International Journal of Advanced Technology in Engineering and Science
Vol. No.4, Issue No. 08, August 2016
www.ijates.com
ijates
ISSN 2348 - 7550

MULTIDIMENSIONAL MODELING COMPARISON ON CONCEPTUAL VIEW

Sushil Kumar[1], Amit Sharma[2]

Asst Professor,IIMT College of Engg,Gr,Noida

ABSTRACT

Many OLAP usages indicate that their usability performance degrades due to wrong interpretation of business dimensions. In this paper, we are focusing about business dimensions by multidimensional data model structures for the DWs. Multidimensionality is just a design technique that separates the information into facts and dimensions by understanding the business processes and the required dimensions . Many approaches have been suggested but we will focus on widely accepted star Schema with slight improvement using Snowflake Schema a variation of star schema, in which the dimensional tables from a star schema are organized into a hierarchy by normalizing them. Multidimensional model present information to the end-user in a way that corresponds to his normal understanding of his business dimensions, key figures or facts from the different scenarios that influence user's requirement .

Keywords: Business Dimensions, Multidimensional Modeling, Data Warehouse, OLAP, Star Schema, Snowflake Schema, Fact.

I. INTRODUCTION

[1]DW generalize and consolidate the data in the multidimensional space. The construction of DW involves data warehouses involves data cleaning, data integration, and data transformation and can be viewed as an important preprocessing step for data mining. Moreover, data warehouses provide on-line analytical processing (OLAP) tools for the interactive analysis of multidimensional data of varied granularities, which facilitates effective data generalization and data mining. A data warehouse is a set of data and technologies aimed at enabling the executives, managers and analysts to make better and faster decisions. DWs to manage information efficiently as the main organizational asset. The principal role of DW in taking strategic decisions, quality is fundamental. Data warehouse systems are important tools in today's competitive, fast-changing era. In the last several years, many firms have spent millions of dollars in building enterprise-wide data The DWs have to inherent support for complex queries however its maintenance does not suppose transactional load. These features cause the design techniques and the used strategies to be different from the traditional ones. Many people feel that with competition mounting in every industry and domain, data warehousing is the latest must-have marketing weapon and panacea —a way to retain customers by learning more about their requirements

Enterprise DW[3] -An enterprise DW provides a centralized database architecture for decision support for the enterprise.

Operational Data Store-It has a broader enterprise wide frame, but unlike the real one. Enterprises DW, data is refreshed in near real time and used for routine business processes.

Data Mart -Data mart is a subset of data warehouse and it supports a specific domain, business process.

1.1 Characteristics of DW[8]

The main characteristics of data warehouse are:

Subject oriented. DW is organized around major subjects, such as, supplier, product, customer and sales.

Separate, DW is always a physically distinct store of data transformed from the application data found from the traditional OLTP environment. Due to this separation, a data warehouse does not require transaction processing, recovery, and concurrency control mechanisms. It usually combines two operations in data accessing: initial loading of data and access of data.

Time Variant. Problems have to be addressed; trends and correlation's have to be explored. They are time stamped and associated with defined periods of time[2].

Not dynamic. When the data is updated, it is done only periodical, but not as on individual basis.

Integrated Performance. The data which is requested by the user has to perform well on all scales of integration. Data cleaning and data integration techniques are applied to ensure consistency in naming conventions, encoding structures, attribute measures, and so on.

Consistency. Architectural and contents of the data is very significant and can only be ensured by the use of metadata: this is independent from the source and collection date of the data.

1.2-data warehouse building process[7]

To construct an effective data warehouse we have to analyze business processes, dimension and business environment. After obtaining the DW logical schema, build it through apptlication of transformations to the source logical schema, and apply the construction of a large and complex information system, can be viewed as the construction of a large and complex building, for which the owner, architect, and builder have different views. These perspectives are merged to form a complex framework that represents the top-down, business-driven, or owner's perspective, as well as the bottom-up, builder-driven, or implementer's view of the information system. The multidimensional model transforms the visualization of a schema into a more business-focused framework. All these structures cubes, measures and dimensions interact with each other to provide an extremely powerful reporting environment.

