International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 05 | May-2016 www.irjet.net p-ISSN: 2395-0072

A Data Warehouse Implementation Using the Star Schema

For an outpatient hospital information system

GurvinderKaurJosan

Master of Computer Application,YMT College of Management Kharghar, Navi Mumbai

---

Abstract *: Today's business requires that companies have access to highly relevant information in a matter of seconds. Modern Business Intelligence tools depend on data structured mostly in traditional dimensional database schemas, typically represented by star schemas. This work explores using the star schema for a data warehouse. An implementation of a data warehouse for an outpatient hospital information system will be presented as an example. Explanations of the many data warehouse concepts will be given.*

Keyword-Database, Data warehouse, Star Schema

I. INTRODUCTION

A data warehouse consist of integrated granular historical data. If there is any secret to a data warehouse it is that it consist of data that is both granular and integrated. The integration of the data allows a corporation to have a true image of enterprise. Instead of looking at data parochially, the data analyst can look at it collectively, as if it had come from a single and unique well-defined source, which most data warehouse data assuredly does not. So the ability to use data warehouse data to look across the enterprise is the first major advantage of the data warehouse. Additionally, the granularity that is the fine level of detail allows the data to be very flexible. Because the data is granular, it can be examined in one way by one group of people and in another process by another group of people. Granular data means that there is still only one set of data. Finance can look at the data one way, marketing can look at it in another way, and accounting can look at the same data in another way. If it fact out that there is a difference of judgement, there is a single sort of fact that can be returned to resolve the difference. Another major profit of a data warehouse is that it is a historical storage of data. A data warehouse is a good place to store several years value of data. It is for these main reasons and more that the idea of a data warehouse has gone from a theory derided by the data base theoreticians of the day to conventional wise thinking in the corporation today. But for all the profit of a data warehouse, it does not come without some degree of pain.

(1) The organization data will get divided into input process which is called as operational data and the system which is used to handle that data called as operational system.

(2) The operational data is not used for decision making process.

(3) So to perform the decision making by top-level management different operational data The main goal of this paper is to introduce the reader to data ware housing concepts and their terms. It will define concepts such as OLAP, OLTP, data marts, dimensional models, fact tables, enterprise-wide data warehouse, dimension tables, and the star schema. The study will also explore the implementation of a data mart for an outpatient hospital information system using the star schema .

II. NEEDS OF DATAWAREHOUSE

Normally in an organization in 1990,each organization produces large amount of data about their customers ,employees and products. To store these data they generally use a single repositions in which storing a data is not a problem but retrieving a data is very difficult rather than time consuming. These problem is called as “DATA IN JAIL”. To solve this problem the organization will get treated like simple system which contain 3 parts . (a) INPUT COMPONENT (b)PROCESS (c)OUTPUT

III.SIGNIFICANCE OF IMPLEMENTING DATAWARE HOUSE

Dataware house is a group of multiple databases. Datawarehouse is a repository which is used to store strategic data or informational data which is used to support decision making. Datawarehouse is defined as subject oriented ,non-volatile ,intergrated,timevarient , collection of data which support in management decision making .[2]

(a) SUBJECT-ORIENTED –It is a subject oriented because it contains multiple databases having some names that of subject names.

(b) INTERGRATED – Dataware house is integrated because it cllects the data from multiple sources .

(c) NON-VOLATILE –In Dataware house we cannot delete a data ,we can modify it but in that case we need to store both modified and older version of data .

(d) TIME VARIENT – In Dataware house we store urgent as well as historic data.

IV. CONSTRUCTION OF DATAWARE HOUSE

Dataware houses and their architectures vary depending upon the specifics of an organization's situation. Three common buildings of Dataware house are:[2],[5]

- (A) Architecture 1- (Basic Architecture)
- (B) Architecture 2 -(Architecture with a Staging Area)
- (C) Architecture 3- (Architecture with a Staging Area and Data Marts)

(A) Architecture 1- (Basic Architecture)[2]

Fig.1An architecture for a data warehouse. End users directly access input derived from several origin systems through the dataware house.

In Fig.1, the metadata and fresh data of a traditional OLTP system is present, as is an additional type of data, summary data. Brief statement are very valuable in dataware houses because they pre-compute long operations in advance. For example, a dataware house query is to retrieve August sales. A brief statement in Oracle is called a materialized view.

