

Multidimensional Analysis of Distributed Data Warehouse of Antiquity Information

Quanhong Sun, Qi Xu* and Qiaoqiao Li

¹Department of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou, 450011, China; ²Department of Electric Power, North China University of Water Resources and Electric Power, Zhengzhou, 450011, China

Abstract: Aiming at the problems in the management and sharing research of antique information, this paper discussed the necessity to build antique information distributed data warehouse. The data warehouse architecture and the multidimensional data model were put forward, by using online analysis processing (OLAP) the multidimensional analysis of antique distributed data warehouse was presented, and a multidimensional analysis process was given. By multidimensional analysis of the distributed data warehouse, the accessing and analyzing platform was provided for antique managers and researchers to provide strong support for the work of antique information management and sharing.

Keywords: Antiquity, distributed data warehouse, multidimensional analysis, OLAP.

1. INTRODUCTION

China is a country with an ancient civilization, which has a long cultural history to retain a large number of precious antiquity and accumulate vast amounts of historical antiquity data. But the management levels of collection antiquity are low, the means of protection are laggard. Although some Relics and Museology units have begun to implement antiquity digitization projects, but which basically limited to the local databases of relevant units, and many historical data are distributed in many different hardware and different databases, even there are many issues that are multiple data types, data deconcentration, different data definitions, complex and diverse storage format and so on in various museums, it is laborious and time-consuming to lookup historical data to make associated comparative analysis from different departments in all Relics and Museology units. Therefore, it is necessary to build a distributed data warehouse to effectively manage and use these historical data, which provides an analysis and decision-making platform for antiquity managers and researchers to realize the sharing of antiquity information by using online analysis processing systems (OLAP) to access and analyze antiquity information.

2. ANTIQUITY INFORMATION DISTRIBUTED DATA WAREHOUSE

2.1. Structure of Distributed Data Warehouse

W.H. Inmon who is honored as father of the data warehouse divided the distributed data warehouse into three types [1]:

a. Distributed data warehouse based on the architecture of local data warehouse and global data warehouse, business is done in different regions or different production systems, the local data warehouse provides and process the data on a remote site, and the global data warehouse provides integrated data in the entire business.

b. Data warehouse environment includes a large amount of data that are scattered on multiple processors. From a logical point of view, there is only one data warehouse, but from a physical point of view, there are many data warehouses that are well connected but stored in different processors. This situation can be called a technical distributed data warehouse.

c. Data warehouse environment is not build up in a coordinated way. First, a data warehouse is established, and then another independent data warehouse is built. This case may be called an independent evolution distributed data warehouse.

This paper adopts the first type, the typical structure is shown in Fig. (1).

2.1.1. Local Data Warehouse

Local data warehouse stores data that have the meaning of local decision, which extracts data from the local operation processing, and provides all services of the typically data warehouse except that the data is local. Local data warehouse comprises historical and integrated data from local sites, date and data structures may be different in different local data warehouses. Usually local data warehouses are built among different regions or different technologies groups.

2.1.2. Global Data Warehouse

Data of the global data warehouse involves the whole antiquity system, which is public, historical and integrated data