Most of the multidimensional database systems used in business framework and decision support applications is particular. Generally, they can be categorized into two categories: 1st is the special traditional relational DBMS which create multi-dimensional schemas such as star schema and snowflake schema by applying the mature theory of relational database systems, the 2nd is the multi-dimensional database systems which are designed specially for online analysis. All dimensional tables are directly connected with the fact table and do not generate connections with other dimensional tables. However, it will need to separate one dimension into many dimensions as per business dimension mappings. Such structure is called the snowflake mode, a slight modification of star adding relational constraints of normalization. Relational database systems are suitable for OLTP4 applications, but it does not guarantee to meet the expectations of online analytical processing applications in real time environment. Relational OLAP3 systems which are inherently ORDBMS can only be classified as relational database systems, because after changing into systems supporting OLTP applications, relational approach can only used, that disappeared the object features. A multidimensional database is a type of database (DB) which is optimized for DW and OLTP applications. Multidimensional databases are mostly generated using the given data from existing RDs, a multidimensional database allows a user to refer problem and questions related to concizing business operations and trends analysis. An OLTP application that processes data from a multidimensional database is formally referred as a multidimensional OLTP application. A multidimensional database or a multidimensional database management system implies the ability to rapidly accept the data in the database so that answers can be generated easily. A number of vendors provide products that use multidimensional databases. An approach to how data is stored and the user interface differs. To multidimensional database systems, applications are eased due to uniform specifications does not exist. They are special database systems which do not support comprehensive query, Four different views regarding the design of a data warehouse must be deemed: the top-down view, the data source view, the data warehouse view, and the business query view. The top-down view allows the selection of the relevant information vital for the DW. This information resembles the current and future business requirements. The data source view shows or reflect the information being captured, stored, and managed by operational systems. This information may be documented at various Hierarchies of detail and accuracy, from individual data source tables to integrated data source tables. Data sources are often modeled by traditional data modeling approach, such as the entity-relationship model or CASE (computer-aided software engineering) tools. The data warehouse view combines fact tables and dimension tables. It represents the information that is stored inside the DW, including predetermined aggregates and counts, as well as information pertaining to the source, date, and time of origin, added to provide historical scenario. Finally, the business query view is the perspective of data in the data warehouse from the viewpoint of the end user.

II. MULTI DIMENSIONAL MODELING

It is a technique for formalizing and visualizing data models as a set of measures that are defined by common aspects of the business processes. Business Dimensional modeling has two basic concepts.

Facts:

- A fact is a collection of related data items, composed of Business measures.
- A fact is a focus of interest for the decision making Business process.
- Measures are continuously valued results that describe facts.
- A fact is a business statistics.

Dimension:

- The parameter over which we have to perform analysis of facts and data.
- The parameter that gives meaning to a measure number of customers is a fact, perform analysis over time. Dimensional modeling has been coherent architecture for building distributed DW Applications. If we come up with more complex queries for our DW which involves three or more dimensions. This is where the multi-dimensional database plays a eminent role. Dimensions are distributed by which summarized data can be used. Cubes are data manipulating units composed of fact tables and dimensions from the data warehouse (DW). Dimensional modeling also has emerged as the only coherent architecture for building distributed data warehouse Applications.

III. MULTI-DIMENSIONAL MODELING USING BUSINESS DIMENSION

Multidimensional database technology has come a long way since its inception more than 30 years ago. It has recently begun to reach the mass market, with major providers now delivering multidimensional database engines along with their traditional relational database software, often at no extra cost. A multidimensional data model is typically referred for the design of corporate *data warehouses* and *departmental data marts*. Such a model can be adopted with *star schema, snowflake schema*, or *fact constellation schema*. The core of the *multidimensional model* is the data cube, which consists of a large set of *facts* (or *measures*) and a number of business *dimensions*. Business dimensions are the entities or perspectives with respect to organizations that wants to keep information and are hierarchical in nature.Multi-dimensional technology has also made significant gains in scalability and maturity to describe the organizations current business requirement. Multidimensional model is based on three key concepts

- Modeling business rules
- Cube and measures
- Dimensions

Multidimensional data-base technology is a key term in the interactive analysis of large amounts of data for decision-making purposes. Multidimensional data model is introduced based on relational elements. Dimensions are modeled as *dimension relation*s. Data mining applications provides knowledge by searching semi-automatically for previously unknown patterns, trends and their relationships in multidimensional databases structures. OLAP software enables analysts, managers, and executives to gain insight into the performance of an enterprise through fast and interactive access to a wide range of views of data organized to reflect the multidimensional nature of the enterprise wide data.

