Singaporean Journal Scientific Research (SJSR)
ISSN: 2231 - 0061Vol.3, No.2 pp.166 - 168

available at: : http://www.iaaet.org/sjsr

A Convenient Way fromNormalized Database toDeformalized Database

Jyoti Singh[1], Bhupendra Singh[2], Y. Sriveni[3]
[1] GCET/CS & E, Greater Noida, India
yrjyoti@gmail.com
[2,3]Amity University/ECE, Noida, India/[2] GCET/CS, Gr. Noida, India
bsingh.tech@gmail.com, sriveni123@gmail.com

Abstract—**Data warehouse is a denormalized and information oriented system used to convert business rules into useful information as it is the collection of archived operational data which is useful in tactical and strategic business decisions. A relationaldatabase consists of a collection of tables that store particular sets of data in a form of normalized database. We use ER Model in relational database to define our overall system on the other hand in data warehouse we use dimensional cubes. In this paper we will discuss the techniques to transform the relational database into data warehouse. Here, we will concentrate on dimensional model as it heavily aggregated on business dimensions to improve retrieval performance rather than optimization of update performance as in the Entity Relationship model in relational database. The method for construction of data warehouse using dimensional model is presented which uses related data of operational system as its base for identifying important objects. We identify relationship between the objects that is different categories of objects. Data warehouse is designed to support ad hoc queries and strategic decision analysis which is not possible with traditional operational systems in which we require multiple tables join that is a huge web of joins using which strategic business decision making is not possible so here we introduced a technique which transforms already existing system into useful information retrieval system used to make tricky and tactical business decisions.**

Index Terms—Entity-Relationship Model, Dimensional Model, Dimension Table, Fact Table, OLTP, OLAP, Snowflake Schema, Star Schema

1. INTRODUCTION

The data warehouse is considered as subject oriented, integrated, and nonvolatile and time variant collection of data, in support of management'sdecision-making [1].

The huge amount of data is available that can be transformed to provide enterprise-wide view of information which is used as the root for strategic business decisions. It is possible to get on the fly information based on business dimensions. For the new enterprise-wide information architectures now emerging [2].Data warehouse captures the overall business process, the different business views and business trends. Data warehouse uses OLAP to view and analyze along different business dimension increasing analytical capability. The current and historical data is are available giving large size of result set to the end user queries leading to high update frequency ,highly optimized database for user interaction and analyzes.

2. COMPONENTS OF DATA WAREHOUSE ARCHITECTURE

The components of Data Warehouse Architecture are:

Relational Database: A relationaldatabase consists of a collection of tables that store particular sets of data.

External sources: The data from external sources is also used and merged with data warehouse for performing different statistical analysis.

Data Staging: Data Warehouses are constructed from different operational systems. Data staging is required as Data warehouse is subject oriented. Here, conversion, combination and refinement of data take place. Finally, data is stored in data warehouse through data loading.

Data Transformation and Loading: To make data suitable for data warehouse data transformation is done. Data is selected, converted, summarized, integrated and measured. Data loading is done on transformed data which is to be loaded in the datawarehouse.

Central Data Warehouse: A Central Data Warehouse is a repository of datawhich is created from operational systems. It acts as a consistent enterprise data model to ensure consistency in strategic decision making. A data warehouse can be viewed as an operational data store as contains data extracted, transformed and loaded form operational systems [1].

Architected Data Marts: Data mart is bound to a particular subject area.

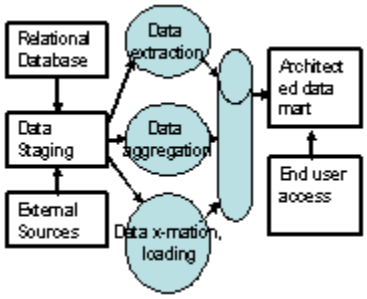

Figure 1. Data Warehouse Architecture

3. DIMENSIONAL MODELING FOR BUSINESS DATAMODEL

A Data Warehouse designed using a Dimensional modeling technique. Dimensional Modeling presents star schema with a central fact table and dimension table connecting to fact table using dimension keys. It also consists of database tables and relations is made along Business dimensions to support optimization of relational database to get accurate and reliable data for strategic business decision making based on metrics and measurements of the overall business process defined by the data warehouse.

