EDMIS: Metadata Interchange System for OLAP

In-Gi Lee, Minsoo Lee, Hwan-Seung Yong

lig@ktf.com, {mlee, hsyong}@ewha.ac.kr
Dept. of Computer Science and Engineering
Ewha Womans University
11-1 Daehyun-Dong, Seodaemoon-Ku
Seoul, 120-750 Korea

Abstract. The management of enterprise-wide information requires that metadata should be shared and globally accessible by all the heterogeneous products found in today's information technology environment. OLAP metadata includes various information about objects and the relationships among them in an OLAP system. OLAP tools are provided by different vendors and run on different hardware and software platforms. But there still is no standard for a logical OLAP model. To use OLAP tools efficiently, users need to be able to move metadata between tools or between tools and a repository. In this paper, we have designed a metadata interchange model that can be shared among OLAP systems and have implemented a prototype called EDMIS as the OLAP metadata interchange system. The OLAP metadata interchange model that we propose in this paper is designed using XML DTDs. This model lists typical mismatches among the data models of commercial OLAP tools and proposes methods to overcome these differences. The OLAP metadata interchange system enables users to browse and search OLAP metadata on the Web. In addition, if a user wants to use the metadata in another OLAP system, the implemented system can automatically create new cubes using metadata in the repository. In order to validate the usefulness of the proposed system, we have used OLAP products such as MS SQL Server Analysis Services, Pilot DSS and Oracle Express.

1. Introduction

As companies are trying to find a way to more easily administrate their databases and enable a rapid decision making process, the concept of data warehousing which enables end users to access data that is distributed within the company is rapidly gaining popularity. The core part of data warehousing is OLAP (On-Line Analytical Processing) technology, which allows the end user to directly access and analyze the multi-dimensional information in a user-friendly manner and make decisions based on the analysis. This technology, in addition to data mining, is one of the most popular uses of the data warehouse [1].

Despite the hype in data warehousing and OLAP, users are still having problems in obtaining consistent and well-organized information regarding OLAP. Currently a large number of OLAP products are available on the market, but in reality there still does not exist a common standard for a logical multi-dimensional model. The absence of a standard among the OLAP products has made it very difficult to exchange data and interface with the products. Even though a user has knowledge of a specific OLAP product, it is very difficult to communicate with other users if they have used different products. To solve this problem, standards efforts are being carried out at the API level. As an example, MD-API of the OLAP Council and OLE DB for OLAP API from Microsoft have been proposed [2]. In addition to this, many vendors are working on OLAP standards but it is very difficult due to the differences in the various OLAP products that have been developed.

This paper explains a common metadata model that can be used among different OLAP products, and shows the design of an XML-based OLAP metadata interchange model based on the metadata model. The proposed OLAP metadata interchange model is an integration of the solutions obtained from the experience of extracting metadata from several OLAP products and creating new cubes. We focus on OLAP metadata required for cube creation and management. We have implemented an OLAP metadata interchange system called EDMIS (Ewha Data warehouse Metadata Interchange System). The OLAP metadata interchange system is used for exchanging cube data that includes dimension, hierarchy and variables etc. It extracts metadata from OLAP servers, stores it, and if needed will automatically create cubes in other OLAP products. This system allows users to interface with different products without going through the complex cube creation process for each product, and enable users to select the OLAP product that can satisfy their purpose and required query performance. It is also developed as a Web-based system to allow users to conveniently manage the metadata through the Web.

The organization of the paper is as follows. Section 2 discusses several standardization efforts. In section 3, an OLAP metadata model is explained and the design of the OLAP metadata interchange model is shown. Section 4 explains the implementation of the OLAP metadata interchange system and gives an example on how the metadata from the OLAP products are extracted and can be used to recreate cubes in other products. Section 5 concludes the paper with a summary of the results and future work.

2. Related Research

It has been very difficult to exchange data among OLAP products due to the absence of a commonly agreed standard for a logical multi-dimensional model. Currently about 40 OLAP products are using different OLAP metadata, vendor specific technology, and specialized user interfaces. Therefore users who have only limited experience with a single OLAP product obtain biased knowledge about OLAP technology. In order to solve this problem, standardization efforts at the API level have been recently carried out. MD-API of the OLAP Council and Microsoft OLE DB FOR OLAP API are such efforts.

