P- ISSN 1991-8941 E-ISSN 2706-6703 *Journal of University of Anbar for Pure Science (JUAPS)* Open Access
2012,(6), (2) :95-100

The Role of Metadata for Effective Data Warehouse

Murtadha M. Hamad **Alaa Abdulqahar Jihad**

University of Anbar - College of computer

ARTICLE INFO

Received: 00 / 00 /00
Accepted: 00 / 00 /00
Available online: 9/12/2012
DOI: 10.37652/juaps.2012.63366

Keywords:
Metadata,
Data Warehouse,
Data Cleaning, Decision Support System.

ABSTRACT

Metadata efficient method for managing Data Warehouse (DW). It is also an effective tool in reducing the time or speed to answer queries. In addition, it achieved capabilities of the integration and standardization, thus lead to faster, clear and accurate decision-making in the right time. This paper provides the definition of metadata concept, and using metadata in Data Cleaning; which it identify the sources, types of fields, and choose the appropriate algorithm. In addition, useful in Decision Support System (DSS); which it improve efficiency of analysis and reduces response time of query.

Introduction

Metadata is data about data; or the description of the structure, content, keys, indexes, etc., of data [1].

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource [2]. In other words, metadata imposes a structure on unstructured information (i.e. documents, maps, audiovisual material, etc.) and adds more structure to already structured (i.e. database) information. The metadata structure is then exploitable for the purposes of finding information, administration, recordkeeping and preservation [3].

Background

There has been a substantial amount of work on the general topic of Data Warehouse. We discuss generally here only the works that propose metadata for the Data Warehouse and Data Warehouse modeling.

In [4] describe a metadata approach for Data Warehouse security, but do not go beyond technical metadata plus business-oriented string labels and descriptions of attribute and table names.

In [5] introduces Data Warehouse architecture with eight layers including a metadata layer. These layers represent the overall structure of data, communication, processing and presentation that exists for end user computing within the enterprise.

In [6] the logical architecture is independent from application and front-end tools. The physical architectures are a mapping of the logical architecture to multidimensional database management system (MDBMS) and relational DBMS (RDBMS).

In [7] introduced different extended relational concepts to model metadata for data warehousing. The differences of the models show a huge advantage of the extended relational model.

Metadata Classifications

There are many kinds of metadata in a Data Warehouse system.

In [8], metadata is classified based on the Data Warehouse architecture layers as follow:

1. Metadata associated with data loading and transformation: It describes the source data and any changes that were made to the data.
2. Metadata associated with data management: It defines the data store in the Data Warehouse. Every object in the database needs to be described including the data in each table, index, and view, and any associated constraints. This information is held in the DBMS system catalog; however, there are additional requirements for the purposes of the warehouse.
3. Metadata used by the query manager to generate an appropriate query: The query manager generates additional metadata about the queries that are run, which can be used to generate a history on all the queries and a query profile for each user, group of users, or the Data Warehouse.

* Corresponding author at: University of Anbar - College of computer-.E-mail address: mortadha61@yahoo.com

The other classification divides metadata into technical metadata, business metadata and information navigator metadata [9]:

1. Technical metadata primarily supports technical staff that must implement and deploy the Data Warehouse .The information contained within the technical directory is compatible with this kind of audience and contains the term and definition of metadata, exactly as they appear in operational databases.
2. The business metadata primarily supports business end users who do not have a technical background, and cannot use the technical metadata to determine what information is stored inside the Data Warehouse.
3. The information navigator metadata is a facility that allows users to browse through both the business metadata and the data inside the Data Warehouse .

Moreover, the metadata can be considered as two classes, namely static and dynamic [10].

1. Static metadata: This kind of metadata is used to document or browse in this system. E.g., metadata of a dimension. The content of this metadata is fixed in the Data Warehouse.
2. Dynamic metadata: vice versa to static metadata, dynamic metadata is metadata that can be generated and maintained in run time. For instance, metadata of a new frequent access query .

The Role of Metadata in the Data Warehouse

Due to the increasing complexity of Data Warehouses, a centralized and declarative management of metadata is essential for Data Warehouse administration, maintenance and usage [11].