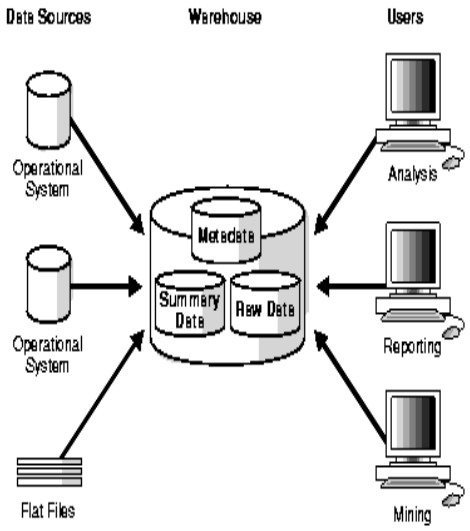

Fig.1 Architecture of a Data Warehouse

(B) Architecture 2 (Architecture with a Staging Area)[2]

In Fig.1, you need to clean and process your information before adding or updating it into the dataware house. Although most dataware houses use a staging area. Architecture with staging area simplifies building summaries and illustrated in

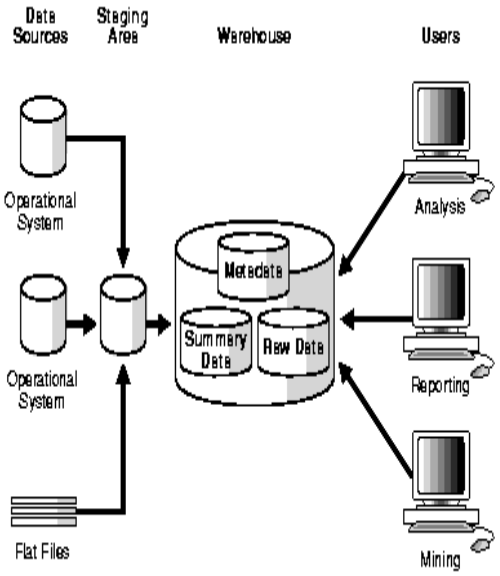

Fig.2 Architecture of a Data Warehouse with a Staging Area

(C) Architecture 3 (Architecture with a Staging Area and Data Marts)[2],[6]

Although the architecture in Fig.2 is quite obvious, you may want to custom your warehouse's architecture for different groups within your organization. You can do this by simply adding data marts .Fig.3 illustrates an example where, sales, inventories and purchasing are separated. In this example, a financial analyst may want to analyze historical data for purchases and sales.