MD-API is an object-oriented database-independent interface to a multi-dimensional data source. This API enables clients to select a multi-dimensional data schema, connect to it, and query the metadata. A set of objects is used to represent a query rather than using a text based language to form a query. Therefore, there is no specific textual language used by the API. All query results are returned as a cube object, and applications can obtain cell data from the cube. OLE DB for OLAP API was designed by extending OLE DB so that users can access multi-dimensional data sources. The API is a collection of COM objects and interfaces to enable efficient communication between producers and consumers of multi-dimensional data. OLAP products can easily communicate independently of the multi-dimensional data environment via the API. A multi-dimensional query language called MDX (Multidimensional Expressions) is also provided to connect the OLAP clients and servers. This query language is similar to SQL and has a large variety of functionalities.

Although the above two standard APIs have been proposed, MD-API is only implemented by a small number of vendors such as Gentia and is currently not being supported by a large number of OLAP product vendors, which is resulting in poor influence on the community. On the other hand, OLE DB for OLAP API is receiving support from almost all of the OLAP vendors, but does not define a detail OLAP model and cannot be considered as a real standard.

We propose an XML-based approach for modeling the OLAP metadata. Other approaches that have recently gained interest regarding modeling of metadata include ontology-based approaches such as DAML, OIL, and topic maps [3]. The use of ontologies provides a very powerful way to describe objects and their relationships to other objects.

3. Design of the Metadata Interchange System

Research on metadata standards includes two areas. One is the metadata model which is used to apply the metadata on the repositories. The other is the metadata exchange model which is used to exchange metadata among different tools.

A metadata model defines the structures and semantics of the metadata in order for vendors to share the metadata in a common way. Figure 1 shows an example of how the metadata can be shared and used by various users and applications. There are several international standards for metadata. The most interesting ones are the Open Information Model (OIM) proposed by the MetaData Coalition (MDC) and the Common Warehouse Metamodel (CWM) proposed by the Object Management Group (OMG). Recently, as the MDC has become part of OMG, these two efforts are now converging.

The implemented OLAP metadata exchange system is based on an extensible OLAP metadata exchange model that can be used as a standard among different OLAP products that have different hardware and platform requirements.

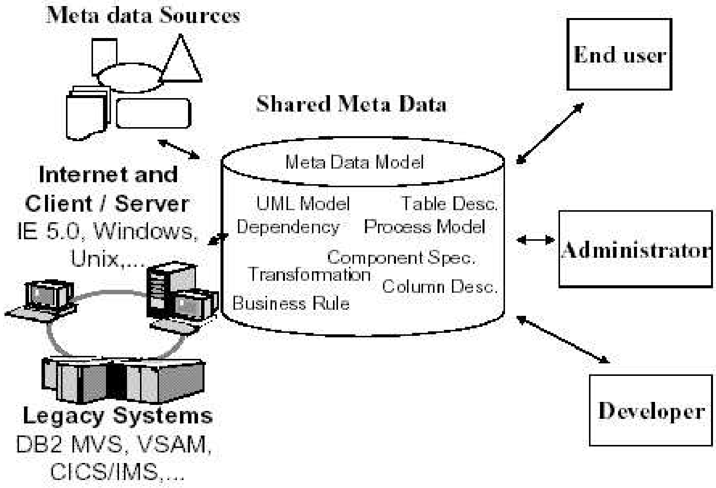

Fig. 1. A shared metadata model

3.1. OLAP metadata model

One of the current standards for the metadata models is the Open Information Model proposed by MDC. OIM is based on an object model that can store specific types of information and is also flexible enough to support new types of information. It is an open standard that can be extended to satisfy the requirements of specific users or vendors. OIM includes five areas: Analysis and Design, Component Description and Specification, Database and Data Warehousing, Business Engineering, Knowledge Management. We discuss the OLAP schema in the Data Warehousing area in the following.

The OLAP metadata model provides a common place to store the multi-dimensional schema information of the various OLAP products. The model is composed of cubes, dimensions, hierarchies, levels, dimension items, variables, and attributes as basic elements. The store is an abstract class that generalizes the different multi-dimensional storage objects. A store can represent a cube, virtual cube, physical cube, partition, and aggregation. The cube is a basic element of multi-dimensional analysis and is generally composed of a fact table and more than one dimension tables. Fact tables contain the measure variables that are related to a combination of the items in the dimension tables. The dimension tables define the item values for a dimension. As an example, there could be the three dimensions such as store, product, and time that each contain specific detail information regarding all of the stores for a company, the products that are sold for the company, and the time information of the fiscal years. The fact table will then record the transactions when a product is sold from a store at a specific time. The fact table would record the number of products sold and the margin from the transaction. Dimension hierarchies are associated with levels and mappings. Dimension levels are used to roll-up or breakdown on details. As an example, a geography dimension can have states, regions, and countries as levels. Mappings are used to indicate the use of a dimension hierarchy by a store.