3.1 The Goals of Multi-Dimensional Data Models

- To enable end-user to access the information in a way that corresponds to his normal understanding of his business, key figures or facts from the different perspectives that relates with the business environment that influence them.
- To facilitate the physical implementation that the software recognizes (the OLAP), thus allowing a program to easily access the data required for processing.

3.2 Usages of Multi-Dimensional modeling use business dimensions

INFORMATION PROCESSING: support for querying, basic statistical analysis, and reporting using crosstabs, graphs, tables or charts. A current trend in data warehouse information processing is to construct low-cost Web-based application tools for global access integrated with Web browsers.

ANALYTICAL PROCESSING Using dimensions with OLAP it includes OLAP operations such as slice-and-dice, drill-down, roll-up, drill-through, drill-across and pivoting. It generally operates on historical data in both summarized and detailed forms. The major strength of on-line analytical processing over information processing is the multidimensional data analysis of data warehouse data.

DATA MINING support with KDD (knowledge discovery in databases) it helps to discovers hidden patterns and associations, clustering, performing classification and prediction, and presenting the mining results using visualization tools etc.

3.3 Logical Multidimensional Model

The multidimensional data model is important because it enforces simplicity. As Ralph Kimball states in his landmark book, *The DW Toolkit*:" The central attraction of the dimensional model of a business is its simplicity that simplicity is the fundamental key that allows users to understand DBs, and allows software to navigate databases efficiently." The multidimensional data model is composed of logical cubes, measures, dimensions, hierarchies, levels, and attributes. The simplicity of the model is inherent because it defines objects that represent real-world business entities. Analysts know which business measures they are interested in examining, which dimensions and attributes make the data meaningful, and how the dimensions of their business are organized into levels and hierarchies. Multidimensional data cubes, are the basic *logical* model for OLAP applications[12]. The focus of OLAP tools is to provide multidimensional analysis to the underlying information. To achieve this goal, these tools employ multidimensional models for the storage and presentation of data. Figure1: Diagram of logical Multi dimensional model