Indeed Kimball has stated that while star schemas have the advantages of greater understandability and superior performance relative to E-R models, their use involves no loss of information, because any E-R model can be represented as a set of star schema models without loss of information [3].

Here is a Question" Why E-R modeling technique is not suitable for designing Data Warehouse Data model." Answer is simple" because Dimensional modeling has some special characteristics/ advantages over E-R modeling."

A. Characteristics/ Advantages of Dimesional Modeling

- Capture Critical Measures
- Views of Dimensions
- Intuitive to Business Users

*User friendly:*Data warehouse is provides user-friendly environment as end user can make statergic decision easily based on business dimensions and can perform metric measurement like aggergation and summarization by drilling up the granularity level on dimensions and drilling down to the detailed level.

*Flexible:*In Data warehouses the data is inserted form relational database and data can be retrieved by f user based on the subject area which ensures flexibility of data retrieved as a result. of the query.

*Scalability:*Since warehouses store information about the database as it progresses over time [4], they tend to grow much more rapidly than relational databases.

*Maximize Effeciency of queries:*It achieves these objectives primarily by minimising the number of tables and relationships between them. This reduces the complexity of the database and minimizes the number of joins required in user queries[6].

B. E-R Modeling Vs Dimensional Modeling

- The relational database uses E-R model to define the structure of the overall system consists of real world objects and relationship between them. Data warehousing uses dimensional model to define the business requirements.
- The E-R Model is based on normalization process i.e. segmentation of available data into multiple tables. Dimensional model is based on denormalization of data i.e. summarized and aggregated data.
- Using E-R model to make a query by end-user requires multiple table joins because of normalization of data. Dimensional model on the other hand rectifies the problem of large joins as data here is summarized depending on business dimensions, metric measurements and granularity level [1, 4].
- E-R model is suitable to remove data redundancy. As Normalization the objective of normalization is to minimize data redundancy (Codd, 1970). This maximizes update efficiency because each change can be made in a single place, but tends to penalize retrieval (Kent, 1978) [5].Dimensional Model provides the critical measures that are viewed along several business dimensions that is centered around a business process.

4. A PATH FROM NORMALIZED DATABASE TO DENORMALIZED DATABASE

The Entity Relationship/ normalized model/ is converted to Dimensional/ deformalized Model which is based on denormalization for the fast retrieval of data by the end-user without any loss of information. Here we have a convenient way convert Entity Relationship Model to Dimensional Model.

A. STEP 1

First, the related data of operational system is used as base for identifying important objects. The entities of the system are identified first by using data model of normal OLTP which is used in transaction processing environment consists of entities which are related to each other. For example, the student academic data model. In this data model the one-to-many relationship can be identified between different objects to establish a connection between objects of the data model.

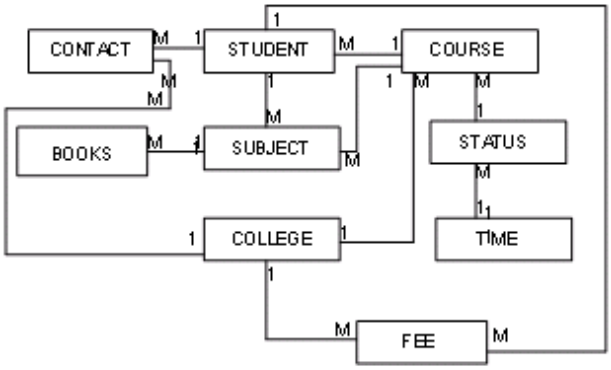

Figure 2. E-R Data Model

B. STEP 2

Second, to identify relationship between the objects i.e. different categories of objects. Using the objects identified from Entity relationship model in first step we categorize the objects depending upon the cardinality ratio between the objects. The entities responsible for some important event and transaction to occur in the business cycle which is important for strategic decision making and market analysis. The key metric measurements to be calculated aggregated and summarized which are required facts in the fact table.

