

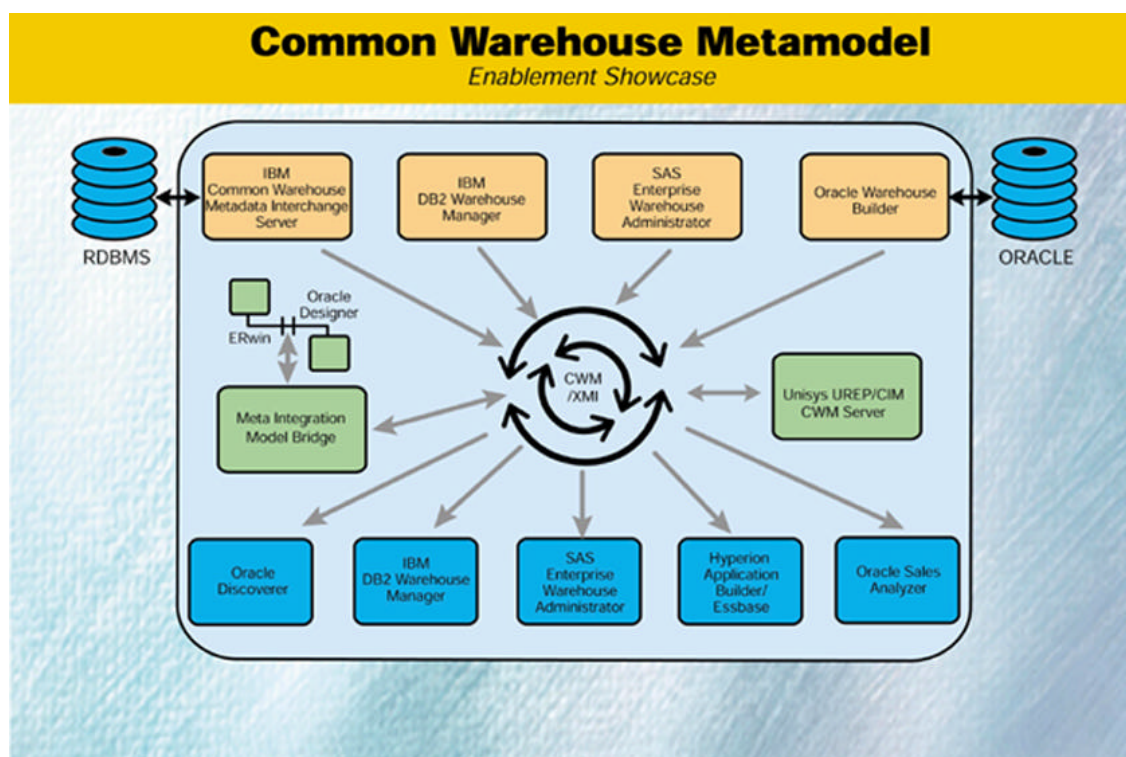
CWM Enablement Showcase: Warehouse Metadata Interchange Made Easy Using CWM

Dr. Daniel T. Chang (dтчang@us.ibm.com)
Member, IBM Database Technology Institute
Chair, OMG CWM Revision Task Force

When IBM, Oracle and Unisys started the CWM (Common Warehouse Metamodel) standardization effort within the OMG (Object Management Group) in May 1998, metadata interchange had been identified for sometime as one of the most critical problems facing the data warehousing industry. No common solution, however, was available at the time or on the horizon. MDIS (Meta Data Interchange Specification) had been published by the MDC (Meta Data Coalition) and implemented by some vendors. However, it was getting little support in the industry because of its limited coverage of metadata types and the lack of open interchange format. As such, warehouse metadata interchange was limited to using a single vendor's proprietary format, such as IBM's Tag Language, or using MDIS between a very limited number of vendors, such as IBM and ETI.

The OMG is an international standards organization with more than five hundred member companies. It is best known for its work on CORBA standards, which made many breakthrough contributions to distributed computing. Much less known, until recently, is its work on modeling and metadata standards that started around 1995. However, with its adoption and publication of UML (Unified Modeling Language) and MOF (Meta Object Facility) in 1997, XMI (XML Metadata Interchange) in 1999, and CWM last June (2000), it has become the premier standards organization on modeling and metadata. In a nutshell, UML provides a language for modeling metadata, MOF provides APIs for manipulating metadata, and XMI provides mechanisms for interchanging metadata in XML. CWM uses UML, MOF and XMI to model, manipulate and interchange warehouse metadata, including both technical and business metadata.