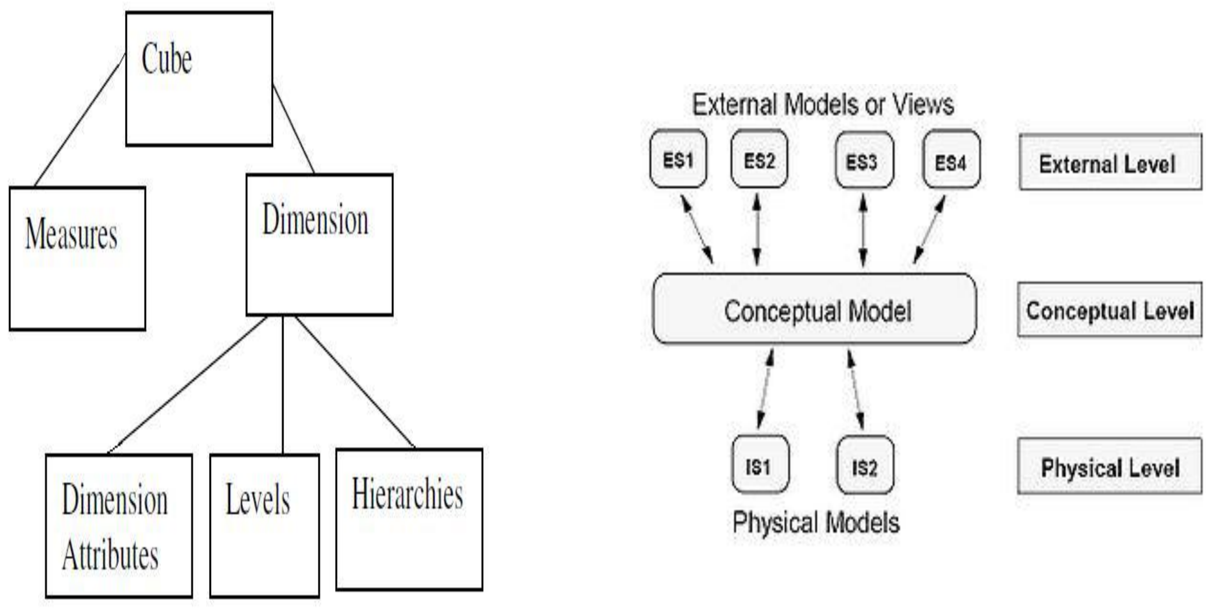


Figure 1: Keys of multidimensional model

A logical model (figure1) for cubes based on the key observation that a cube is not a self-existing entity, but rather a view over an underlying data set. Logical cubes provide a means of organizing measures that have the same shape, that is, they have the exact same dimensions. The relational model forces users to manipulate all the elements as a whole, which tends to lead to confusion and unexpected result sets. In contrast, the multi-dimensional model allows end users to filter each dimension in isolation and uses more friendly terms such as Add, Keep and Remove. Users can quickly and easily create multi level queries. The multi-dimensional query model has one important advantage over the relational querying techniques. Each dimension can be queried separately. This allows users to divide and analyze what would be a very complex query into simple manageable steps. The multidimensional model also provides powerful filtering capabilities. Additionally, it is also possible to create conditions based on measures that are not part of the final report. Because the dimensional query is independent of the filters, it allows complete flexibility in determining the structure of the condition. The relational implementation of the multidimensional data model is typically a star schema, or a snowflake schema.

3.4 Conceptual View

Conceptual view describes the semantics of a domain, being the scope of the model. For example, it may be a model of the interest area of an organization or industry. This consists of entity classes, representing kinds of things of significance in the domain, and relationships assertions about associations between pairs of entity classes. A conceptual view specifies the kinds of facts or propositions that can be expressed using the model. In that sense, it defines the allowed expressions in an artificial 'language' with a scope that is limited by the scope of the model. Early phases of many software development projects emphasize the design of a conceptual data model. Such a design can be detailed

into a logical data model[6]. In later stages, this model may be translated into physical data model. However, it is also possible to implement a conceptual model directly.

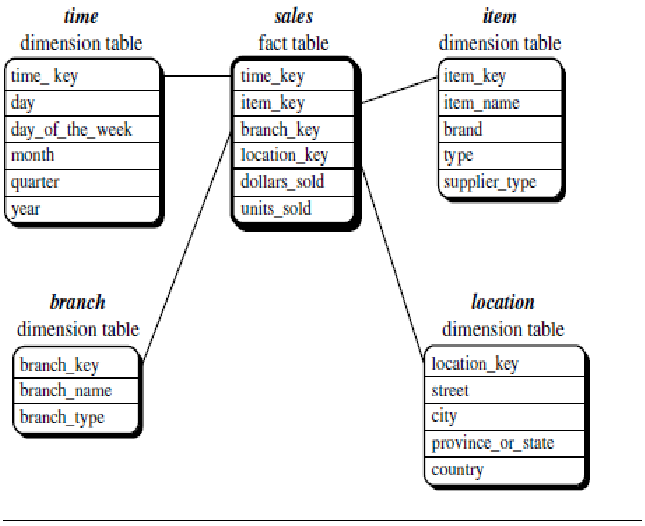
Multidimensional Conceptual View provides a multidimensional data model that is intuitively analytical and easy to use. Business users' view of an enterprise is multidimensional in nature. Therefore, a multidimensional data model conforms to how the users perceive business problems.