C. STEP 3

Third, performing the aggregation of objects to improve the retrieval performance. The parent-child hierarchy is followed to aggregate and summarize data. The path from parent to child and from child to parent is to be identified. The path is of the entities connected together or merged in sequence by one-to-many cardinality ratio from parent to child and many to one from child to parent. For example, the following are object and the possible hierarchies. The business dimensions and facts are known by categorizing objects depending upon there relationship.

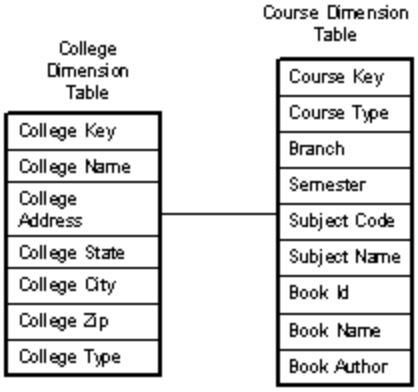

Figure 3.Merging Hierarchical Dimensions

D. STEP 4

Fourth, is to produce dimensional model using star schema design technique. The objects are merged together based on there primary key-foreign key relationship forming a hierarchy as dimensional modeling uses denormalisation climbing up hierarchy at each granularity level. The dimension attributes are grouped together to get some quantitative measurements which are grouped together for aggregation and summarization of data The attributes are grouped using primary key of dimension table.

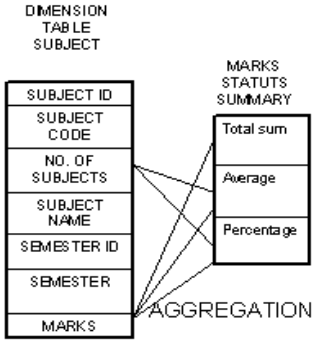

Figure 4. Aggregate summarized key metric measurement

STAR SCHEMA "It consists of one central fact table and dimension table surrounding it which are connected to fact table through their primary keys. Dimension table contain attributes and fact table contain keys of dimensional tables and aggregate facts and measurements."

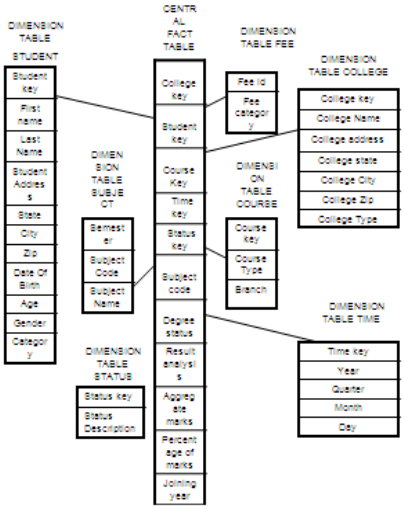

Figure 5. Student Analysis Star Schema

REFERENCES

[1] G. W.H. Inmon, Building the Data Warehouse, John Wiley and Sons,1992.

[2] JaideepSrivastava, Senior Member, IEEE, and Ping-Yao Chen, Student Member, IEEE,"Warehouse Creation-A PotentialRoadblock to Data Warehousing", EETransaction on Knowledge and Data Engineering, Vol. 11, No. 1, Januaray/Feburary 1999.

[3] Dimensional Modeling and E-R Modeling In The Data Warehouse By Joseph M. Firestone, Ph.D.White Paper No..Eight , June 22, 1998

[4] Ponnaiah, Paulraj, Data Warehousing Fundamentals, Wiley-Students Edition, 2001.

[5] From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design,Daniel L. Moody,Department of Information Systems,University of Melbourne, Parkville, Australia 3052,Mark A.R. KortinkSimsion Bowles & Associates 1 Collins St, Melbourne, Australia 3000.

[6] Multi-Dimensional Modeling with BI, A background to the techniques used to create BI InfoCubes,Version 1.0,May 16, 2006.