



Role of Metadata in Data Warehousing for Effective Business Support

SHIV KUMAR GUPTA¹ and RITU VIJAY²

¹School of Engineering and Technology, Jaipur National University, Jaipur (India).

²Department of Computer Science and Electronics, Bansthali University, Rajasthan (India).

*Corresponding author: E-mail: skgupta50@yahoo.com

(Received: September 29, 2011; Accepted: October 10, 2011)

ABSTRACT

Today large organizations are being served by different types of area processing and information systems. It is important to create an integrated repository of what these system contain and do it in order to use them collectively and effectively. The repository contains metadata of source systems effectively. The repository contains metadata of source systems data warehouse and also business data. Metadata, usually called data about data, is an important part. Metadata is supposed to be a helping hand to all co-workers in an organization that work directly or indirectly with the data warehouse. The main purpose of this paper is to determine through feedback from the business end user.

Key words: Metadata, DSS, DW, ETL, OLAP.

INTRODUCTION

In recent years data warehouse (DW) has emerged as a powerful technology. A data warehouse is a relational database that is designed for query and analysis rather than transaction processing. It usually contains historical data derived from transaction data, but it can include data from the other sources. It separates analysis workload from transaction workload and enables an organization to consolidate data from several sources.

In addition to relational database, a data warehouse environment includes extraction, transportation and loading (ETL) solution, an online analytical processing (OLAP) engine, client analysis

tools and other applications that manage the process of gathering data and delivers it to business user.

According to W. H. Immon, data warehouse is a subject oriented, integrated, time variant, non volatile collection of data in support of management's decisions.

According to Ralph Kimball ("The Data warehouse Tool kit") a data warehouse is a copy of transactions data specially structured for query and analysis". That is, users can access the data, as they want for the analysis by querying, reporting, analysis, data mining and visualization of warehouse information. "A data warehouse combines various data source into a single source for end user access.

End user can perform adhoc querying, reporting, analysis, data mining and visualization of warehouse information. The goal of data warehouse is to establish a data repository that makes operational data accessible in a form that is readily acceptable for decision support and other applications". (Efraim Turban, Jay E. Aronson 2001). The major technologies related to data warehousing that are used for decision support are online analytical processing (OLAP), data mining and data visualization.

OLAP

On line analytical processing is "a category of software technology that enables analysts, managers and executives to gain insight into data through fast consistent, interactive access to a wide variety of possible views of information that has been transformed new data to reflect real dimensionality of the enterprise as understood by the user" (Inmon, 1992) that is the Technology facilitates the multi dimensional analysis of data, which in turn useful in decision support system.

Data mining

Data mining is the term used to describe knowledge discovery in databases. Sometimes we have extract data from large databases, which can have storage of data from several years (data warehouse. An end user when has little or no programming shell asks some adhoc queries and gets the required answers data) quickly. (Ebraim Turban, Day E. Aronson 2001).

Data visualization

It consists of pictures for analysis and decision making that includes digital images, geographical information systems, graphical user interfaces, multi dimensions, tables and graphs, virtual reality, three dimensional presentations, and animation. (Efraim Turban, Jay E. Aronson 2001). There are four separate distinct components (from the data ware tool kit by Ralph Kimball) to be considered in the data warehousing.

- ' Operational sources system
- ' Data staging area
- ' Data presentation area
- ' Data access tools

Data presentation area is a where data is

organized stored and made available for direct querying by the end users.

Metadata is commonly understood as any information needed in information technology in order to analyze, design, build, implement and then use computer systems. In the case of information systems, metadata particularly facilitates managing, querying consistent use and understanding of data.

Many recent efforts within both academic and industrial community have concentrated on issues related to metadata. The generation, storage, and management of metadata promise to better support the exploitation of the huge amount of data available in every conceivable form. Metadata is one important concept that plays a crucial role in the data ware housing environment all the metadata for the data warehousing environment is managed in the data presentation area which consists mainly of data access tools.

Metadata

- ' It is directly to help DSS analysis to locate the contents of DW.
- ' It is a guide to the mapping of the data as the data is transformed from the operational environment to the DW environment.
- ' It is guide to the algorithms used for summarization between the current detailed data and lightly summarized data in the highly summarized data.

Metadata is referred to as "technical" or "business focused". (DW management hand book – Richard J. Kachur).

Technical Metadata Cover the data movement, transportation aggregation and the presentation of the data.

Business Metadata focuses on what the information means to the business. Business metadata define information about the business and what this data means.

The end user or DSS analyst (Decision support system analyst) is the one who unlocks the secrets of information that are held within the bits and bytes of the data warehouse. i.e. the objective

of the end user is the information processing (WH Inmon 1994).

End - user's metadata serves a business purpose, it translates a cryptic name code that represents a data agreement in to the meaningful description of the data element so that end users can recognize and use the data. For example, metadata would clarify that the element "ACCICD" represents "account code for small business" (<http://www.gantthead.com/>).