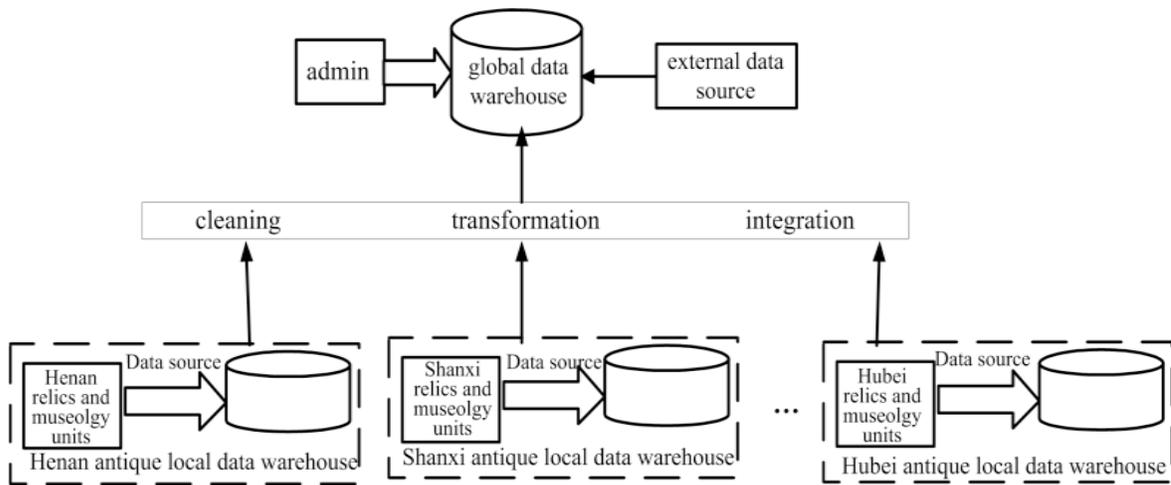


Fig. (1). Typical structure of distributed data warehouse.

within the whole antiquity system. Unlike local data warehouse, data of the global data warehouse is mainly from the local data warehouse and external data sources except for the actual operation system, which does not simply accumulate data of local data sources, but reorganizes, synthesizes and integrates data of the local data warehouse. The overlap and public data between the global data warehouses must be coordinated, which generally can not be distinguished when local data warehouse is built, so it is best to store the data in the global data warehouse.

2.1.3. Data Mode

Global data warehouse contains public and integrated data within the antiquity system that is the processing comprehensive data from local data warehouses spreading everywhere. How to map data from locally operating systems into the global data warehouse is a key technology to build a successful distributed data warehouse environment. Global data warehouse has a common data structure, which reflects all the public data in the enterprise. Because the different local data warehouse perhaps adopts different logical models and physical models and so on, so data mapping from each local site to global data warehouse is generally different. From the local data warehouse to the global data warehouse data warehouse, data mapping design is a gradual iteration process, which is not accurate in the beginning, but along with accumulation of user feedback knowledge, the mapping of local layer will gradually become more reasonable over time.

2.2. Data Sources

Antique information is collected from each of the relics and museology units, universities and research institutions that have their own business systems. From the point of view of global data warehouse, data mainly comes from interrelated data dispersed in different position, which generally refers to the mild integrated or highly integrated data. These data can be integrated based on the needs of decision-makers, such as comprehensive data by province, city, region. Data of each local data warehouse are mainly from the various units and regions, taking Henan antique system as an example, data of Henan antique data warehouse is mainly

from Henan Museum, Zhengzhou Museum, Nanyang Museum of Han Stone Gravings, cultural relics and archaeology institute of Henan Province, Cultural Relics Bureau of the prefectures, cities and counties, universities and colleges represented by Zhengzhou University and Henan University and so on. A part of global and local data sources are also external data, when the data is unified to load into a data warehouse and data marts, which needs to be converted, extracted, filtered, and cleaned.

2.3. Architecture of Antique Information Distributed Data Warehouse

Building antique information distributed data warehouse can provide a decision analysis platform for heritage information manager and decision-maker and a studying platform for personnel of archaeological research institutes, which can help them to make correct decisions in management and protection of antique information. According to the theory of distributed data warehouse discussed in [1], combining the characteristics of antique information, this paper designs a four-level distributed data warehouse architecture, which is shown in Fig. (2). From Fig. (2), the data warehouse from the national level to the regional level can be called global data warehouse relative to low-layer data warehouse, but the data warehouse from provincial level to county level can be called local data warehouse relative to upper-layer data warehouse. the bottom-up arrow means that all data is transmitted from the bottom upward, heritage data are gathered in higher-level data warehouse after cleaning, conversion, reproduction and integration. Horizontal arrows indicate that the data warehouse can extract the data from the antique database system at the same levels.

