8-16-1996

A Multidimensional Data Model For Data Warehousing Applications

Anil Kini
Department of Business Analysis and Research,Texas A&M University, a-kini@tamu.edu

A Multidimensional Data Model For Data Warehousing Applications

Anil Kini
Department of Business Analysis and Research
Texas A&M University
College Station, TX 77843
: (409) 845-6998
E-: a-kini@tamu.edu
http://acs.tamu.edu/~adk6816

Decision Support Systems and Data Warehousing

Decision Support Systems (DSS) are computer based systems that are used to assist human decision makers in solving semi-structured and unstructured problems. (Sprague, 1980). However, the major problem with developing effective decision support systems is the availability of data which can be represented in a form that is easily understood by the decision maker (Inmon, 1992), (Ricciuti, 1994). Relational databases are based on the relational model (Codd, 1970) and are the *de facto* standard for storing and retrieving data and are widely used in most organizations. However, relational databases are optimized for transaction processing and are extremely efficient in storing and updating the data. Most applications of relational databases have aimed at facilitating or meeting the requirements for transaction processing, operational control, or management control. Therefore the major concerns of these systems have been at the lower levels of the data, dealing primarily with raw data.

The relational model is quite inadequate for analyzing data,. Most business managers need to analyze information on-line and in a dynamic fashion (Frank, 1994), (Hodges, 1995). In a recent paper, Codd (1993) states that the relational database systems were never intended to provide powerful functions for data synthesis, analysis, and consolidation.

Data warehousing is a concept that is gaining increasing popularity because of the need for quality data. Data warehousing is the creation of a specialized database to help support decision making (Inmon, 1994). Because of the problems associated with relational databases, the solution now is to create a specialized database which extracts data from the conventional databases and stores them in a single large server or database. This data store or warehouse typically consists of enterprise data from diverse production systems and is an approximation of the entire enterprise data.

The data warehouse is usually updated by batch processing on a periodic basis, and is therefore a more static or historical model of the enterprise data. Once the data warehouse is populated, the data is essentially static and does not change till the next time the warehouse is updated. The data warehouse is created to satisfy the needs of decision makers and has a different structure and representation that is more intuitive and responsive to managerial queries.

Data Warehousing and Multidimensional Databases

The data warehouse is usually developed as a central data store which provides data in the format that is understandable by the users. The data warehouse provides a traditional, highly manageable data center for DSS, ensuring that the data is readily available and quickly accessible by the users [Demarest, 1994]. The data warehouse is not just a copy of the data in other systems. It is a unique, enriched data set that is optimized for decision support. One approach to provide this unique and enriched view of the data is to use multidimensional databases.

Multidimensional databases (MDDB) are specialized databases designed to facilitate multidimensional data analysis and decision support (Finkelstein, 1995). MDDBs store numeric or quantitative data which is categorized over several qualitative dimensions. For example, a MDDB may store sales data (quantitative) for several product lines, in several different cities, for each month (3 qualitative dimensions). The user will then be able to retrieve data on any specific combination of these dimensions as well as perform aggregations and consolidations. For example, a manager might be interested in obtaining the sales figures for Product X in Houston for Dec. 1994 or obtain the Total Sales for Product X in Houston for 1994 (Aggregation).

A MDDB has a database management system (MDDBMS) which provides the user with the capability to analyze this data by providing tools for flexible, ad hoc data analysis. This end user oriented data analysis system provides users with the capability for sophisticated data analysis without requiring programming language knowledge or support form the Information Systems (IS) personnel. This system insulates the user from having to master the intricacies of data storage and access mechanisms.

Architecture of the Multidimensional Database System

In this section we describe the architecture of the multidimensional database system with its functional elements and the interactions between these elements. Figure 1 illustrates the architecture of the system and outlines the information flow as well as the control flow between the components of the system. We will discuss the multidimensional database management system (MDDBMS) component in this paper. The MDDBMS is the component that provides the facility to build, manage and query the multidimensional database. The MDDBMS consists of three sub-components, the Structure Definition Processor (SDP), the Mapping Processor (MP), and the Query Processor (QP).

The SDP is used to define the multidimensional data structure and to define the dimensions and the interactions among the dimensions. The SDP will be implemented based on the proposed data model. This component will be used by the database designer to create the database schema. The SDP populates the multidimensional data dictionary once the entire database schema has been defined.

The MP serves as the processing interface between the relational and the multidimensional components. Once the data structure of the multidimensional database is specified, the MP extracts data from the relational database and converts the data into the multidimensional format. This data is then stored in the multidimensional database and is now accessible by the query processor.