The advantages of managing metadata [12]:

- Consistency of definitions: One department refers to "revenues," another to "sales." Are they talking about the same activity? One subsidiary unit talks about "customers," another about "users" or "clients." Are these different classifications or different terms for the same classification? Effective Meta data management can ensure that the same data "language" applies throughout the organization.
- Clarity of relationships: Meta data management illuminates the associations and interactions among all components of the warehouse environment: business rules, tables, columns, transformations, and user views of the data, to name a few. By clarifying relationships throughout the Data Warehouse environment, managed Meta data enables warehouse managers and knowledge workers to see the bigger picture—to fully understand the meanings of the data assets, and to accurately predict and manage the impact of changes to the environment.
- Availability of information: Meta data exists "behind the scenes," revealing the origin of data, who defined it, when it was modified, and much more. Traditionally hidden, Meta data must now be made visible to company knowledge workers on demand.

Metadata Creation and Management

Metadata, in a DW project, play a key role. Nevertheless, metadata management is a big challenge to many DW projects, mainly because there exists much heterogeneity among tools and products for creating and managing metadata in a Data Warehousing environment. That is because there is not in the industry a unified standard for metadata definition and interchange.

In spite of these obstacles, due to the importance of metadata in an analytical environment, some works have been developed to support the metadata management in DW. However, none of the woks analyzed integrates metadata creation and management to the development methodology [13].

Metadata and Data Cleaning

In order to supply a decisional database, metadata is needed to enable the communication between various function areas of the warehouse and an ETL tool (Extraction, Transformation, and Load) is needed to define the warehousing process. The developers use a mapping guideline to specify the ETL tool with the mapping expression of each attribute [14], see table (1) examples for the use of metadata.

Resource Discovery

Metadata serves the same functions in resource discovery as good cataloging does by [2]:

1. Allowing resources to be found by relevant criteria;
2. Identifying resources;
3. Bringing similar resources together;
4. Distinguishing dissimilar resources; and
5. Giving location information.

In Figure (1) using Metadata in Data Cleaning, Metadata benefit data cleaning process in two directions:

1. The sources of data (dirty data)
   - Scan the sources and knowledge of characteristics
   - Apply the ETL process to extract data
   - Extract of data types in the sources

2. The target data (clean data)
   - Assist the cleaning process for ETL and transform data
   - Discovery of data types in the target data to be cleaning and have appropriate transformation
   - Identification of possible values in the target data to determine the anomalous values and errors in the data source
   - Identify the appropriate formula for the target data

Metadata and DSS

The metadata warehouse is a resident resource that provides the metadata and structure associated with the information repositories.

After the keywords are submitted, the metadata warehouse is consulted to determine in what context the keywords are found in the information repositories. It is important to note that this resource only provides information about the contents of the information repositories and not their actual contents [15].

The metadata is also required by the query manager to generate appropriate queries [16].

Show in figure (2) metadata is the first thing the DSS analyst looks at in planning how to perform informational/analytical processing [1]. For example, the IT professional uses metadata on a casual basis when he uses the operational database; the DSS analyst uses metadata regularly and as the first step of an analysis when he uses the data warehouse.

Discussion

After the practical application and the work of metadata, and use in the process of the cleaning, found it is useful in improving the efficiency of processing in speed and they determine the required properties for the table to be processed accurately. The tables (2, 3, 4) is proposed Metadata for files in Data Warehouse.

In addition, metadata will help determine the type of specific data to select the appropriate algorithm, when does not use of metadata will have to work program for each field and the program is dedicated to this field only. Unlike, if the metadata is used, the program will process the different fields.

Metadata reduce response time of query. Metadata is directory of the files, the fields and types, existence of metadata facilitates the implementation of query on the required information (see figure (3)), and the missing of metadata will reduce the efficiency of implementation of the query and the results.

Conclusion

After the practical application and use of metadata, showing the importance and benefits of the metadata of the organization in data cleaning and DSS.

The metadata provides an inventory of data assets, helps determine and maintain the value of data, helps in determine the reliability and currency of data, supports decision making, helps keep data accurate and helps verify accuracy to support good decision making.