3. DESIGN OF ANTIQUE INFORMATION DISTRIBUTED DATA WAREHOUSE

3.1. Classification of Antiquity

Antiquity classification is a very important task, which is the needs of antique information management, research and protection. There are a lot of classifications about antiquity, which mainly include age classification, regional classification, morphological taxonomy, texture classification, func-

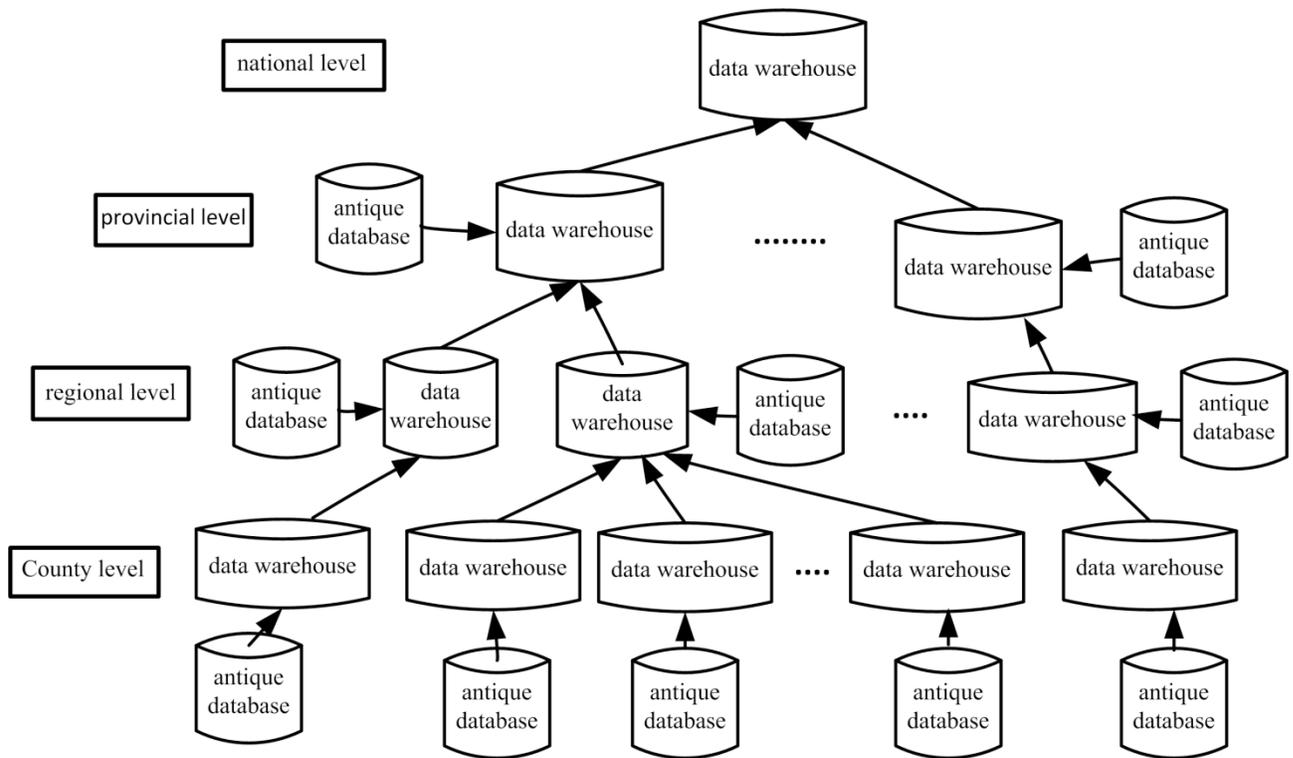


Fig. (2). Structure of antique information distributed data warehouse.

tion classification, properties (nature) classification, source classification and so on, the specific description of each category can be found in the introduction of the National Cultural Relics Bureau [2, 3].

3.2. Analysis of the Subject

An important feature of the data warehouse is subject-oriented to organize data. The subject is the framework and direction to constitute a data warehouse system, which is a standard to classify data at a higher level. Therefore, in the design process of the data warehouse, the analysis of subject is very important. According to the characteristics of information and cultural heritage classification, antique managers and professional researchers primarily concern with the following objects, which are antiquity, geographic area, collection units, texture, usage, complete residual degree, grade, age, nature, source, value and so on. The above objects are related to cultural relics, and so antiquity is identified as the subject of heritage data warehouse, the above objects are as dimensions related to artifacts to be studied to support decision-making needs of managers and researchers from different angles.