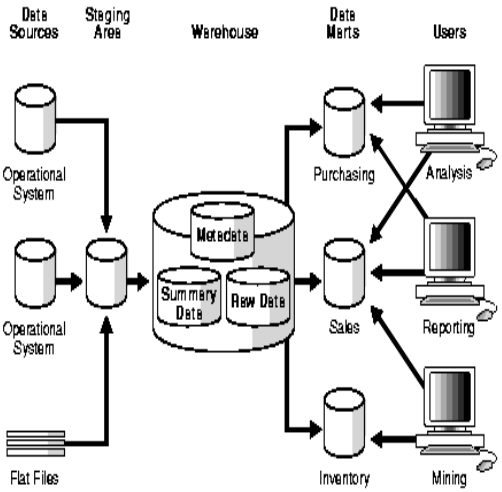

Fig.3 Architecture of a Data Warehouse with a Staging Area and Data Marts

V. DIMENSIONAL MODELLING :

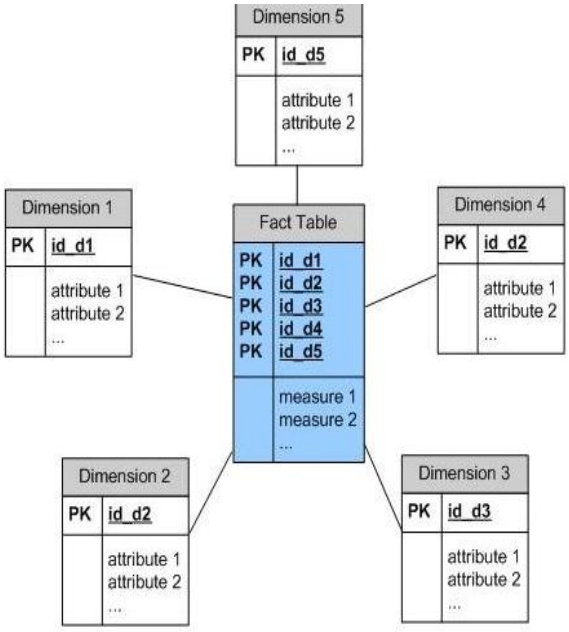

Fig.4 STAR SCHEMA

It is a process which convert the information package into a digrametical representation. To represent the data of dataware house diagrammatically we use STAR SCHEMA. It contain following two types of table: (a) FACT TABLE-it contain primary key of fact table,primary key of dimension table and fact that we want to analyse. (b)DIMENSION TABLE –it contain primary key of dimension table and attribute of dimension table .[3]

(A) Basic structure of STAR SCHEMA

If a single fact table is surrounded by as well as connected to multiple dimension tables then this structure is called as STAR SCHEMA.Star schemas are denormalized, meaning the normal rules of normalization applied to transactional relational databases. The benefits of star schema denormalization are:[3]

- ***Simpler queries*** - used to retrieve data from a highly normalized transactional schemas.
- ***Simplified business reporting logic*** - It simplifies common business reporting logic that is period-over-period and reporting.
- ***Query performance gain*** - Star schemas can provide performance improvement for read-only reporting applications.
- ***Fast aggregations*** - The simpler queries against a star schema can result in good performance for aggregation test operations.
- ***Feeding cubes*** - Star schemas are used by all OLAP systems to build OLAP cubes efficiently; in fact, most major OLAP systems provide a ROLAP mode of operation which can use a star schema directly as a medium without building a proprietary cube structure.

EXAMPLE

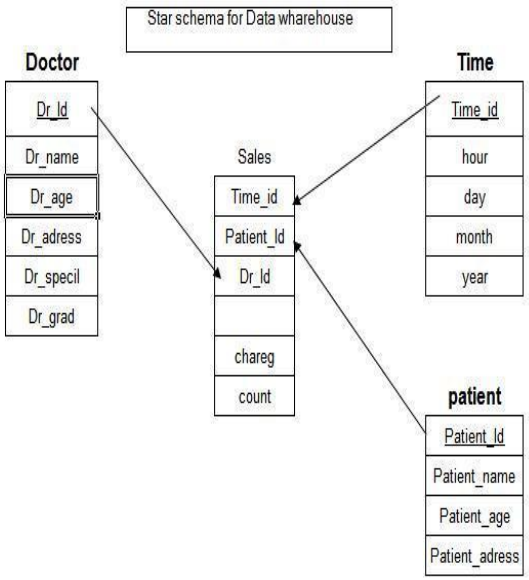

Fig.5 Star Schema*for an outpatient hospital information system*

The disadvantage of the star schema is that data integrity is not enforced as well as it is in a highly normalized database[3]

(1) Poor Data Integrity – Due to the non normalized structure of these tables, information can be replicated, creating several anomalies in the data.

(2) Long time loading dimension table – When the data integrity is low and replication values high, loading time of the tables increases.

(3) More disk space

(4) Additional processing – Usually some controlling processes are added to avoid the data integrity issue.

*(5)**Harder Complex Queries*** – Since the data schema is built specifically to examine a set of data and its de-normalized organization makes it harder to develop new complex queries.

(6) No Many-to-Many – This schema has no many-to-many relationships

VI.RELATIVE WORK

The main problem is the database has limited storage for saving the data . That leads to keep less data and weak in examine because of database attributes are imperfect as well. Perfect in examine and huge storage also fast responds for quires and better behaviour. To meet these needs there is a new type of technique that has evolved.[4]

(A) Federated database

Conceptual Federated Dataware architecture is supporting the tightly coupled integration of Datamarts into a global Dataware schema. While a global schema hides data heterogeneity from the users, the local Datamarts retain information and schema . The canonical information model of our Federated architecture supports Datamarts implemented on relational and multidimensional physical platforms alike. As such, the architecture is ideal for independent organizations that want to distribute their Datawarehouse. Our future and current work involves the implementation of a prototype demonstrating the ability of our approach. The prototype is implemented using off-the-shelf software as far as possible. Particularly, the SQL Server 2005 and Oracle 10g database systems host the dimension and metadata repositories.The major challenges to overcome are two fold, namely the optimization of the distributed query plans and the robust security of the system.[4]The evaluation of queries over the global schema is the hard factor for achieving a satisfactory improved performance in Datamart federations . The prototype will include optimization algorithms for decomposition of query and query planning similar to those developed in . Secondly, the transmission of the query results is the second major factor on the usefulness of the federated data ware approach. In case that one or some of the local Datamartsdont give any respond to a query request, the federated system should still evaluate the query result.

