

# Metadata Based Visualization System for Multidimensional Data

Yan Peng

School of management, Capital Normal University, Beijing, 100048, China

**Abstract**—With the ability to measure and simulate more processes at finer scales, the number of data dimensions and data objects has grown significantly in today's datasets. Much existing research has focused on different visualizations for different data exploration tasks. In this paper, based on the financial management data of MAIN company, a financial management system is analyzed and designed, meanwhile, MAIN's important financial data are extracted, classified, and sorted out. with the analysis of financial data, a multidimensional (MD) data model based on the metadata was implemented; the system architecture and visualization exhibition pattern are designed based on BI(Business Intelligence) technology and an existing metadata management system; The Visualization Exhibition System was implemented by ASP.NET and Silverlight Technology. This system provides multiple perspectives and comprehensive query analysis functions for corporate decision makers and managers. It also provides a feasible solution of BI front-end display system for small and medium-sized enterprise.

**Keywords**—*metadata multidimensional; data visualization; data warehouse; data analysis*

## I. INTRODUCTION

Datasets are becoming increasingly complex. Knowledge discovery from these big datasets are becoming a challenging task. Various enterprises have set up their data warehouses, which have already saved massive amounts of data. While these data have great potential in terms of analysis, there is a difference between raw data and the knowledge that can be extrapolated from it. This is where data visualization comes into play[1]. Managers are overwhelmed with data and standard tools are often insufficient to enable efficient data analysis. Meanwhile, visualized exhibition tools are expensive and do not always meet the demand of the enterprises. Visualization transforms data into readily comprehensible images and is an indispensable part of the knowledge discovery process[2]. Designers of multidimensional(MD) models must structure the information that is available into facts and dimensions. Facts are usually measures of business processes of some kind and dimensions represent the different ways in which the data can be viewed and sorted[3]. Business intelligence(BI) systems extend this by using heuristic and statistical models to find interesting features and make predictions. In order to adapt to a particular dataset, however, such models require tuning and learning. This leads to a significant need for cognitively effective and efficient techniques for interactive exploration of the data[4].

Based on the enterprise financial metadata management system, this paper discussed a MD analysis approach, and

combined specific financial requirements with web based MD data analysis approach. A Metadata Based Visualization exhibition system is implemented. The system provides various kinds of statistical form and figures with MD data system analysis operations. Meanwhile, a set of reference methods and tools are provided which will facilitate decision makers with an easier data analysis.

The rest of the paper is organized as follows: Section 2 presents the related research work. Section 3 describes the design of the proposed system architecture, and Section 4 describes the functionalities of the systems and illustrated visualization output, and we conclude the paper in Section 5.

## II. BACKGROUND AND RELATED RESEARCH

### A. The Effect of the Metadata

The metadata is the "data about data"(i.e. the content, quality, condition or other character information about the data). Metadata can be obtained from many technical, such as database catalogs, ETL tools, exhibition tools, etc.

The basic function of the metadata is describing the content and properties of the information resource object and it can reflect the full view of the information resource object perfectly. A typical data warehouse structure consists of operation environment layer, data warehouse layer and business logic layer in an integration project. In a data warehouse, metadata plays an important role: On the one side, during the process of data integration, the data which was extracted from each data source must follow the same pattern in order to be saved into the data warehouse. The congruent relationship, transformational rules of the data source and data warehouse could be saved in the metadata knowledge base; on the other side, during the process of data warehouse project, according to the uniform data model, the first step is set up the data mart. Then build the data warehouse based on the data mart.

Many organizations have formulated all kinds of metadata management standards in past ten years. Among them, the OIM(Open Information Model) of the MDC(Meta Data Coalition) and the CWM(Common Warehouse Model) of the OMG(Object Management Group) have been adopted by most software companies. And as the consolidation of OMG and MDC, these two standards would be compatible with each other step by step. All the tools will obey the CWM standard so as to pave the way for development of metadata.