3.3. Data Warehouse Design Model

3.3.1. Local Data Warehouse Model

Currently there are two multidimensional logical data model commonly used that are the star schema and the snowflake pattern [4, 5]. Considering the characteristics of antique information data warehouse, this system adopts the star schema as the logic model of data warehouse. Because the above objects are related to artifacts, data warehouse information centers on antique, so the paper regards antiq-

uity as the subject to discuss the logical model of data warehouse. This model is shown in Fig. (3).

3.3.2. Design of Global Data Warehouse Model

Global data warehouse is a whole on the logic and can't show the distribution of data, but which is physically distributional and consists of a number of local data warehouses, the global data warehouse only slightly integrates data of local data warehouses, the detailed data come from the data source corresponding to source data table. The local data warehouses store meaningful data for local decision, which is extracted from the local data processing operations, and supports local decision. Global data warehouse stores data that is meaningful on a global sense, which extract data from local data warehouses, own local operational processing environment and head office, the model is as shown in Fig. (4).

Correspondence between global data warehouse and local data warehouse were achieved by source data in the global data warehouse. Its source data table (YSJB) is shown in Table 1, because the provincial and regional data warehouses are local data warehouse for the superstrata, which are global data warehouses for the substratum, so the provincial and regional layer data warehouse also have Table 1 as source table data.

Now, taking Henan data warehouse as an example, the global data warehouse data model is illustrated. the detailed usage data is stored in Henan local warehouses that can use the detailed data to analysis, synthesis, extract useful information for decision-making, recorded antique information of Henan local data warehouse is shown in Table 2. Because the Henan local data warehouse is provincial, which is a local

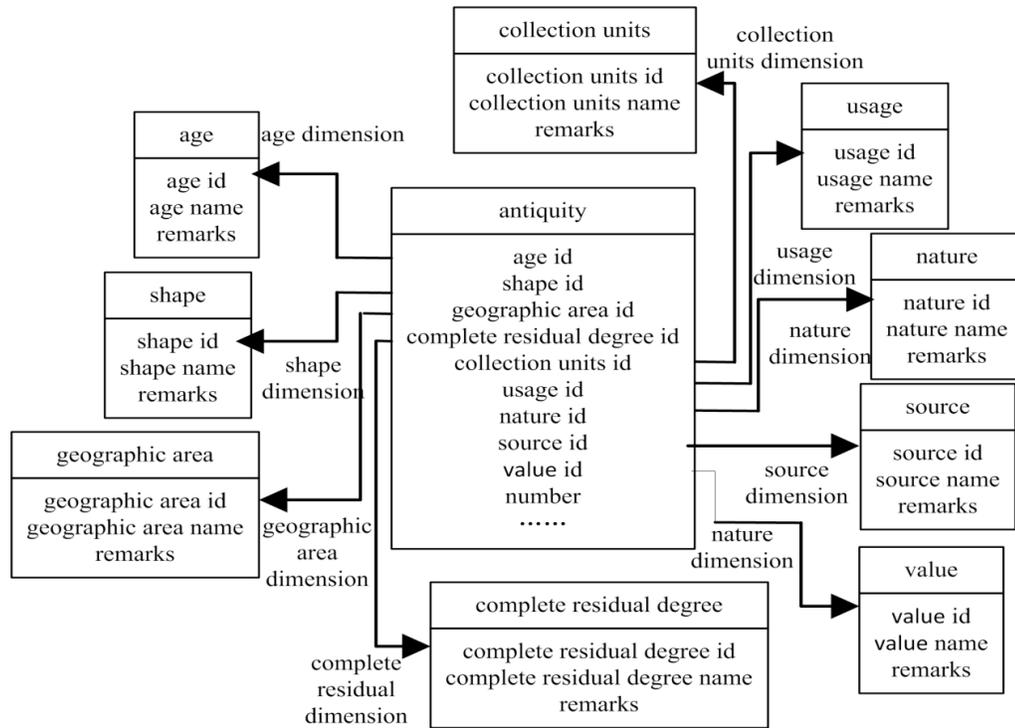


Fig. (3). Star model structure of local heritage information data warehouse.

