

Conceptual Study on Data Marts – A Building Block of Data Warehouse

Pratibha Maurya

Assistant Professor, Amity University, India

Corresponding Author: Pratibha Maurya

ABSTRACT

The data mart is a newly evolving area, which is to be used. Its emphasis and ease promise the rapid deployment of decision support ability with the rapid return on investment required by the pace of contemporary business. The origin of data mart is from the data warehouse concept. The opportunity has become highly engrossed — concentrating on a single subject area rather than the complete enterprise. This paper presents a brief impression of data marts.

KEYWORDS: Data Mart, Data Warehouse, Dependent data mart, Independent data mart, Star Schema, Snowflake Schema, Starflake Schema

Date of Submission: 06-04-2018

Date of acceptance: 21-04-2018

I. INTRODUCTION

Data Mart is a subcategory of the data source, usually oriented to a definite purpose or major data theme that may be distributed to support business needs [1]. The conception of a data mart can be put on to any data whether they are operating data, evaluation data, spatial data, or metadata. It is used to answer very precise questions for a specific group of data users such as organizational divisions of marketing, sales, operations, collections and others. Finance has their data mart, marketing has theirs, and sales have theirs and so on.

The discrete departments own the hardware, software, data and programs that establish the data mart. The rights of possession allow the departments to sidestep any means of control or discipline that might synchronize the data found in the dissimilar departments. Every department has its own understanding of what a data mart should look like; each department's data mart is peculiar to, and specific to its individual needs.

II. DATA MART VS DATA WAREHOUSE

"...The data warehouse is nothing more than the union of all the data marts..." [2]. The data mart and data warehouse both tend to imply the existence of the other in specific form. But limited people agree that the design of a data mart inclines to start from a scrutiny of user requirements and that a data warehouse tends to begin from an investigation of what data previously exists and how it can be composed in such a way that the data can later be used. A data warehouse is a principal aggregation of data (which can be dispersed physically) and a data mart is a data repository that is customarily derived from a data warehouse and that highlights ease of access and usability for a specific designed purpose.

In overall, a data warehouse tends to be a premeditated but somewhat unfinished concept although a data mart tends to be strategic and intended at meeting an immediate need.

III. ADVANTAGES OF DATA MART OVER DATA WAREHOUSE

Data marts are exclusively useful to make access to exact frequently access data very ease. It can give a collective picture or a certain aspect in the business by a particular group of users. Since data marts are smaller compared to a full data warehouse, reaction time could be reduced and the cost of imbedding could also be less expensive.

The most distinguished difference of a data mart from a data warehouse is that the data mart is shaped based on a very precise and predefined purpose and need for a grouping of certain data. A data mart is configured, such that it makes access to pertinent information in a specific area very easy and fast.

IV. REASONS FOR CREATING A DATA MART

- Easy access to frequently needed data
- Creates collective view by a group of users
- Improves end-user response time
- Ease of creation
- Lower cost than implementing a full Data warehouse
- Potential users are more clearly defined than in a full Data warehouse [3].

Data marts have benefits over data warehouses. For example, the data mart emphasizes on data legibility: The business team receives only the data it needs, in a form that matches its collective understanding of the business. The data mart makes full use of today's LAN-based client/server technologies, assimilates with the information workers' toolset, and rides the price-performance curve of those technologies. A data mart offers a homogeneous population of information workers with similar business prototypes, business vocabularies, and responsibilities.

V. DESIGN SCHEMA

Data marts can be designed with star schema, snowflake schema or starflake schema. The star schema is the utmost simple of all the styles associated to data mart and data warehousing. It contains only of few fact tables. The snowflake schema is a deviation of the star schema and the storing method is of multidimensional nature. The starflake schema is a fusion mixture of together the star and snowflake schemas.

5.1 Star Schema

In the star schema plan, a single object explicitly the fact table sits in the mid and is radically linked to other surrounding objects called dimension lookup tables like a star. Each dimension is represented as a single table. The primary key in each dimension table is related to a foreign key in the fact table. [4]

Suppose our data warehouse retains store sales data, and the different dimensions are time, store, product, and customer.

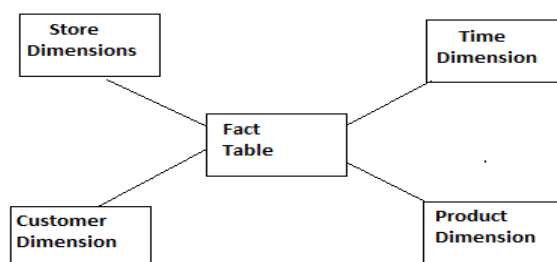


Fig 1: Sample star schema

The lines between two tables specify that there is a primary key / foreign key relationship amongst the two tables. Note that different dimensions are not related to one another. All events in the fact table are associated to all the dimensions that fact table is linked to. In other words, they all have the equivalent level of granularity.

A star schema can be simple or composite. A simple star contains of one fact table while a complex star may have more than one fact table.