### B. MD Data Visualization

Graphical representations of data in a spatial array can facilitate the comprehension and analysis of many types of data[5]. Applications now allow the rapid generation of maps, charts, timelines, graphs, word clouds, search interfaces, and many others capabilities[6]. Statistical software packages (e.g., MatLab, SAS, SPSS) that have utilities to plot three (categorical or continuous) variables. In general, software packages provide many useful tools for data exploration including display rotation, zooming, and point identification. However, these tools also create many research challenges that the scientific community must tackle. First, new approaches are needed for determining how best to visualize particular kinds of scientific data. A strong start in this direction can be seen in the "Periodic Table of Visualization Methods" developed by Lengler and Epler[7], which shows a number of visualization techniques. A second challenge is to create and maintain data and information provenance for visualizations [8].

## III. SYSTEM ARCHITECTURE

### A. System Design

In consideration of the system design principles and the design of data warehouse, as well as the functional requirement of the enterprise financial business intelligence platform, we set up the main framework of the Visualization Exhibition System of MD data analysis based on the metadata, shown as Figure I.

The Metadata plays an important role in the entire process; it controls every step. This system set up the MD data set based on AB company's metadata management system. Based on the fact table of MAIN's metadata management system, choose the necessary dimensions and metrics, and then combine them into the multi-dimension that we need to analyze.

The metadata management system records the description information of the data, data processing, and application structure. Its main goal is to provide the all-sided introduction of the data resource. It ranges from the generalization of the conception of real world to the detail physical description of some specific database management system. Metadata plays an important role in the design and operation of the data warehouse. It expresses all the objects and all aspects of the data warehouse. It is the data of all the operation, management, and data is the kernel of the data warehouse.

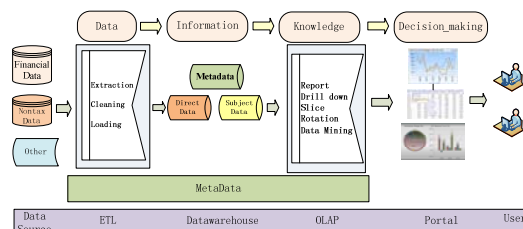


FIGURE I. ARCHITECTURE FOR MD DATA MODEL IN VISUALIZATION EXHIBITION SYSTEM

Metadata is the foundation of communication between the business personnel and technician. The technician gets the physical model and relationship of the data in the data warehouse that it is the base of technical realization; business personnel could see the business relations in the data warehouse from the metadata that is the start of data analysis. As referred on the strong function of metadata above, our MD data analysis model is based on the metadata that is used as the base modeling tool in order to obey the demand of company MAIN's financial data mining and decision analysis system user, as well as increase the usability of the system.

The system can explore the overall strategic target of the enterprise and reflect the measured parameters of the change of the key drive factor to win the competitive edge of the enterprise.

### B. MD Analytical Stars

The system data modeling is based on a star model, shown as figure II. The facts table is in the center of the star model. Every dimension has its own attributes. Dimension table is connected with fact table by primary key. For instance, as shown in figure II, Sale was considered as fact. Time, Product, Customer, and Area are four dimensions which have their own data table. The tables are connected by primary key. The majority of the star queries running in the data warehouse used a query execution strategy known as the star transformation. The star transformation provides very efficient query performance for star queries.