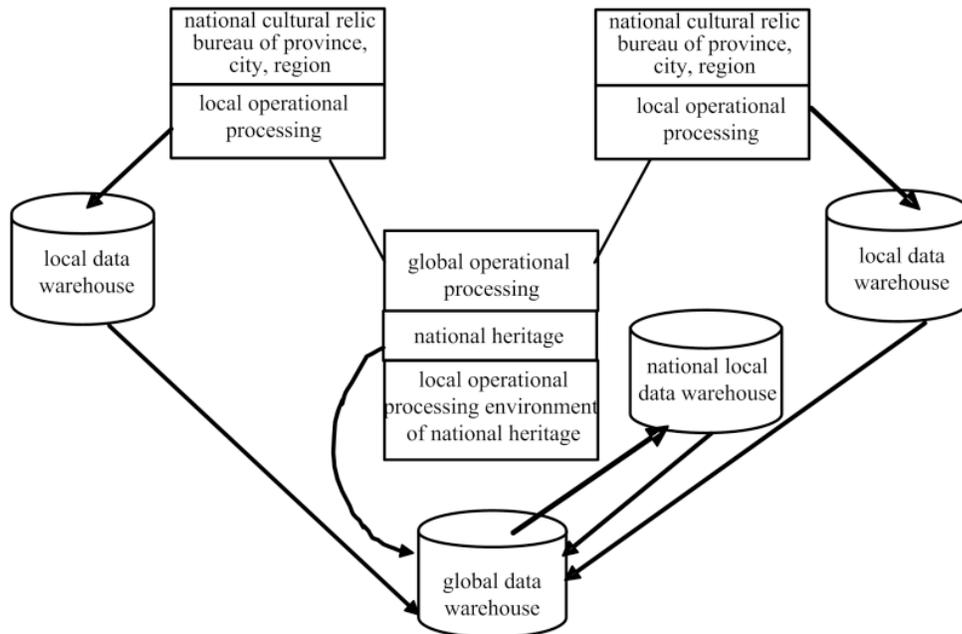


Fig. (4). Local and global data warehouse data warehouse data extraction.

Table 1. Source data table structure.

Age	Type	Source	Integrity	Degree
....
The Qing dynasty	immovable heritage	Henan	medium	two-level
The Tang dynasty	movableheritage	Shanxi	good	three-level
....

Table 2. Henan cultural relics information.

Type	Integrity	Source	Time	Degree
....
Bronzes	good	Nanyang	The Qing dynasty	two-level
jades	good	Zhumadian	The Tang dynasty	two-level
....

Table 3. Heritage information of the national data warehouse.

Field Name	Field Code	Type	Length	Unit
Source	SJLY	char	30	
the server name of Each unit	FWQMC	char	10	
Server IP	FWQIP	char	10	

data warehouse for national data warehouse and a global data warehouse for county and regional data warehouse, therefore the data granularity recorded in Henan data warehouse is not minimum, a lower comprehensive data can be looked over in various regions or counties by information source of the provincial local data warehouse.

Global data warehouse synthesizes various detailed data of the local data warehouse, which can analysis and synthesis detailed data of the local data warehouse to extract the global information, the heritage information of global data warehouse is shown in Table 3.

According to Morphologyclassification, movable heritages can be divided into bronze, jade and so on, immovable cultural relics can be divided into garden art, the Great Wall and so on. Thus, the detail information of local data warehouse is easily integrated into global data warehouse data for summarizing, for example, the state-level global data warehouse can analysis movable and immovable antiquity based on the specific antique information of each local data warehouse, and check thespecific information by the source table.

4. MULTI-DIMENSIONAL ANALYSIS OF DISTRIBUTED DATA WAREHOUSE

OLAP [6, 7] (Online Analysis Processing) can Online access, process and analysis data for specific topics, which is the main application of data warehouse to support sophisticated analysis operations. OLAP can focus on decision support and provide intuitive query results to let the analysis staff, managers or executives access information quickly, congruously, interactively from multiple perspectives to gain a deeper understanding of the data. In this paper OLAP can be understood as a collection of multidimensional analysis tools.