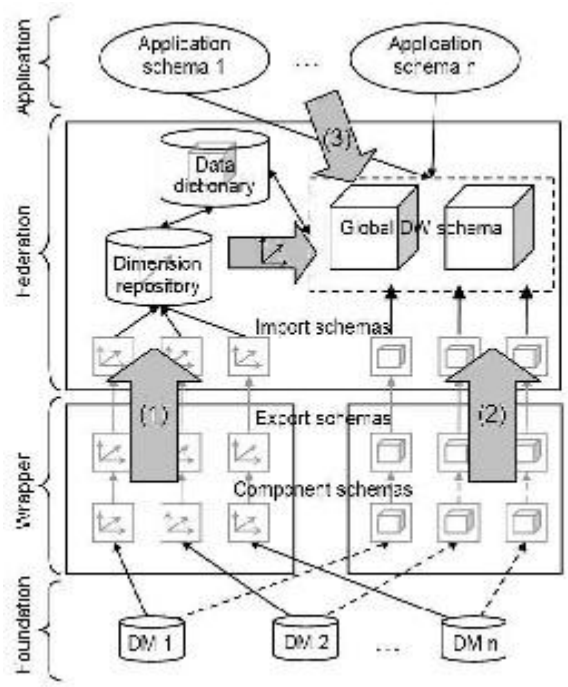

Fig.6 Federated Data Warehouse architecture

(B) Structure of Star Schema

The structure of the data warehouse is usually represented by a star schema, facts and dimensions, which are shown in the tables of physical data warehouse.[4] Fac1t table is located in the centre of the data warehouse and contains foreign keys for all dimension tables . It has many-one with the size of the table. In other words, every dimension records belongs to the thousands of metrics in the fact table. Figure shows the star schema structure.

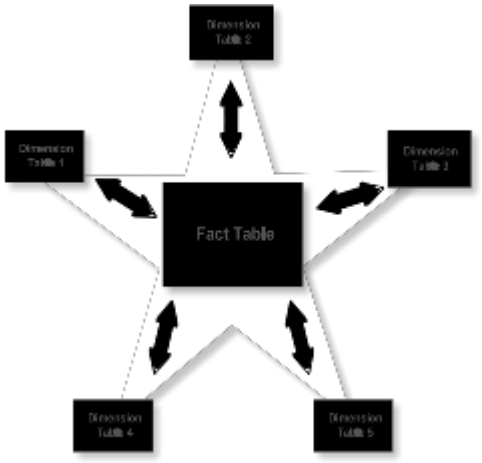

Fig.7 Structure of Star Schema

VII.CONCLUSION:

The database has many issues in build,designing and maintainance for information system. Since this time the database is used as storage for data and information and solve the problem about saving them safely. The dramatically increase in governments and companies transactions and operations meet by increase in government and companies data leads to increase in their data storage [4]

In these days government and companies are looking forward to huge repository to save data and information. For this the data warehouse is the best storage to keep their data. Star schema structures data by create fact table in centre surrounded by denormalized dimensional tables. This project creates tool by using JSP and MySQL to convert database into star schema for data warehouse. Star schema is made fast respond for quire and for better performance in data warehouse.[4]

References

[1]http://waset.org/publications/4000/the-comparison-of-anchor-and-star-schema-from-a-query-performance-perspective

[2]https://docs.oracle.com/cd/B10500_01/server.920/a96520/concept.htm

[3]https://en.wikipedia.org/wiki/Star_schema

[4]http://www.ijser.org/paper/Convert-Database-Structure-into-Star-Schema-Structure-for-Data-Warehouse.html

[5]Inmon, W. H., (1996), Building the Data Warehouse, Second Edition, John Wiley & Sons, Inc.

[6]Inmon, W.H., Building the Data Warehouse. John Wiley, 1992