The MD-API uses its own OLAP data model and has been unsuccessful in receiving support from the vendors. The OLE DB FOR OLAP API follows a standard metadata model. Because the different OLAP products use different OLAP metadata models, it is very important to define a standard OLAP metadata model. This paper proposes a flexible OLAP metadata interchange model that enables exchanging of metadata among vendors and also could be extended as a metadata model.

3.2. An XML-based OLAP metadata interchange model

The OLAP metadata exchange model is described in detail in terms of the DTDs of the XML in the following subsection. Afterwards, the application method of this model on several commercial products is investigated.

3.2.1. Model design

The OLAP metadata interchange model is defined based on the metadata model explained in the previous section. There are several reasons for using XML in the interchange model. XML is vendor independent, programming language independent, and its structure can be defined via DTDs and XSDs. The OLAP metadata interchange model was designed by observing and factoring out the common characteristics of several products and enabling the model to handle the differences among those products. The model includes common metadata that is shared by several products (i.e., the least common denominator) and special metadata that is dependent on specific products. The special metadata can be stored in the metadata store and can be used when necessary. As an example, Attributes are used to show the query results or used as characteristics to select the set of members. However, not all OLAP products can support this Attribute. Even though some products support it, each product supports it in its own way. In this case, all of the supported Attributes were organized together to prevent the loss of data. The differences are supported by approaches discussed in the next section on how the model is applied to each product. The OLAP metadata interchange model can include and interchange the characteristics of several products in this manner.

The interchange model is composed of the basic elements such as DATABASE, CUBE, DIMENSION, MEASURE, HIERARCHY, LEVEL, MEMBER, and ATTRIBUTE.

The DATABASE is the highest level structure which is composed of PROPERTY, DATASOURCE, and CUBE. The CUBE is composed of PROPERTY, MEASURE and DIMENSION, and an optional element called STORE has the information about the storage structure. The PROPERTY element contained in each element is very similar but is organized in different ways according to the circumstances. For the CUBE, the PROPERTY includes the cube type which shows if it is a virtual cube or not, the description of the cube, and other optional items.

The DTD of the MEASURE shown in Figure 2 includes the values and the types of the real data used for analysis. The data type is usually a numeric type and has attributes such as numeric_precision, numeric_units, and numeric_scale. The properties of MEASURE include the measure_aggregator which is the basic operation used for aggregation, the caption, and unique_name. Each FIELD of MEASURE includes the dimensions that are used for the analysis with consideration of the series multi-cube approach.

```
<!ELEMENT MEASURE(PROPERTY,FIELD+,DATATYPE)*>
<!ATTLIST MEASURE
      measure_name      CDATA #REQUIRED
      created_on        CDATA #IMPLIED
      last_schema_update CDATA #IMPLIED
      schema_updated_by  CDATA #IMPLIED
>
<!ELEMENT PROPERTY(database_name+, cube_name, measure_unique_name?,
measure_caption?, measure_guid?, measure_desc*, measure_aggregator )>
```

Fig. 2. MEASURE element of the OLAP metadata interchange model

```
<!ELEMENT DIMENSION(PROPERTY, HIERARCHY*,DATATYPE?,ATTRIBUTE*)>
<!ATTLIST DIMENSION
      dimension_name    CDATA #REQUIRED
      created_on        CDATA #IMPLIED
      last_schema_update CDATA #IMPLIED
      schema_updated_by  CDATA #IMPLIED
>
<!ELEMENT PROPERTY(database_name+, cube_name*, dimension_unique_name?,
 dimension_desc*, dimension_ordinal, dimension_type, dimension_cardinality,
default_hierarchy, is_virtual, is_drillthrough_enabled?,
dimension_unique_settings?, is_SQL_enabled? )>
```

Fig. 3. DIMENSION element of the OLAP metadata interchange model

Figure 3 shows that a DIMENSION is composed of HIERARCHY, DATATYPE, and ATTRIBUTE elements. Although there usually is just a single hierarchy in a dimension, some products support several hierarchies. Many products do not support DATATYPE. Some products support ATTRIBUTE at the dimension while others support it at the level or member. Looking at PROPERTY, the cube name is a required element. If the metadata is extracted from a product supporting multi-cubes instead of hyper-cubes, there can be multiple cube names. Dimension_ordinal is a number that represents the order among the dimension sets that form the cube. Dimension_cardinality is the number of members in the dimension. Dimension_type indicates that the dimension is either a time dimension, quantitative dimension, regular dimension, or other dimension.

Figure 4 shows that the HIERARCHY is composed of LEVEL, ATTRIBUTE, DATATYPE elements and has a similar structure as the dimension.