The OMG holds five technical meetings each year around the world. Last December 11th, it held its technical meeting in Orlando, right at the tail of the Florida presidential election controversy. Little known to the outside world and to the data warehousing industry at large was an event that took place on Tuesday and Wednesday of that week. The event was the CWM Enablement Showcase (see the attached figure), which was a major breakthrough for warehouse metadata interchange.



The CWM Enablement Showcase involved six vendors (IBM, Unisys, Hyperion, Oracle, SAS, and Meta Integration) and many different types of warehouse tools including: relational and multidimensional data stores (DB2 UDB, Essbase, and Express), warehouse builder tools (DB2 Warehouse Manager, Oracle Warehouse Builder, and SAS Enterprise Warehouse Administrator), OLAP and end-user tools (Hyperion Application Builder, Oracle Discoverer, and Oracle Sales Analyzer), and metadata repository and tools (UREP/CIM and Meta Integration Model Bridge). The Showcase was structured, and shown in the figure, along two major dimensions. Horizontally, tools were arranged according to whether they played the role of metadata producers (top), metadata consumers (bottom), or both (middle). Vertically, tools were arranged depending on whether they manipulated relational metadata (left), OLAP metadata (right), or warehousing metadata (middle).

The Showcase was demonstrated along three major threads of warehouse metadata interchange, all using the same sample (retail sales) scenario developed by the team. The scenario consists of three dimensions (product, store, and time) and four sales measures (quantity, gross revenue, gross cost, and gross profit), which can be represented either as a relational star schema or as an OLAP schema. The first thread demonstrated the interchange of relational metadata between DB2 UDB (producer, using the CWM Interchange Server), Oracle Discoverer (consumer), and Meta Integration Model Bridge (both consumer and producer). The second thread demonstrated the interchange of OLAP metadata between Oracle Warehouse Builder (producer) and Hyperion Application Builder or Oracle Sales Analyzer (consumer). The third thread demonstrated the interchange of warehouse metadata between DB2 Warehouse Manager (producer) and SAS Enterprise Warehouse Administrator (consumer). In all cases, UREP/CIM can be used to store and manage the metadata being interchanged.

In May 1998 it was difficult to imagine that multiple vendors could interchange even relational metadata (catalog, schema, table, column, etc.), which are well known and relatively straightforward, due to the lack of widely supported warehouse metadata standards. Two and a half years later, in December 2000, with the adoption by the OMG of CWM as the warehouse metadata standard, it became almost trivial for multiple vendors to interchange not only relational, but also OLAP and warehouse process metadata, as demonstrated by the CWM Enablement Showcase. The warehouse process metadata being interchanged there was pretty complex, representing the metadata involved in a customizable replication process from a source relational database to a target relational database. As such, the types of metadata covered included relational, transformation, warehouse process, software deployment, and business information.

The key to the success of CWM is UML and XML. UML provides a single and powerful modeling language for modeling all types of warehouse metadata; XML provides a simple and universal data format for interchanging all kinds of warehouse metadata. By nature, data warehousing deals with a very complex environment that involves many different types of data sources and targets (relational, record-based, object-oriented, multidimensional, and XML), various types of transformation and analysis tools (OLAP, data mining, information visualization, and business nomenclature), as well as warehouse process and operation management. UML is capable of modeling the metadata for all of the above entities, thus serving as the single modeling language and enabling a model driven approach for metadata interchange. Also by nature, data warehousing involves many different tools from many different vendors, big and small. Therefore, for metadata interchange to be successful and prevalent, the interchange format used must be universal and must be cheap to implement and support. XML is ideal for this purpose, being simple and universal.

The CWM Enablement Showcase was a resounding success and a major breakthrough for warehouse metadata interchange. Nevertheless, this is only the beginning. With the decision by the MDC to fold its metadata standardization effort into the OMG and the fact that JCP (Java Community Process) is extending OMG metadata standards into the Java platform (e.g., JMI, JOLAP and JDMAPI), acceptance and support for CWM in the industry is rapidly growing. One should not be surprised to find CWM providing the common solution to warehouse metadata interchange problems in the very near future. Looking a little further into the future, as W3C adopts newer standards on XML (e.g., XML Schema and XML Query Language) and as CWM evolves to incorporate these technologies, warehouse metadata interchange using CWM should become even easier and more powerful. For more information on CWM, please access <http://www.cwmforum.org>. IBM's DB2 Warehouse Manager supports CWM for metadata interchange today. DB2 Warehouse Manager is a tool for defining and managing data warehouses and the ETL steps used to populate and maintain the data warehouse. More information on DB2 Warehouse Manager can be found at <http://www-4.ibm.com/software/data/db2/warehouse>.