The QP is the querying facility of the database and allows the user to view the different dimensional views and to query the database. The QP will also use the graphical data model to represent the multidimensional data.

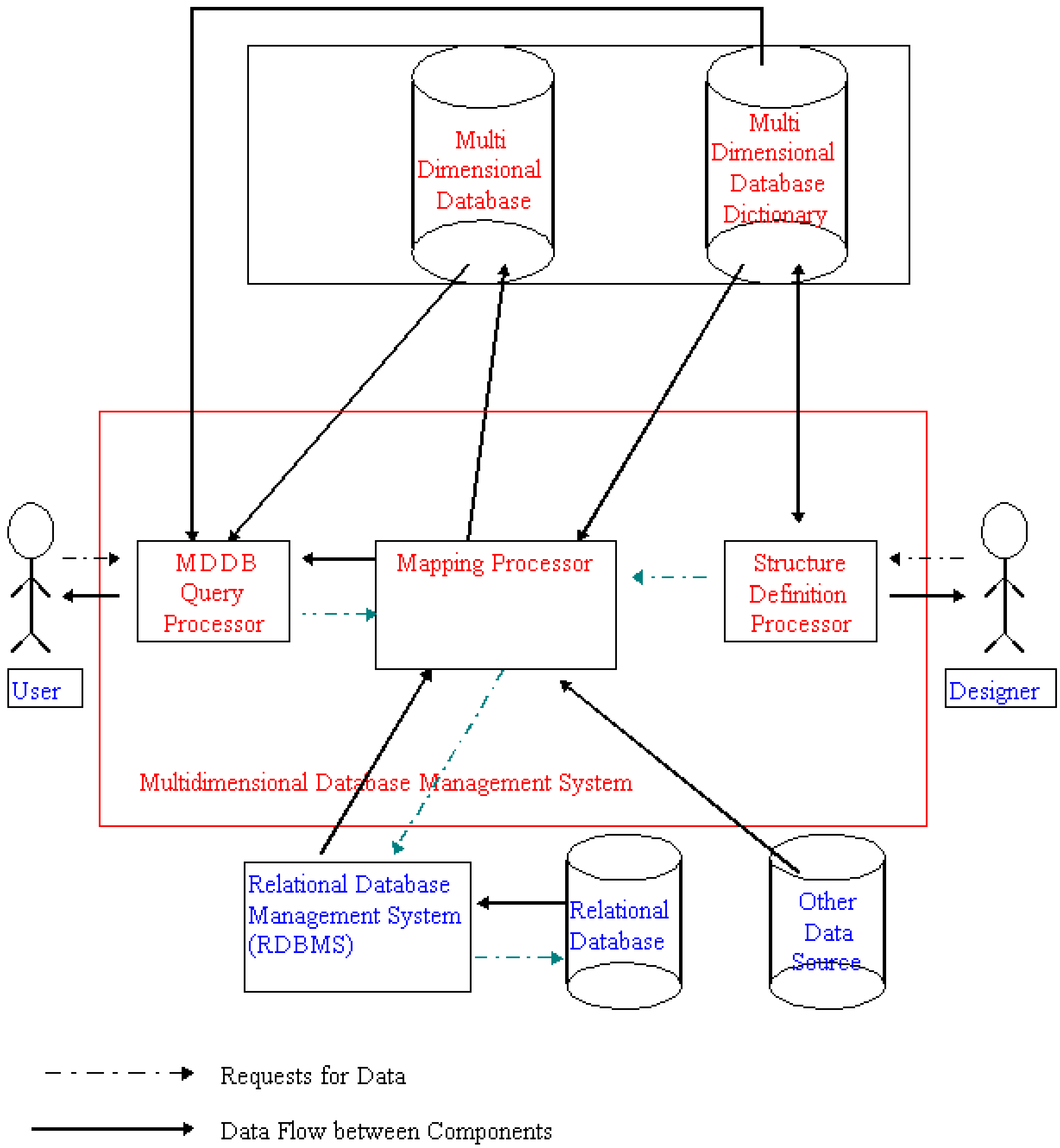

Fig. 1: Architecture of the multidimensional database system**A Data Model for Multidimensional Data**

In this section, we describe the proposed data model, and illustrate the structures proposed for multidimensional data. This model is based on the semantic data model (Su, 1983) and is modified to represent multidimensional databases. We will briefly describe the structures and illustrate the data model. For illustration purposes, consider an example database which contains Sales data of various products for company X. The company needs to store data about the sales for different products over several periods of time for its various regions. This simplified database will be used to illustrate the data model. Four of the structures in the data model are described below:

Summary Variable (SV): This is the variable of interest in the MDDB. It is the quantitative variable for which values are stored in the database. The summary variable is a numeric measure that is of interest to the analyst as it has a major impact on the analysis. In the previous example, the summary variable is the **Sales** (in units) of the products.