The end user metadata (business metadata) contains:-

- ' Physical structure
- ' Information about data source
- ' Mapping of the source to the data warehouse information about data transportation.
- ' Alias information, status and volumetric information.

Observation and several open problems of metadata

Data should be accurate and timely is the key for any successful DWH. Some of the factors that lead to the poor quality of data are –

- ' Problems in data acquisition.
- ' Data coming from different sources may not integrate properly.
- ' Redundant, inaccurate data continues to grow in the data warehouse.
- ' Ineffective data update process.

Poor data quality such as inaccurate, conflicting, missing information leads to erroneous and costly business assumptions. Lack of timely information also results in slow reaction to changes in the business climate, contributing to cost opportunities and reduce sales. In dealing with this problem data warehousing projects invest 60 to 85percent of their total project time and resources. Without this effort, doubt over the validity and the accuracy of the data stored in the DWH often causes projects to fail. Credibility of data is essential. Not only must the data be quality before it enters the DWH, but it also must be audited to prove correct source were selected when doubt creeps into the minds of our business users(Richard Kachur).

According to Richard Kachur, in the data

warehouse environment, data quality, management must focus on implementing a process by which credible, accurate and timely information is extracted from preferred subject of internal and external sources.

Purpose and remedial measures

Primary consumers for the contents of the DW are the end user or DSS analysts within the organizations. The efficient use of DW for management decisions making depends on the effective use /retrieval of the data warehouse information by the end user or DSS analysts or any other knowledge workers from the organization.

As the problem area, we think in dealing with the problem/error in the data warehousing environment pays a key role.

The data staging area, data presentation area and data access tools contain their own metadata to be maintained. All the phases of the data warehouse consist of meta data to be stored. Metadata is centralized and managed as a metadata repository/directory from the data access layer.

On investigation of the role of metadata end user perspective, which lead to the following questions? "What will be the impact of metadata on the data quality in the warehousing environment?"

The above question indicates that metadata plays a special role in maintaining the data quality. The main purpose of this determine through feedback from the end users and the other knowledge workers, if the metadata is needed for achieving the data quality.

This is about the impact of metadata on the data quality. The problem area is huge as the data warehouse consists of three different phases and each phase has metadata management. Data retrieval / access phase, end user accesses the data warehouse for the data. In order to achieve the result, we have to review three indicators of data quality those are data accuracy, consistency and timeliness, for how these concepts depend on metadata. These data retrieval errors are mapped to the data quality, have come up the with the following hypothesis in the form of table shown below.

Table 1:

Metadeta/Data Quality	Data Accuracy	Data Consistency	DATA Timeliness
Basic Components	X	X	X
Mapping Informations	X	X	-
Extract History	X	X	-
Versioning	X	-	X

In the above table shows the dependency of data quality indicators on the particular metadata. Here 'X' denotes dependency and '-' denotes that particular situation is "not applicable".

The case study is best suited to study the complex situation in a systematic way and we can get the information needed by different levels of employees from the organization.

CONCLUSIONS

We can summarized as detail above the data quality indicator depends upon the all the metadata basic components, mapping information,

extract history, and versioning. Most end users are not using the related tools; organizations make them learn by creating interest in them.

The area discussed in the paper also contains a number of research issues which will need to be dealt with before commercial information system can leverage the maximum advantage for metadata and associate technology.

ACKNOWLEDGEMENTS

I am thankful to Dr. Ritu Vijay, Reader & H.O.D., AIM & ACT, Bansthali University, Rajasthan, India.

REFERENCES

1. William H. Inmon, Building the Data Warehouse, Fourth Edition, Wiley Publication, (1992).
2. W. H. Inmon and C. Kelley, "RdbMS: Developing the Data Warehouse", QED Publishing Group, Boston, Massachusetts (1993).
3. Ralph Kimball, Margy Ross. "The data warehouse toolkit: the complete guide to dimensional modelling", Wiley Publication, 2nd Edition, ISBN: 9780470479575, (2002).
4. Richard J Kachur, "The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling", Wiley Publication, ISBN-10: 0471200247, Edition: 2, April 26, 2002 (2000).
5. Meta Data Europe 99: Implementing, Managing and Integration MetaData, London U.K. March99. Technology Transfer Institute. <http://www.ttiuk.co.uk/events/5.cfm>.
6. Gartner Group. Repository Market Update 1999 Research Note 01 June 99, M-08-3721.
7. Object Management Group (OMG). Common Warehouse Metadata Interchange – Request for Proposal ,1998. OMG Document ad/98-09-02.
8. M. Staudt, A.Vaduva & T.Vetterli., "Metadata management and data warehousing", Technical R, Report, 21 (1999).
9. Brackett, Michael H, "The Data warehouse challenge", Taming data chaos, Newyork: Wiley, (1996).
10. Ecierson, Wayne W, "Data quality and the Bottom Line", TDWI (online) Retrieved September 13, 2005 (2002).
11. White paper on "Maintaining Data Quantity" November 2005, [http:// www.infogain.com](http://www.infogain.com)