References

1. W. H. Inmon, Building the Data Warehouse, Third Edition, John Wiley & Sons, 2002.
2. Understanding Metadata, NISO Press, USA, National Information Standards Organization, ISBN: 1-880124-62-9, 2001.
3. Metadata Resources Guide, Information Management, Government of Alberta, ISBN 0-7785-3109-, May 2004.
4. Katic, N.; Quirchmayr, G.; Schiefer, J.; Stolba, M.; Tjoa, A M.: A Prototype Model for Data Warehouse Security Based on Metadata. Proceedings DEXA 98.
5. Ken Orr, Data Warehousing Technology, The Ken Orr Institute, A white paper, 1996.
6. M. Wu, A. P. Buchmann, Research Issues in Data Warehousing, BTW 1997: 61-82.
7. O. Mangisengi, A M. Tjoa, R. R. Wagner, Metadata for Data Warehouses Using Extended Relational Models Proc. of third IEEE Computer Society Metadata Conference, April 1999.
8. Thomas M. Connolly and Carolyn Begg, Database Systems: A Practical Approach to Design, Implementation, and Management, 3rd Edition, Addison-Wesley, ISBN 0-201-70857-4., 2004.

9. Que, The Official Client/Server Computing Guide to Data Warehousing, Que Books, 1997.
10. Thanh N. Huynh, Oscar Mangisengi, and A Min Tjoa, Metadata for Object-Relational Data Warehouse, Vienna University of Technology, Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'2000) Stockholm, Sweden, 2000.
11. Thomas Stöhr, Robert Müller and Erhard Rahm, An Integrative and Uniform Model for Metadata Management in Data Warehousing Environments, Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'99), Heidelberg, Germany, 1999.
12. Ascential Software, Meta Data Management in the Data Warehouse Environment, U.S., 2001.
13. Liane Carneiro and Angelo Brayner, X-META: A Methodology for Data Warehouse Design with Metadata Management, University of Fortaleza, 2000.
14. Rami Rifaieh and Nabila Aïcha Benharkat, Query-based Data Warehousing Tool, DOLAP'02, November 8, 2002, McLean, Virginia, USA.
15. Isabel F. Cruz and Kimberly M. James, A User-Centered Interface for Querying Distributed Multimedia Databases, the National Science Foundation under CAREER Award IRI-9896052 and CISE Research Instrumentation Grant 9729878, 1999.
16. BERIL PINAR, A COMPARISON OF DATA WAREHOUSE DESIGN MODELS, A MASTER'S THESIS in Computer Engineering Atilim University, JANUARY 2005.

Table (1) Examples for the use of reorder metadata to address data quality problems

Problems	Metadata	Explanation
Illegal values	cardinality	e.g., cardinality (gender) = 2
	max, min	max, min should not be outside of permissible range
Missing values	null values	e.g., Age=NULL.
Duplicates	uniqueness	attribute cardinality = number of rows
	attribute values	sorting rows based on main attributes; if two rows equal then indicates duplicates

Table (2) Metadata of Sales table

Field_id	Field_name	Description	Type	Cardinality	Is it unique	Allow_null	Valid values	Min	Max
1	Sales_id	Sales ID	Num	-	Y	N			
2	Item_id	Item ID	Num	-	N	N			
3	Cust_id	Customer ID	Num	-	N	N			
4	Time_id	Time ID	Num	-	N	N			
5	Sales_amt	Sales amount	Num	-	N	N			
6	Total_cost	Total cost	Currency	-	N	N			

Table (3) Metadata of Customer table

Field_id	Field_name	Description	Type	Cardinality	Is it unique	Allow_null	Valid values	Min	Max
1	Cust_id	Customer ID	Num	-	Y	N			
2	Cust_name	Customer name	Character	-	N	N			
3	Birthday	Birthday	Date	-	N	Y			
4	Gender	Gender	Character	2	N	N			
5	Marriage_C	Marriage case	Character	4	N	Y			
6	Num_of_child	Number of child	Num	-	N	Y			
7	Blood_C	Blood class	Character	6	N	Y			

8	Salary	Salary	Currency	-	N	Y			
9	Phone	Phone	Num	-	N	N			
10	Street	Street	Character	-	N	N			
11	City	City	Character	-	N	N			
12	State	State	Character	-	N	N			

Table (4) Metadata for queries

Id-query	Name-query	Tables used	Fields_names	Conditions	Order by
1	Query1	Sales	Time_id, Sales_amt	Sales_amt> 50	Time_id
2	Query2	Sales, customer	Sales_id, Cust_id, Time_id, Sales_amt	City="Heet"	Sales_id
3	Query3	Sales	Sales_id, Time_id, Sales_amt, Total_cost	Sales_amt>25 AND Total_cost>$100	Sales_id