4.1. Multidimensional Data Model

Multidimensional cube data model can also be called multidimensional data model. A multidimensional cube is a logical data organization form and a data object to achieve

OLAP analysis. In antique data warehouse, from Fig. (3), the data sources used in artifacts data model consist of age, morphology, geographic area, nature, purpose, completeresidual extent, source, value, unit, rank, etc. These data sources correspond to the appropriate dimension tables, which connect to the antique fact table through its code to generate the artifacts cube data model [8, 9].

4.2. Multidimensional Analysis Motion

OLAP mainly analyzes the data in the multidimensional cube by the motion of slicing, dicing to make that antiquity managers or researchers view data from a variety of dimensions, multiple sides and many data integration degree to find that the law behind the data to support managers and researchers to make decision [10].

4.3. Multidimensional Analysis Process Design

For heritage information distributed in different places, if low-level detailed data is just needed, the data can be only extracted from a national data warehouse (global data warehouse) for analysis, if the higher level detailed data is required, first the provincial data warehouse is found out by looking up source data table in national data warehouse, then the regional data warehouse is found out by looking up source data table in provincial data warehouse, just like that, looking up down, finally the distributed data warehouse can be carried out multidimensional analysis. When we are analyzing multidimensional data, analysis procedure is usually as follows. first the interested contents are selected by carving method, and then data is drilled (upturned, drilled down) to reach the appropriate level, finally the important data is selected by viewing data through rotating multi-angle to slice, if slicing data is too many, which is needed to divide into smaller data to make analysis and comparison. The whole analysis process is shown in Fig. (5).

Through multidimensional analysis of the distributed data warehouse, the different granularity information distributed in the local and global data warehouses can be checked to support heritage managers and researcher to make decision.