3.5 Star schema architecture with business dimension scenario

it consists of a fact table for a particular business process (for example: Sales analysis would take Sales as fact table) with a single table for each dimension table. Star Schema is the special design technique for multidimensional data representations. It Optimize data query operations instead of data update operations. Star Schema is a relational database schema for representing multidimensional data. It is the simplest form of data warehouse schema that contains one or more dimensions and fact tables[15]. It is called a star schema because the entity-relationship diagram between dimensions and fact tables resembles with a star like structure where one fact table is connected to multiple dimensions. The center of the star schema consists of a huge fact table and it points towards the dimension tables. The advantage of star schema is slicing down, performance increase and easy understanding of data.

Steps in designing star schema

- Identify a business process for analysis.
- Identify measures or facts.
- Identify the dimensions for facts.
- List the columns that describe the each dimension.
- Determine the lowest level of summary in a fact table[15].



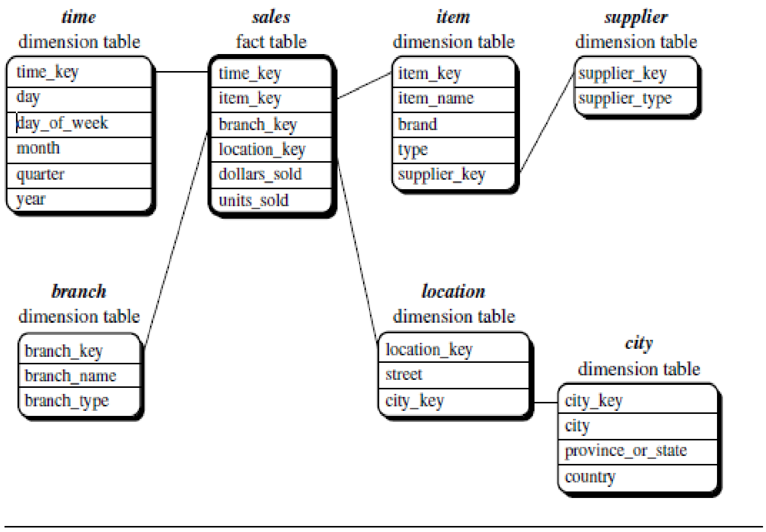
Star schema of a data warehouse for sales.

3.6 Snowflake schema

Snowflake schema: The snowflake schema is a variant of the star schema, where some dimension tables are *normalized*, and enhanced further splitting the data into additional tables[16]. The resulting schema graph forms a shape similar to a snowflake.

3.7 Important aspects of Star Schema & Snow Flake Schema

In a star schema every dimension will have a primary key and also a dimension table will not have any parent table. Whereas in a snow flake schema, a dimension table will have one or more parent tables. Hierarchies for the dimensions are stored in the dimensional table itself in star schema. Whereas hierarchies are broken into separate tables in snow flake schema[16,17]. These hierarchies help to drill down the data from topmost hierarchies to the lowermost hierarchies. Snowflake schema is the normalized form of star schema.



Snowflake schema of a data warehouse for sales.

IV. PROPOSED MODEL

In this section we will summarize the basic concepts of object oriented multi-dimensional model [19]. The multi dimensional model is the core of the comprehensive object oriented model of a DW containing all the details that are necessary to specify a data cube, the dimensions, the classification hierarchies, the description of fact and measures attributes.

