International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 3 Issue 4, April-2014, pp: (409-414), Impact Factor: 1.252, Available online at: www.erpublications.com

Conversion from star schema of a data warehouse into its equivalent graph multi dimensional data model: A Conceptual Perspective

S. K. Abdul Rahim

Department of Computer Science & Engineering, Bengal College of Engineering & Technology, Durgapur, India

Abstract: This paper proposes the design of Star Schema of a Data Warehouse and conversion to its equivalent Graph based Multi Dimensional Data Model (GMDDM Model). Design of Star Schema facilitates in constructing Graph based Multi Dimensional Data Model. The GMDDM model defines a set of graph based constructs that are used to specify the conceptual level design of Data Warehouse. Data warehouse representation is complex itself. Thus the pictorial representation of the Data Warehouse should increase the understandability to the warehouse designer. The warehouse representation through graph G [V, E] may be an effective approach that reduces the inherent complexity of any multi – dimensional data model. But there is no need to develop this model from scratch, rather, in contrast, it may be considered as a software layer on the top of any standard data model representation. Here the scope of the project work is to form a software tool that is able to convert the Star Schema of a Data Warehouse to its equivalent Graph based Multi Dimensional Data Model (GMDDM Model).Hence the software tool helps the warehouse designer to interact with the graph without writing any code and thus ease the entire complex design mechanism. The project work contains a detailed description of the Star Schema of a Data Warehouse say Hospital Management System and conversion from Star Schema to its equivalent Graph based Multi Dimensional Data Model and the conversion rules from Star- Schema to its equivalent GMDDM model.

Keywords: Star Schema, Graph Multi–Dimensional Data Model (GMDDM), Object – Oriented Schema, Warehouse Designer, Software Layer, Software Tool, Graph Data Model.

I. INTRODUCTION

A Data Warehouse as in [3] is a repository of subjectively selected and adapted operational data which can successfully answer any ad – hoc, statistical and analytical queries. It is situated on the top of a decision support system. A data warehouse as in [6] is Subject – Oriented, Integrated, Time – Variant, Non – Volatile collection of data in support of management's decision making process. Data Warehouse design framework as in [11] should span in three perspective namely, the conceptual level, the physical level and the logical level.

The conceptual level which deals with the high level representation of the world in order to capture the user ideas using rich set of semantic constructs. The physical level deals with the details of the representation of the information or data storage techniques in the specific DBMS. The logical level acts as an intermediate between the two, trying to balance a storage independent paradigm and a natural representation of the information. Conventional OLTP (Online Transaction Processing) database applications are developed to meet the day - to - day database transactional requirements and operational data retrieval needs of the entire user community. Ordinary Transaction processing systems are not able to

analyze online a large number of past transactions. Any multidimensional data such as spreadsheet can not be processed by conventional SQL type DBMS. For a complex real life problem, e.g. Online Shopping Management System, the complex queries (e.g. display the total sales of each product on all regions and also find the corresponding region wise total sales of each product online).

For such type of complex real life problem, Data Warehouse based OLAP (Online Analytical Processing) is used. OLAP systems as in[9] which are contrary to the regular, conventional Online Transaction Processing (OLTP) are capable of analyzing online a large number of past transactions or data records (ranging from mega bytes to giga bytes and tera bytes) and summarize them. This type of data is usually multidimensional in nature. This multidimensionality is the key driver for OLAP technology, which happens to be central to Data Warehousing. DW data as in [7] can be dynamically manipulated using on-line analytical processing (OLAP) systems. DW and OLAP systems rely on a multidimensional model that includes measures, dimensions, and hierarchies.

For a complex real – world problem, as in [9] the data is usually multidimensional in nature. Even though one can manage to put such data in a conventional relational database in normalized tables, the semantics of multidimensionality will be lost and any processing of such data in the conventional SQL will not be capable of handling it efficiently. As such multidimensional query on such a database will explode into a large number of complex SQL statements each of which may involve full table scan, multiple joins, aggregation, sorting, and also large temporary table space for storing temporary results. Data warehousing based OLAP tools are developed to meet the information exploration and historical trend analysis requirements of the management or executive user community. The conventional, regular database transactions or OLTP transactions are short, high volumes and provide concurrent and online update, insert, delete in addition to retrieval queries and other procedures, processing or reporting.