The XML DTDs shown here only represent the basic structure of the OLAP metadata interchange model. The next subsection explains how to actually apply this model to several products.

```
<!ELEMENT HIERARCHY(PROPERTY, LEVEL*, DATATYPE?, ATTRIBUTE*)>
<!ATTLIST HIERARCHY
      hierarchy_name     CDATA #REQUIRED
      created_on        CDATA #IMPLIED
      last_schema_update CDATA #IMPLIED
      schema_updated_by  CDATA #IMPLIED
>
```

Fig. 4. HIERARCHY element of the OLAP metadata interchange model

3.2.2. Model application methodologies

The OLAP metadata interchange system was designed to handle the special characteristics of the products by extracting the metadata from each of the products. The model needs to be applied in different ways based on how the product organizes its metadata. These methods are incorporated as a module of the OLAP metadata interchange system. When the OLAP metadata interchange system creates a cube for specific products, transformations are necessary to map it to the model of the product.

The following explains the transformations that are required to overcome the differences in the products shown in Table 1. Cubes, Attributes, and Dimensions are mapped accordingly.

Table 1. Analysis of OLAP product models

Component / Products	**Storage**	**Cube**	**Attribute**	**Dimension**
Cognos Powerplay	MOLAP	Hypercube	Not Supported	Multi level (Special category)
Hyperion Essbase	MOLAP	Hypercube	Dimension attribute	Single Hierarchy Multi level
Informix Metacube	ROLAP	Block Multicube	Supported	Single Hierarchy Multi level
MS OLAP Services	HOLAP	Block Multicube	Level attribute	Single Hierarchy Multi level
Oracle Express	MOLAP (HOLAP)	Series Multicube	Not Supported	Multi Hierarchy Single level

4. System Implementation

An OLAP metadata interchange system was developed to exchange cube data among different OLAP products. The extracted metadata from the OLAP products is transformed according to the OLAP metadata interchange model and stored. Users can search and edit the metadata in a Web environment. Additionally, users can automatically create cubes in the desired OLAP products without needing to go through the complex process relevant to each product. The user interface is a Web-based interface so that users can easily interchange data among the OLAP products anytime anywhere.