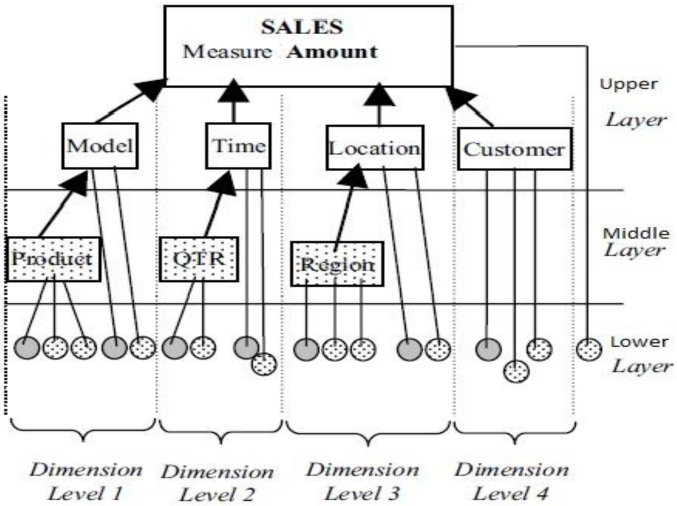


Figure 3: Hierarchical View of Proposed Model Schema

At the lowest layer, each vertex represents an occurrence of an attribute or measure, e.g. product name, day, customer city etc. A set of vertices semantically related is grouped together to construct an Elementary Semantic Group at lower layer. On next, several related elementry are group together to form a Semantic Group at middle layer – the next upper layer constructs to represent any context of business analysis. A set of vertices of any middle, those determine the other vertices of the lower, is called Determinant Vertices. This layered structure may be further organized by combination of two or more middle as well as lower group to represent next upper level layers from the topmost layer the entire database appears to be a graph with middle as vertices and edges between middle layer object. Dimensional Semantic Group is a type of middle layer object to represent a dimension member, which is an encapsulation of one or more lower layer group along with extension and / or composition of one or more constituent middle layer groups. Fact Semantic Group (FSG) is a type of group represents facts, which are an inheritance of all related lower, middle and a set of upper defined on measures. In order to materialize the Cube, one must ascribe values to various measures along all dimensions and can be created from FSG.

Example

Let consider an example, based on Sales Application with sales Amount as measure and with four dimensions – Customer, Model, Time and Location with the set of attributes {C_ID, C_NAME, C_ADDR}, {M_ID, M_NAME, P_ID, P_NAME, P_DESC}, {T_ID, T_MONTH, Q_ID, Q_NAME, YEAR} and {L_ID, L_CITY, R_ID, R_NAME, R_DESC} respectively. Model, Time and Locationdimensions have upper level hierarchies say Product, QTR and Region respectively. Then in the notation of GOOMD model, there will be four middle layer groups {DCustomer, DModel, DLocation, DTime} with hierarchy.

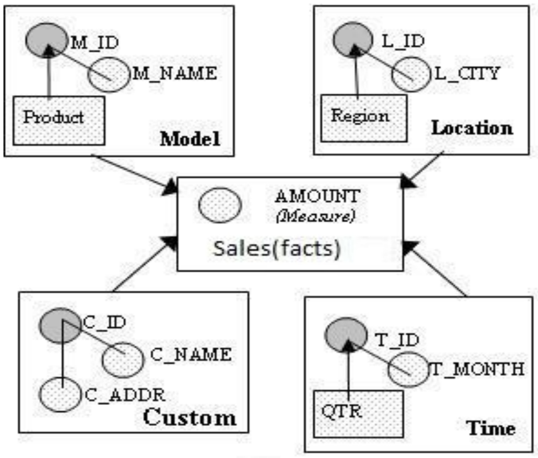


Figure 4: Schema for Sales Application in Proposed Model

model also provides algebra of OLAP operators those will operate on different semantic groups. The Select operator is an atomic operator and will extract vertices from any middle layer groups depending on some predicate P. The Retrieve operator extracts vertices from any Cube using some constraint over one or more dimensions or measures. The Retrieve operator is helpful to realize slice and dice operation of OLAP. The Aggregation operators perform aggregation on Cube data based on the relational aggregation function like SUM, AVG, MAX etc. on one or more dimensions and are helpful to realize the roll-up and drill down operations of OLAP.

V. COMPARISONS OF CONCEPTUAL DESIGN MODELS

Property 1 (Additivity of measures): DF, starER and OOMD support this property. Using ME/R model, only static data structure can be captured. No functional aspect can be implemented with ME/R model.

Property 2 (Many-to-many relationships with dimensions): StarER and OOMD support this property. DF and ME/R models do not support many-to-many relationships.

Property 3 (Derived measures): None of the conceptual models include derived measures as part of their conceptual schema except OOMD model.

