metadata, metamodels, and meta-metamodels, as well as query management and connection technology for available information systems. MetaMatrix provides this dramatic new technology in a scalable, extensible, fault-tolerant architecture which is easy to set up and manage.

Competition is brutal. Time frames for responding developing and deploying applications grow shorter and shorter. Enterprises must use their data seamlessly and quickly. Businesses that fail to change and adapt, using an EII solution, cannot keep up in the hyper-competitive new economy.

Potential innovators have approached the challenge of managing and integrating disparate data resources in many ways in the past, mainly “one off” solutions where enterprises cobbled together data sources through specialized programming. These solutions worked in the infancy of information integration efforts, but proved hard, if not impossible, to scale in the global, networked economy. Information integration on this scale, the enterprise and the global scale, requires a new approach. The Enterprise Information Integration approach uses meta-metadata, or information about metadata, to achieve a new level of seamless and uniform access to disparate data sources. An approach that makes an enterprise’s disparate data sources appear as if they were one source.

MetaMatrix presents the first solution to the EII problem. The MetaMatrix product suite, with its flagship MetaMatrix Server™ and MetaBase™ repository, provide a platform through which an enterprise can manage and provide uniform access to disparate information systems for timely application development in a hyper-competitive inter-networked world.

This and other MetaMatrix White Papers are available for download at www.metamatrix.com/whitepapers.html