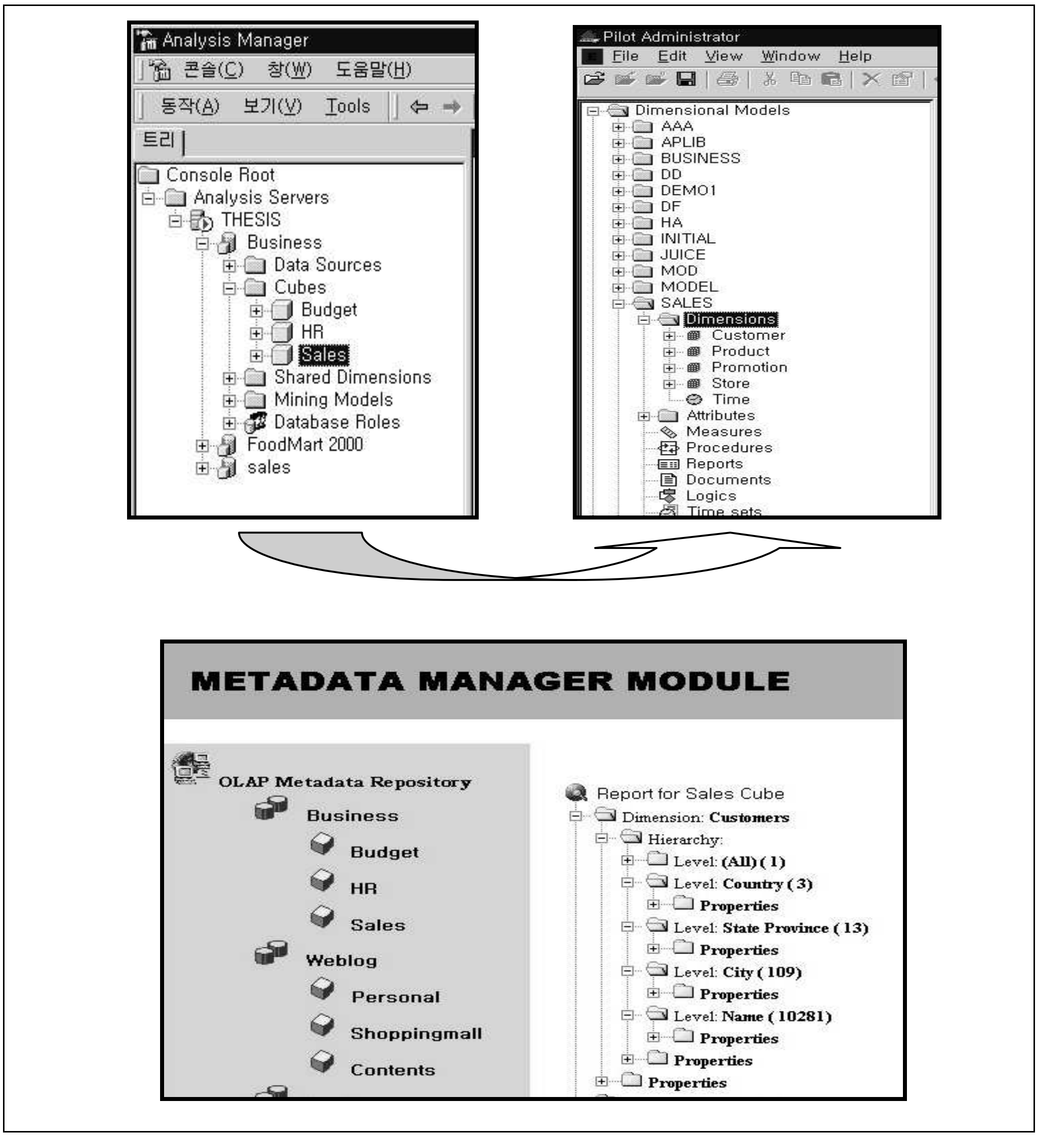

Fig. 5. Cube creation in Pilot

The implementation environment is as follows. The OLAP metadata interchange system uses Object Design eXcelon as the XML data server running on Windows 2000. Excelon can efficiently store and manage the dynamic structure of the XML data. OLE DB FOR OLAP API was used for the metadata extraction, and C++ was used as the programming language. The MS Internet Information Server was used as the Web server and MS Internet Explorer was used as the client. ASP, XQL, and XSL was used for searching and editing the metadata in a Web environment.

The OLAP metadata interchange system is composed of four modules: the metadata extraction module, the metadata transformation and storage module, the metadata manager module, and the cube creation module. The metadata extraction module will extract the metadata that is desired by the user from various OLAP products. The metadata transformation and storage module applies the OLAP metadata interchange model to the extracted metadata information and transforms it into XML and stores it into the eXcelon Server. The metadata manager module enables users to search and edit the metadata in the eXcelon Server through a Web-based interface. The cube creation module enables users to automatically create cubes based on the metadata in a specific OLAP product.

The system has been used with Microsoft OLAP Services, Pilot's Decision Support Suite, and Oracle's Express for evaluation. Figure 5 shows the metadata being exported from OLAP Services and the cube being created in Pilot. The extension of the system to support other products is easy and straightforward.

5. Conclusion

On-Line Analytical Processing(OLAP) systems is one of the key technologies for data warehouses. However, the approximately 40 OLAP products were developed with different OLAP metadata, making it difficult for OLAP products to exchange data or interface with each other. This paper proposes an OLAP metadata interchange model along with an implemented OLAP metadata interchange system.

The following points differentiate the proposed OLAP metadata interchange system with other research results. First, by analyzing many of the existing OLAP products, several problems that occur during the standardization process in the OLAP area were identified and incorporated into the model. Second, the design of the OLAP metadata interchange model can incorporate the models of the currently used products and supports the metadata model transformation process. Third, the metadata interchange model uses XML, making it easier to transmit structured documents over the Web. Fourth, users need not learn about the complex cube creation process of each product. With just a single cube, the user can experience the different analysis environments provided by several different products.

Currently, the most debated area of OLAP technology is the standardization effort. OLAP products will be able to talk to each other in the near future by extending the results of this paper to design a common query language as well as common metadata and interfaces.

References

[1] E. Thomsen, OLAP Solution, John Wiley &Sons, 1997

[2] Cho Jae Hee, Park Sung Jin, OLAP Technology, Sigma Consulting Group, 1999

[3] DAML.org, DARPA Agent Markup Language (DAML), http://www.daml.org/

[4] "Putting Metadata to Work in the Warehouse",http://www.cai.com/products/platinum/wp/wp_meta.htm

[5] Pilot Software, *White Paper*, "An Introduction to OLAP: Multidimensional Terminology and Technology", http://www.pilotsw.com/olap/olap.htm

[6] Oracle Express Server, http://www.oracle.com/ip/analyze/warehouse/servers/index.html

[7] Microsoft, OLE DB for OLAP 2.0 Beta Specification,http://www.microsoft.com/data/oledb/olap/spec/

[8] OLAP Council, MDAPI specification version 2.0,http://www.olapcouncil.org/