Table 1. Comparison of Conceptual Design Models

Property	DF	starER	ME/R	OOMD
1	✓	✓	×	✓
2	×	✓	×	✓
3	×	×	×	✓
4	×	✓	×	✓
5	×	✓	✓	✓
6	✓	✓	✓	✓
7	WAND	×	GRAMMI	GOLD, MS VISIO Rational Rose

Property 4 (Non-strict and complete classification hierarchies)

Although DF and ME/R can define certain attributes for classification hierarchies; starER model can define exact cardinality for non-strict and complete classification hierarchies. OOMD can represent non strict and complete classification hierarchies.

Property 5 (Categorization of dimensions -specialization/ generalization): All conceptual design models except DF support this property.

Property 6 (Graphic notation and specifying user requirements):

All modeling techniques provide a graphical

notation to help designers in conceptual modeling phase.ME/R model also provides state diagrams to model system's behavior and provides a basic set of OLAP operations to be applied from these user requirements. OOMD provide complete set of UML diagrams to specify user requirements and help define OLAP functions.

Property 7 (Case tool support): All conceptual design models except starER have case tool support[18].

VI. CONCLUSION

This paper helps us to enlighten our comprehensibility with the multidimensional structure related to business processes and dimensions. Multi-dimensional data model combined with facts and context dimensions using Star and Snow Flake schema. This paper relates the various multi-dimensional modeling according to the multi-dimensional space, language aspects and physical representation of the traditional Database Model and establish relationship multidimensional data to object oriented data.

REFERENCES

1. Jiawei Han and Micheline Kamber Data Mining: Concepts and Techniques, Second Edition. Morgan Kaufmann publications
2. S.Kelly.Data Warehousing in Action.John Wiley &Sons(1997).
3. Kimball, R. "The Data Warehouse Toolkit: Practical Techniques for BuiIding Dimensional Data Warehouses". John Wiley and Sons 1996. ISBN 0-471-15337-0
4. S.Chaudhuri,U.Dayal.An overview of data warehousing and OLAP technology. SIGMOD Record 26,1 (1997).
5. G.Colliat.OLAP, relational and multi-dimensional database systems.SIGMOD Record 25, 3 (1996)
6. M. Golfarelli, D. Maio, and S. Rizzi. The dimensional fact model: a conceptual model for data warehouses.
7. M. Jarke, M. Lenzerini, Y. Vassilious, and P. Vassiliadis, editors. Fundamentals of Data Warehousing.
8. L. Cabibbo and R. Torlone. A logical approach to multidimensional databases. In Proc. of EDBT-98, 1998.
9. E. Franconi and U. Sattler. A data warehouse conceptual data model for multidimensional aggregation. In Proc. of the Workshop on Design and Management of Data Warehouses (DMDW-99), 1999.
10. McGuff, F. "Data Modeling for Data Warehouses"October, 1996 from http://members.aol.corn/fmcguff7dwmodel/dwmodel.html
11. Gyssens, M. and Lakshmanan, L.V.S. "A foundation for multi-dimensional databases," Technical Report, Concordia University and University of Limburg, February 1997.

12. M.Blaschka,C.Sapia,G.H¨ ofling, and B. Dinter. Finding Your Way through Multidimensional Data Models. In DEXA ’98, pages 198–203, 1998. http://www.pentaho.org (16.06.2006), 2006.
13. Antoaneta Ivanova, Boris Rachev Multidimensional models- Constructing DATA CUBE International Conference on Computer Systems and Technologies - CompSysTech’2004
14. Multidimensional Database Technology by Torben Bach,Pedersen,Christian S. Jensen,Aalborg University
15. Rakesh Agrawal, Ashish Gupta, and Sunita Sarawagi.Modeling multidi-mensional databases. Research Report, IBM Almaden Research Center, San Jose, California,1996.
16. P. Vassiliadis and T.K. Sellis, “A Survey of Logical Models for OLAP Databases,” ACM SIGMOD Record, vol. 28, no.4, 1999.
17. L. Cabibbo and R. Torlone. A logical approach to multidi-mensional databases. In Proc.of EDBT-98, 1998
18. Deepti Mishra, Ali Yazici, Beril Pinar Başaran “A Casestudy of Data Models in Data Warehousing”.
19. Anirban Sarkar, Swapan Bhattacharya “Object Relational Implementation of Graph Based Conceptual Level Multidimensional Data Model”