II. STAR SCHEMA

A Star-Schema as in [10] is a logical structure that has a fact table containing factual data in the center, surrounded by dimension tables containing reference data (which can be denormalized). In data warehousing and business intelligence , a star schema is the simplest form of a dimensional model, in which data is organized into facts and dimensions. A fact is an event that is counted or measured, such as a sale or login. A dimension contains reference information about the fact, such as date, product, or customer. A star schema is diagramed by surrounding each fact with its associated dimensions.

The resulting diagram resembles a star. The Star-Schema exploits the characteristics of factual data such that facts are generated by events that occurred in the past, and are unlikely to change, regardless of how they are analyzed. As the bulk of data in a data warehouse is represented as facts, the fact tables can be extremely large relatively to the dimension tables. As such, it is important to treat fact data as read-only reference data that will not change over time. The most useful fact tables contain one or more numerical measures or facts that occur for each record. Star Schema can be used to speed up query performance by denormalizing reference information into a single dimension table. Dimension tables, by contrast, generally contain descriptive textual information.

Dimension attributes are used as the constraints in data warehouse queries. Star schemas are optimized for querying large data sets and are used in data warehouses and data marts to support OLAP cubes, business intelligence and analytic applications, and ad hoc queries. Within the data warehouse or data mart, a dimension table is associated with a fact table by using a foreign key relationship. The dimension table has a single primary key that uniquely identifies each member record (row). The fact table contains the primary key of each associated dimension table as a foreign key. Combined, these foreign keys form a multi-part composite primary key that uniquely identifies each member record in the fact table. The fact table also contains one or more numeric measures.

AN EXAMPLE OF STAR SCHEMA OF HOSPITAL MANAGEMENT SYSTEM:

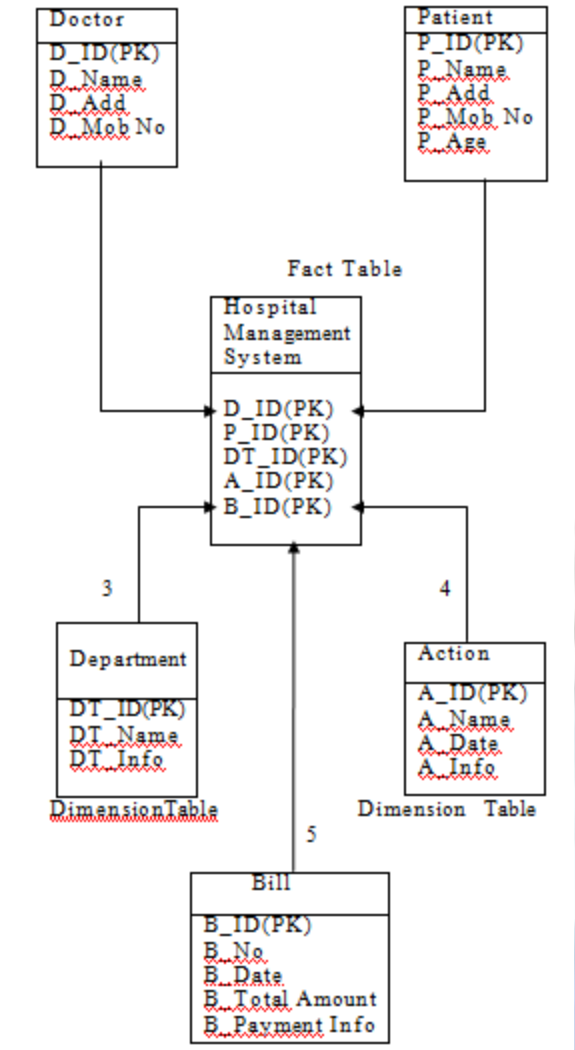

Fig.1 Star Schema

III. DESCRIPTION OF GMDDM MODEL

The GMDDM model as in [12] is the basis of the object oriented model of a Data Warehouse. It contains all the details that are necessary to specify a data cube. It includes as in [6] a description of the dimensions, the classification hierarchies, and a description of the context of analysis i.e. Fact and the quantified attributes of facts (measures). In this context, a fact as in [5] is a collection of related data items, consisting of measures and business context data e.g. total sales of a Shopping House within a certain period. A fact table is a central table. It contains mostly raw numeric items, narrow rows (corresponding to small volume of data inserted), a few columns at most and large number of rows. It is accessed via dimensions. Dimensions as in [5] are the parameters over which OLAP are performed e.g. Customers, Banks, Time, location, company etc. A dimension table defines business in terms of already familiar to users. It contains wide rows (corresponding to large volume of data inserted), with lots of descriptive text, small tables. It is joined to a fact table by a foreign key. It is heavily indexed The Graph data model as in [3] allows the entire multidimensional database to be viewed as a Graph (V, E) in layered organization. At the lowest layer, each vertex represents an occurrence of an attribute or a data item, each basic attribute is to be represented as separate vertex. A set of vertices semantically related is grouped together to construct an Elementary Semantic Group (ESG). So an ESG is a set of all possible instances for a particular attribute or data item. Several related ESGs are grouped together to form a Contextual Semantic Group (CSG). The edge within CSG represents the association between different ESG in the said CSG. The most inner layer of CSG is the construct of lowest level of granularity of the contextual data in Multidimensional database formation. This layered structure may be further organized by the combination of two or

more CSGs as well as ESGs to represent next upper level layers and to achieve further higher level granularity of contextual data.

IV. COMPONENTS OF GMDDM MODEL

A. Elementary Semantic Group (ESG)

An elementary semantic group as in [2] is an encapsulation of all possible instances or occurrences of an attribute or measure represented in a graph ESG (V, E) where the set of edges E is a null set and the set of vertices V represent the set of all possible instances of an attribute or a measures.

B. Contextual Semantic Group (CSG)

A contextual semantic group as in [2] is an encapsulation of two or more related ESGs to represent one elementary context of business analysis. A CSG is the construct for different level of granularity of the contextual data and is used to exhibit respective level of details in Multidimensional database formation.

C. Dimensional Semantic Group (DSG)

A dimensional semantic group in [1] is an inheritance or encapsulation of one or more CSGs along with ESGs.

D. Fact Semantic Group (FSG)

A fact semantic group is in [1] an inheritance of all related DSGs with relevant business context data and a set of measures. One FSG can be represented as a graph.

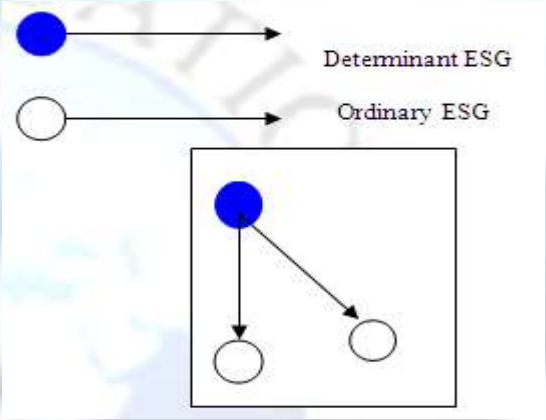

Fig. 2: Example of DSG

V. RULES FOR CONVERSION FROM STAR-SCHEMA TO GMDDM MODEL

Step1: Convert every Dimension table into a 'DSG'

Step2: Convert every attribute of dimension table. into a 'ESG' in a 'DSG'

Step3: Convert every Primary Key in a dimension table into a 'Determinant ESG' in a 'DSG'

Step4: Convert 'Fact' table into a. 'FSG' in GMDDM.

Step5: Connect every 'DSG' with the'FSG' as every Dimensional table is connected with the Fact table' in the Star Schema through Foreign Key (FK). i.e. every Primary Key (PK) of every Dimension table will be converted as a Foreign Key (FK) of Fact table.