Dimensional Attribute (DA): The base attributes or categories that characterize the summary variable are called dimensional attributes. For example, the summary variable Sales can be dimensioned by three attributes **Product**, **Region**, and **Period**.

Dimensional Instances (DI): These are the instances of the dimensional attributes. For example, the dimensional attribute Product can have 4 instances, <Processor, Memory, Modems, Hard Drive>.

Dimension Hierarchies (DH): These are hierarchies that describe the dimensional variables in further detail. Most dimensions themselves can have a complex structure that is usually hierarchical. For example, the dimension Region can be broken down into a State dimension, and further into a City dimension. If a dimension is broken down into hierarchies, then the dimension is no longer a dimensional variable. The dimension variable is always the lowest node in its hierarchical structure.

In this paper, we have illustrated only a few of the structures necessary for multidimensional data. A more detailed description of this data model is available in Kini (1996). Fig. 2 illustrates the graphical view of the multidimensional data model for company X as defined previously.

Operations On Multidimensional Data

Multidimensional databases provide several operations to facilitate data analysis. We will briefly

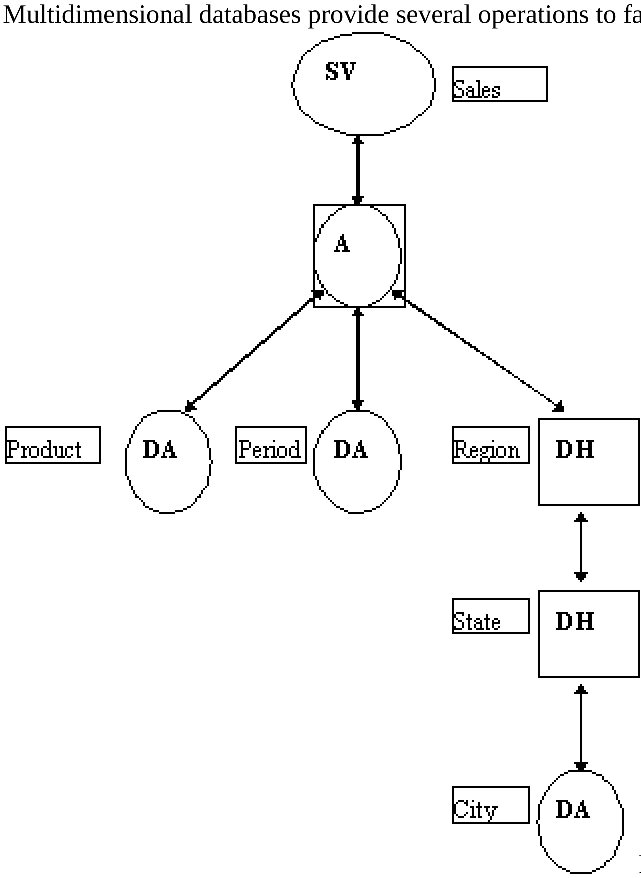

Fig 2: A graphical view of the multidimensional data.

describe three operators for multidimensional data.

Consolidation (Roll Up) refers to the process of moving up a dimension hierarchy to obtain more aggregated views of the data. This process is also called roll up, as it refers to the movement from lower to higher levels in the hierarchy. For example, a simple roll up involves consolidating Cities into States, and States into Regions.

Drill-Down: This operation refers to the process of drilling down to obtain detail data. Drill down is useful when analyzing a cause or effect for some observed phenomena in the aggregate data

Slice and Dice: This is the ability to look at the database from different viewpoints. For example, the analyst may wish to view Sales data dimensioned by Product and Region only with Period values being consolidated.

Conclusion

In this paper, we illustrated the architecture of a multidimensional database system and described a data model to represent multidimensional data. The concept of data warehousing is gaining popularity as the means of providing efficient decision support. Multidimensional databases are ideally suited for use as the underlying database for developing data warehouses for decision support. This proposed model will be used to represent and develop multidimensional databases, and will provide an abstraction tool for designing these databases.

"References available upon request from author"