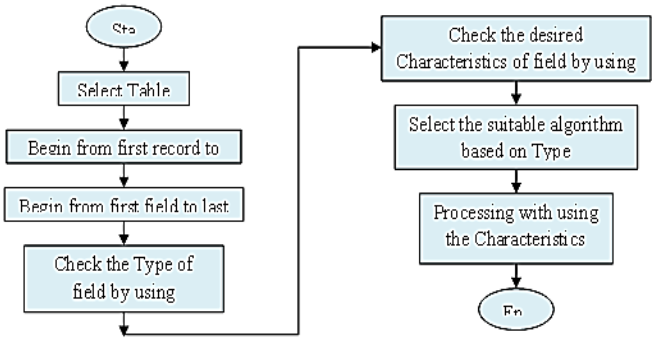

Figure (1) Using Metadata in Data

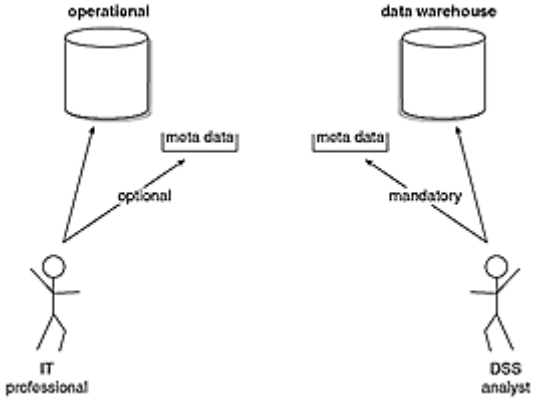

Figure (2) Using metadata in operational database and data warehouse.

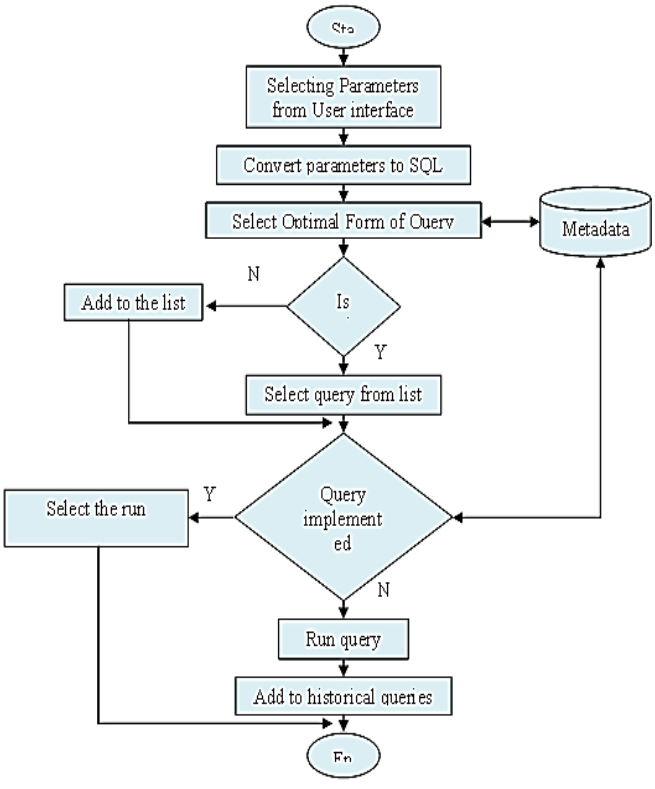

Figure (3) Flowchart of query solution

P- ISSN 1991-8941 E-ISSN 2706-6703 *Journal of University of Anbar for Pure Science (JUAPS)* Open Access
2012,(6), (2) :95-100

دور الميتاداتا لمستودع البيانات الفعّال

مرتضى محمد حمد علاء عبد القهار جهاد

E.mail: **mortadha61@yahoo.com**

الخلاصة:

الميتاداتا طريقة كفوءة لإدارة مستودع البيانات (DW). وهي اداة فعالة في تخفيض الوقت وتسريع الاجابة عن الاستفسارات. بالإضافة الى ذلك، تنجز قابليات التكامل وتوحيد المقاييس، هذا يؤدي الى تسريع اتخاذ القرار الدقيق والواضح في الوقت الصحيح. يزود هذا البحث تعريف مفهوم الميتاداتا، واستعمال هذا المفهوم في تنظيف البيانات، التي تميز المصادر، وأنواع الحقول، وتختار الخوارزمية الملائمة. بالإضافة الى ذلك، فائدة الميتاداتا في نظام مساندة القرار (DSS)، والتي تحسن كفاءة التحليل وتخفض وقت الاجابة عن الاستفسار.