Step6: Store the pre-calculations in the Fact table.

VI. EQUIVALENT GMDDM MODEL

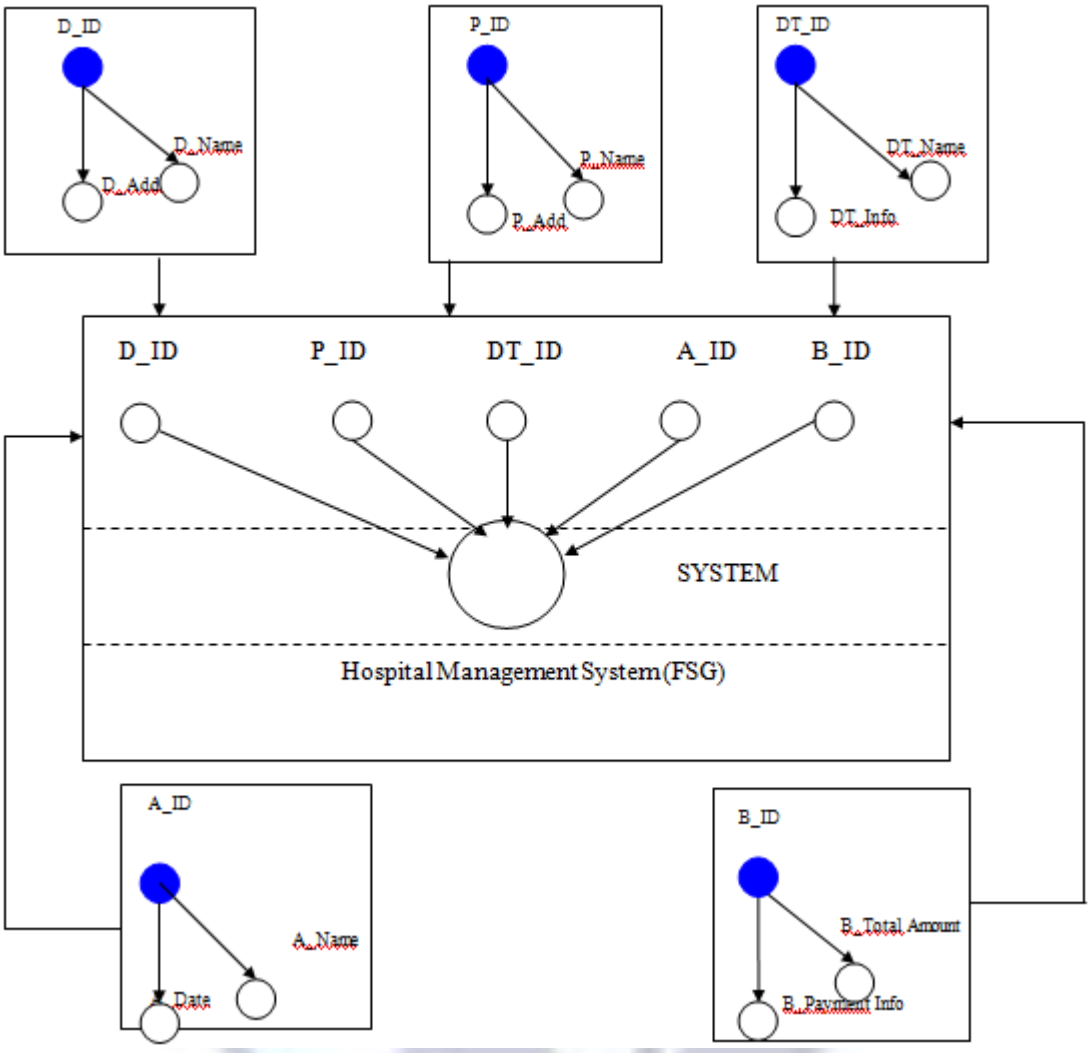

Fig.3: Gmddm Model of Hospital Management System

VII. CONCLUSION

In this paper, an attempt has been made to represent the Star Schema of a Data Warehouse. An attempt also has been made to convert the Star Schema of a Data Warehouse into its equivalent Graph Multi-Dimensional Data Model in which the entire multidimensional database can be viewed as a Graph [V, E] in layered organization.. We have also illustrated the proposed conversion rules from Star Schema of a Data Warehouse to its equivalent GMDDM model. Further, the future scope of the project work is to develop a software tool that is able to convert the Star Schema of any Data Warehouse into its equivalent Graph Multi-Dimensional Data Model So, the software tool helps the warehouse designer to interact with the graph without knowing anything about Object – Oriented representation. It facilitates the entire complex design mechanism to be very easy.

REFERENCES

[1]. Anirban Sarkar, Swapan Bhattacharya. "The Graph Object Oriented Multidimensional Data Model: A Conceptual Perspective ", SEDE 2007: 165-170.
[2]. S. Choudhury, N. Chaki, S. Bhattacharya, "Graph Object Oriented Model and Query Language: A Semi – Structured Approach ", IEEE Sponsored International Conference on Info. Tech: Coding & Computing (ITCC 2001), Nevada, USA, pp.685 – 689, 2001

[3]. S. Choudhury , N. Chaki , S. Pramanik , S. Bhattacharya , "Conceptual Level Graph Theoretic Design and Development of Complex Information System " , IEEE Sponsored International Conference on Info. Tech: Coding & Computing (ITCC 2000), Nevada, USA, pp.449 – 454, 2000

[4]. Anirban Sarkar, Sankhayan Choudhury, Nabendu Chaki, Swapan Bhattacharya,] "Conceptual Level Design of Object Oriented Data Warehouse: Graph Semantic Based Model"