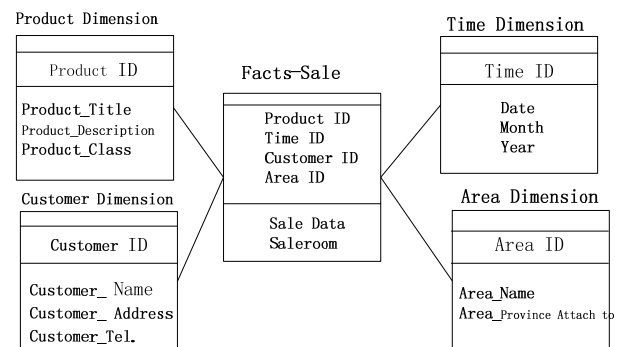


FIGURE II. STAR MODEL

### C. Interface Implementation

Firstly, only one facade class is opened as a data interface unit for users to access the interface. Thus, it can support larger data read out load. Secondly, the file format of the data interface is set to provide standard data exchange service. The standard data follows the national standard, the industrial standard of financial information management. It covered many interactive data formats, such as Excel, TXT, XML and so on. Thirdly, in consideration of the difference and commonality of the heterogeneous databases during the process of data exchange, we design the data exchange format to conduct the peer exchange of data. Finally, the data interface model is designed according to above demands, shown as the Figure III. To ensure the high efficiency of the interface, remote service program cannot be called. A

singleton pattern was used to ensure that only the definition interface class was used; one produce only has one connection, and various interface processes could share this connection. Furthermore, the connection could be disabled in the right time automatically. That means the external calling program would not need to know how to connect or disconnect. So we packaged all the operations of the database in this interface.

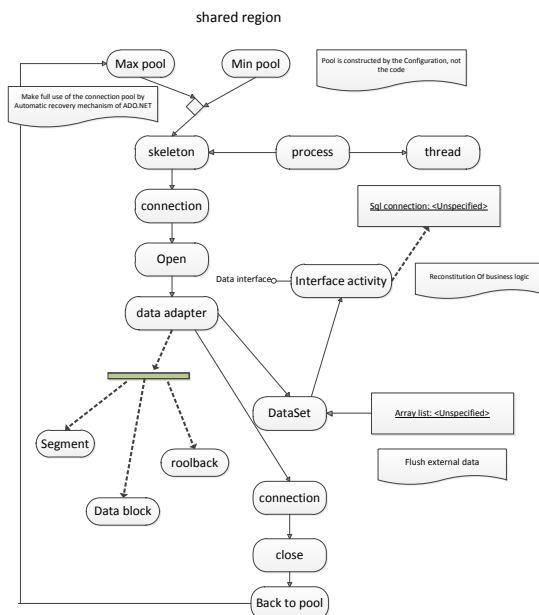


FIGURE III. DESIGN OF DATA INTERFACE

#### D. Exhibition Board Unit

Exhibition Board Unit is the core unit of the display system. It can be presented as a map, table, parameter activeX control, time navigation, etc.

When logging into the main interface, the system could display the original project and related analysis results according to the current login user permissions. Users can view the previous analysis results and, also, can modify by specific needs or redesign the query conditions. Users can also create new projects, query, and analysis for the project. The query and analysis results will be stored in the database, which are used to prepare for the next check. Users can create new projects according to need. Systems will automatically load topics about new projects and the fact table and dimension table about the topic.

### IV. SYSTEM OPERATION DISPLAY

Because the system was built for a Chinese Enterprise, the interface is in Chinese.

#### A. Main Interface of the Metadata Management System

Because this system and MD datasets were implemented based on enterprise metadata management system, the main

interface of metadata management system shows firstly as figure IV.



FIGURE IV. MAIN INTERFACE OF THE METADATA MANAGEMENT SYSTEM

Here, the metadata management system is the tool of MD data modeling. The Metadata management system contains subjects of the enterprise. Each subject has fact tables, dimension tables, and data transformation rules. These are the departments of establishing a MD data set. As shown in figure 4, all the apartment subject data of currently logging user is loaded on the left side of the metadata management system. It is in the form of tree structure. Each subject includes two fixed tree nodes, which are a fact table and a dimension table. The fact table node downloads all the fact tables of the current subject, and the dimension table downloads all the dimension tables of the current subject. Both fact table and dimension table can have operations such as adding, deleting, changing, etc. In addition, the measurement of the subject, the fact table and dimension table can be set and modified.

Based on the metadata management system, users can flexibly build their own MD datasets.