5.2 Snowflake Schema

A snowflake schema is a logical prearrangement of tables in a multidimensional database such that the entity relationship diagram looks like a snowflake in shape. It is represented by central fact tables, which are connected to multiple dimensions. It is an addition to the star schema, where every point of the star shatters into more points. In a star schema, every dimension is represented by a single dimensional table, whereas in a snowflake schema, that dimensional table is regularized into multiple lookup tables, each demonstrating a level in the dimensional hierarchy. [5]

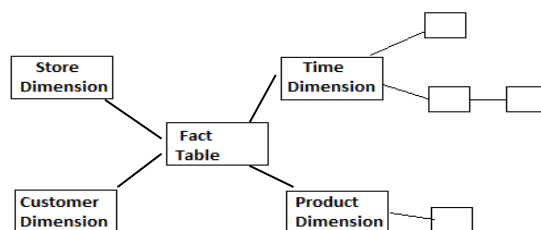


Fig 2: Sample snowflake schema

For example, the Time Dimension that consists of two different hierarchies:

1. Year → Month → Day
2. Week → Day

This gives four lookup tables in a snowflake schema: The first one is a lookup table for year, second lookup table for month, third lookup table for week, and the fourth lookup table for day. The Year is connected to Month, which is then connected to Day. Week is only connected to Day.

The main benefit of the snowflake schema is the enhancement in query performance due to reduced disk storage requirements and joining smaller lookup tables. The main hindrance of the snowflake schema is the supplementary maintenance efforts desired due to the increase number of lookup tables.

5.3 Starflake Schema

It is a mixture schema containing properties from both star and snowflake schema.

VI. TYPES OF DATA MART

There are two elementary types of data marts: dependent and independent. The categorization is based principally on the data source that feeds the data mart. Dependent data marts draw data from a principal data warehouse that has already been created. The Independent data marts, in contrast, are separate systems built by depiction data directly from operational or external sources of data, or both.

The main change between independent and dependent data marts is how the data mart is populated. This step is known as the Extraction-Transformation-and Loading (ETL) process, which involves moving data from operational systems, filtering it, and loading it into the data mart.

With dependent data marts, this process is slightly simplified because structured and summarized (clean) data has previously been loaded into the central data warehouse. The ETL process for dependent data marts is generally a process of categorizing the right subset of data applicable to the selected data mart subject and moving a copy of it, maybe in a summarized form. The Independent data marts deal with all the features of ETL process, similarly as a central data warehouse does. The number of sources is likely to be less and the amount of data related with the data mart is less than the warehouse, given focus on a single subject.

The motivations behind the creation of these two types of data marts are also typically different. Dependent data marts are usually built to accomplish improved presentation and availability, better control, and minor telecommunication costs causing from local access of data pertinent to a specific department. The formation of independent data marts is usually driven by the need to have answer within a shorter time.

VII. DRAWBACKS OF DATA MART

Data marts overvalue the capability and performance of LAN-based data-management tools. For example, flat-file DBMSs and small-scale RDBMSs do not have the handling power or the facilities to neither support a high-performance query environment, nor can you connect them to production data sources in a way that permits effective and timely large-scale extracts. Data marts also disregard the technical limits of mainframe-based production systems in providing extract processing time and power. It also neglects the secondary systems and network management clutter that is created, when numerous data marts demand substantial data extractions from multiple production systems.

Data marts report the decision-support requirements of merely small companies with insufficient knowledge workers, single markets, and simple product lines. On its own, data mart cannot meet the requirements of international enterprises with many distinctive knowledge edge worker communities, numerous products and markets, and continuous reorganization in reply to market conditions.

VIII. CONCLUSION

A fruitful data mart strategy can alleviate the risk, limit the expense, and decrease the time required to provide data warehouse functionality. As it is scalable, the data mart can work well for organizations of any size and level of difficulty. While it is correct that the data mart significantly reduces the risk related with building a decision support system, it however requires substantial skill and knowledge to properly implement one. The ability of constructing the data mart is still beginning and experienced practitioners remain few in number. The fastest way for an ambitious data mart designer to acquire the desired skills is to study the lessons learned by knowledgeable professionals.

REFERENCES

- [1]. Talus, *A Practical Guide to Building Data Marts* ftp://ftp.wiley.com/public/computer_books/updates/guide.pdf
- [2]. K. Drewek, *Data Warehousing: Similarities and Differences of Inmon and Kimball*, <http://www.b-eye-network.com/view/743>
- [3]. R. Chhabra et al, *Data Mart Designing and Integration Approaches International Journal of Computer Science and Mobile Computing, Vol.3 Issue.4*, April- 2014, pg. 74-79.
- [4]. <https://www.1keydata.com/datawarehousing/star-schema.html>
- [5]. <https://www.1keydata.com/datawarehousing/snowflake-schema.html>.
- [6]. Douglas Hackney, *Understanding and Implementing Successful Data Marts* (Reading, MA: Addison-Wesley, 1997).
- [7]. H. Jiawei and K. Micheline *Data mining concepts and techniques*, Morgan Kaufmann, 2001.
- [8]. Ponniah Paulraj, *Data warehousing fundamentals: a comprehensive guide for IT professionals*, A Wiley-interscience publication, John Wiley and Sons INC.2001
- [9]. www.dwinfocenter.org.
- [10]. www.sciencedirect.com

Pratibha Maurya." Conceptual Study On Data Marts – A Building Block Of Data Warehouse." International Journal of Computational Engineering Research (IJCER), vol. 08, no. 04, 2018, pp. 30-33.