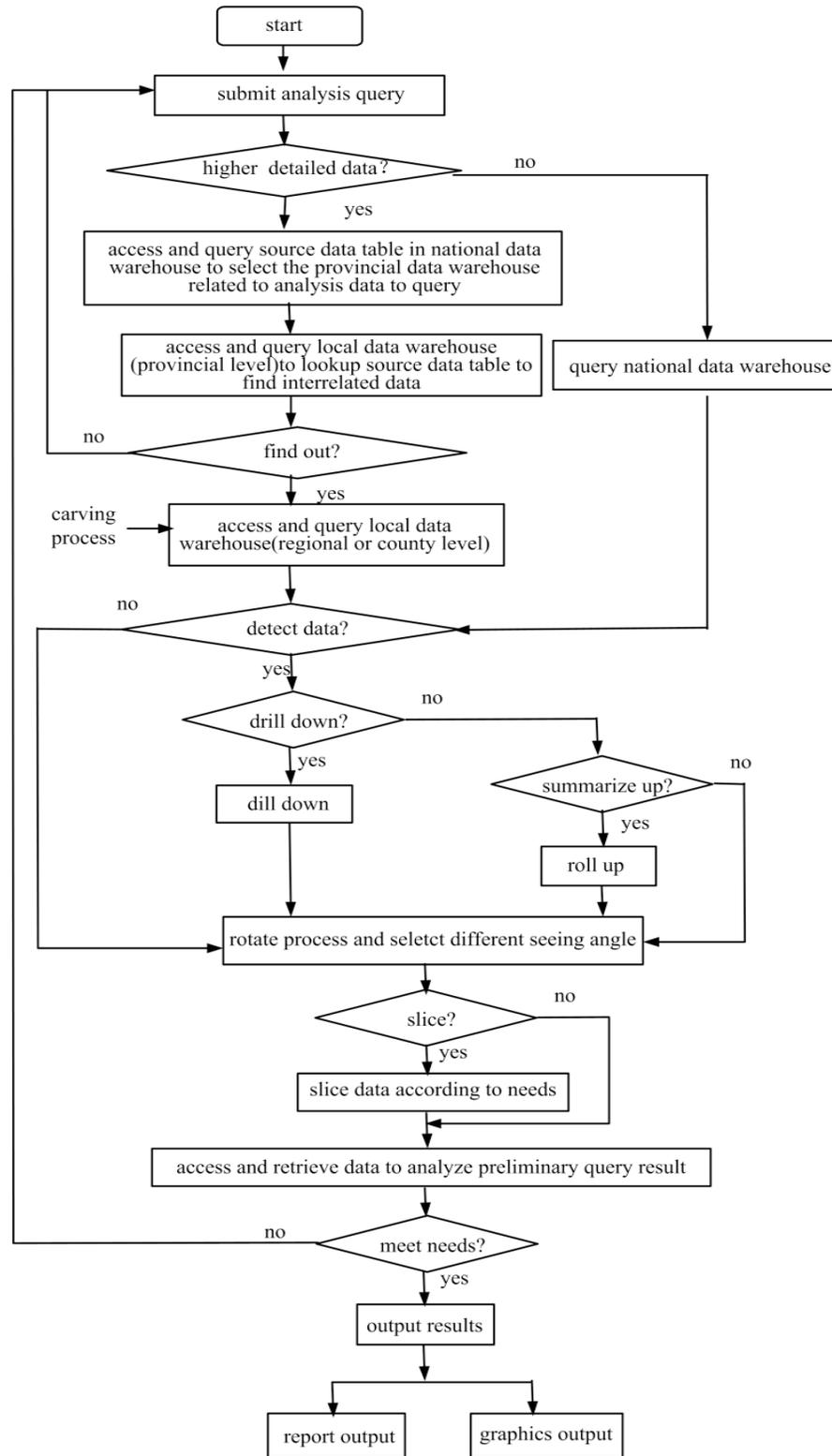


Fig. (5). Dimensional analysis flow char of heritage DDW.

CONCLUSION

For these problems in the cultural relics information sharing research and management, the paper put forward to combine distributed data warehouse and OLAP technology to apply the antique system to solve management and sharing

problems. First the architecture of heritage information distributed data warehouse is proposed, and then according to antique classification proposed by experts and scholars of State Administration of Cultural Heritage, then the design of the global data warehouse and local data warehouse are discussed, and a DDW multidimensional analysis algorithm is

proposed. Such system and algorithm can make heritage managers and researchers look over antique information with various different comprehensive degree from different angles and different aspects from different places to support analysis and decision to antique information and realize sharing of antique information.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Henan Science and Technology Project (Henan Finance Office to teach 108).

REFERENCES

- [1] W. H. Inmon, *Building the Data Warehouse*, New York: John Wiley & Sons, 1996.
- [2] <http://www.nach.gov.cn/>, 2005.
- [3] P. Tao, "Construction of antique information data warehouse and research on antique classification," *Computer and Applied Chemistry*, vol. 22, no. 4, pp. 258-260, 2005.
- [4] D. Theodoratos, and T. Sellis, "Data warehouse configuration", In: *Proceedings of the 23rd VLDB Conference Athens*, Greece, pp. 126-134, 1997.
- [5] Y. Yan-yan, and H. Mei. "Data warehousing technology and reusable component library system", *Computer Science*, vol. 5, no. 26, pp. 56-60, 1999.
- [6] S. Wang, *Data Warehouse Technology and OLAP*, Beijing : Science Press, pp. 79-80, 1998.
- [7] Y. Frank, and S. C. Tseng. "Design of a multi-dimensional query expression for document warehouses", *Information Sciences*, vol. 174, no. 2, pp. 55-79, 2005.
- [8] Ted Gannon, Don Bragger, *Data Warehousing with Intelligent Agents*, vol. 1, no, pp. 82-100, 1998.
- [9] E. A. Rundensteiner, A. Koeller, and X. Zhang. "Maintaining Data Warehouse over Changing Information Sources," *Communications of the ACM*, vol. 43, no. 6, pp. 57-62, 2000.
- [10] F. S. C. Tseng. "Design of a multi-dimensional query expression for document warehouses," *Information Sciences*, vol. 174, no. 2, pp. 55-79, 2005.

Received: September 22, 2014

Revised: November 30, 2014

Accepted: December 02, 2014

© Sun *et al.*; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.