#### B. Exhibition of MD Data Analysis Results

According to the requirements of the maritime financial data analysis, the system designs thirteen kinds of chart types, which are column chart, bar chart, pie chart, line chart, column line chart, area chart, combination chart, point chart, bubble chart, dashboard chart, packing chart, percent packing chart, geographic distribution chart.

##### 1. The Comparative Analysis Chart

Shown as figure V, it is the year-on-year analysis chart of non-tax income. The analysis results are displayed in bars and lines comparatively.

2. The Dashboard Chart The chart can show the income and its key factor conveniently, quickly, and clearly. It could show the year-on-year non-tax income upside down clearly according to the analysis results, and can figure out the key factor which influences the enterprise income most. These results can be the guidance for enterprises to increase their income.

The analysis result of enterprise annual budget execution shown as figure VI. Figure VI is the dashboard chart of enterprise's annual budget execution. Annual budget is an important task of the finance department; it plays an important role in enterprises. Comprehensive budget can refine the

enterprise strategy. It is the foundation of performance management and used as an important basis in scheduler sources. In addition, the budget management is important for business risk control, improving income and cutting down cost.



FIGURE V. YEAR-ON-YEAR ANALYSIS OF NON-TAX INCOME



FIGURE VI. DASHBOARD CHART OF ANNUAL BUDGET

## V. CONCLUSIONS

Based on business financial data, this work establishes data warehouse with Oracle by using Visual Studio 2010 as the development platform, ASP.NET combined with Silverlight and c# as develop tools. The Web-based exhibition system is implemented by Analysis services and XML. The MD data visualization exhibition system displays the data analysis results with varying forms of graphs and charts on the webpage. The system interactivity, efficiency, and maintainability are improved. It provides a feasibility reference for other enterprises to establish a low-cost data analysis system. With plenty of graphics displayed, the decision maker can observe the data analysis results more intuitively. The system support IE7.0 or above browser, B/S structure and with authority management to ensure the data and application security of financial system. Only the user with corresponding data permissions has the authority to manipulate the data. The system uses metadata to describe the data structure, which increases the coordination between departments and improves the efficiency of the relevant department.

## ACKNOWLEDGEMENT.

Thanks for the support of Beijing Natural Science Foundation (9142002) & Research Project of science & technology Innovation Platform of Capital Normal University

## REFERENCES

- [1] David A. Aoyama, Jen-Ting T. Hsiao, TimeLine and visualization of multiple-data sets and the visualization querying challenge, *Journal of Visual Languages and Computing* 18 (2007) 1-21
- [2] C. D. Hansen, C. R. Johnson, *The Visualization Handbook*, Elsevier Academic Press, 2005.
- [3] Oliver Rubel, Sean Ahern, E. Wes Bethel, Coupling visualization and data analysis for knowledge discovery from multi-dimensional scientific data, *Procedia Computer Science* 1 (2012) 1757-1764
- [4] Mark JohnSifer, John Michael Potte, Exploring hierarchical multidimensional data with unified views of distribution and correlation, *Journal of Visual Languages and Computing* 24 (2013) 289-312
- [5] Jesús Pardillo, Jose-Norberto Mazón, Juan Trujillo, Extending OCL for OLAP querying on conceptual multidimensional models of data warehouses, *Information Sciences* 180 (2010) 584-601.
- [6] Matthew J. Pastizzo, Robert F. Erbacher, Multi-Dimensional Data Visualization, *Behavior. Research Methods* Vol. 34, No.2(2002)158-162.
- [7] Peter Fox and James Hendler, Changing the Equation on Scientific Data Visualization, *SCIENCE* Vol. 331 No.2(2011)705-708.
- [8] A. Perer, B. Shneiderman, "Integrating statistics and visualization: Case studies of gaining clarity during exploratory data analysis," *ACM Conference on Human Factors in Computing Systems (CHI 2008)*, 265-274.