[5]. S. Choudhury , S. Pramanik , S. Bhattacharya , " Graph Theoretic modeling of semi – structured Information system based on functional abstraction ", Proc. Of IASTED International Conference on Applied Modeling and Simulation, 1998, pp. 528 – 522

[6]. Anirban Sarkar, S. Choudhury, N.Chaki, S. Bhattacharya, "Implementation of Graph Semantic Based Multidimensional Data Model: An Object Relational Approach", International Journal of Computer Information Systems and Industrial Management Applications ISSN 2150-7988 Volume 3 (2011) pp. 127-136

[7]. Elzbieta Malinowski, Esteban Zimanyi,"Designing Conventional, Spatial, and Temporal Data Warehouses: Concepts and Methodological Framework.". Print ISBN:978-3-540-744047,2006-2007

[8]. Surajit Chaudhuri Microsoft Research, Redmond Umeshwar Dayal Hewlett-Packard Labs, Palo Alto , "An overview of data warehousing and OLAP technology" Newsletter ACM SIGMOD Record Homepage archive Volume 26 Issue 1, March 1997 Pages 65-74 ACM New York, NY, USA.

[9]. C.S.R Prabhu ,"Data Warehousing: Concepts, Techniques, Products and Applications: Second Edition, 2002" ISBN – 81-203-2068-9

[10]. Gajendra Sharma, "Data Mining, Data Warehousing and OLAP", second edition, 2008-2009, ISBN: 978 – 81 – 89757 – 47 - 2

[11]. Anirban Sarkar, "Conceptual Level Design of Semi-structured Database System: Graphsemantic Based Approach", International Journal of Advanced Computer Science and Applications, The SAI Pubs., New York, USA, Vol. 2, Issue 10, PP 112 – 121, November, 2011. [ISSN: 2156-5570(Online) & ISSN : 2158-107X(Print)]

[12]. Sk. Abdul Rahim, Baisakhi Chakraborty , Joyati Debnath and Narayan Debnath,, "Design Graph Multi-Dimensional Data Model of a Data Warehouse and conversion of its equivalent Object – Oriented Schema ". Iscc,pp.000079-000084,2013 IEEE Symposium on Computers and Communications (ISCC), 2013,ISBN : 978-1-4799-3755-4.