Application of data warehouse technology in power analysis

Xiaoguang Hong, Juan Chu

Computer Department of Shandong University, Ji Nan 250061, Shandong province, China

hxg@sdu.edu.cn, chj@dareway.com.cn

***Abstract*: Data warehouse technology transforms the operational data store to general and compositive information. It also provides effective way for analysis and statistic to the mass data, and helps to do the decision-making. The paper introduces the data warehouse and the online analysis process. According to the power management information system, the paper analyzes the power-oriented multidimensional database modeling and the presentation of multidimensional views. The system also builds power-oriented data warehouse using DB2 database, Warehouse Manager, Cube View and Alphablox of IBM. Then, this paper discusses the concrete applications of data warehouse in completing the extract, transform and load (ETL) of power data. The discussion includes the optimization of the system after creating the MQT tables using Cube View and the presentation of the analysis system. The system can help power corporations to do the decision-making.**

***Key words:* data warehouse; power management; olap; multidimensional views; power system**

I. INTRODUCTION

Along with the prevalence of the computer application technology, the foundation of the power management information system advances rapidly. The basic application systems have been founded and are running steadily, such as the power management information system (MIS), Automatic Meter Reading system, the Distribution Management System, the Automatic Mapping Management information System, the Substation Automation and the office automation (OA) system. It is an important resort to build higher level application making for the analysis and the decision-making using the preceding computer technologies and the basic application systems. It can help the electric power corporations enhance the management level and the circulation efficiency, improve service quality and finally upgrade the enterprise competition. The method to solve the problem is building the data warehouse.

The decision-making system based on OLAP has been applied in the areas such as banks, hospitals and chain sales. In [8], the OLAP is applied in the telecom management system. In [9], the paper gives an instance using OLAP in insurance operation. In the setting of power sales analysis model, the paper builds a power sales-oriented data warehouse using Warehouse Manager, Cube View and Alphablox of IBM according to the power management system [1].

The rest of the paper is organized as follows. Section Ⅱ introduces the data warehouse technology. Section Ⅲ presents the data warehouse design and implement in power analysis system. Finally, section Ⅳ discusses related work and draws some conclusions.

II. DATA WAREHOUSE TECHNOLOGY

"A data warehouse is, primarily, a data collection which is to favor decision-making of enterprises or corporations, subject-oriented, compositive, cannot be renewable and can be changed at any time."[3] Because data warehouse doesn't have strict mathematic theory base, it leans to be an engineering project. In technology, it can be divided into key technologies such as data extraction, data storage and management, data presentation according to its work process. Data warehouse is a

location of data for storage and analysis. OLAP is a technology allowing client applications to access these data effectively. OLAP builds the analysis-oriented multidimensional data model based on the integrated data of the data warehouse. It also analyses and compares the subjects from several different angles of views using the multidimensional analysis method. It provides a validated analyzing method through its powerful analysis function and a high performance decision-making for users who have definite analysis scope and analysis requests [6].

III. DATA WAREHOUSE DESIGN AND IMPLEMENT

A Research background

Through the data presentation of power supply and power sales, it can provide the power basic guideline data. The presentation may be about the whole corporation, subsections, depart transformer substations, depart voltage levels, depart time segments, and depart lines. After the data mining of the basic data, it can carry out a lot of analysis, such as the power sales analysis in different time segments, the power used in different areas, the power sales analysis in different electric prices, power usage analysis in different industries, and power usage analysis in different units and the big users' analysis. Data source is the power supply automotive system.

B System frame

Power analysis system contains modeling, ETL, cube-building and final presentation. To achieve these aims, the paper uses DB2 ESE, DB2 Warehouse Manager, DB2 Cube Views and DB2 Alphablox [7]. The whole frame is shown in figure 1.

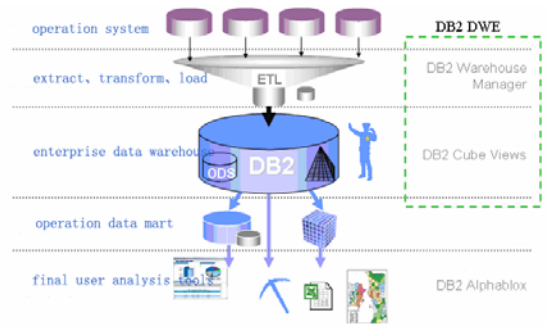

Figure 1 system frame

C The foundation of the snowflake type data logic model

After seeing into the Sales Atomization System, the Automatic Meter Reading system and so on, the paper picks up the power sales as the subject and designs the snowflake type model according to the related factors affecting the subject. The factors that affect the power sales subject are: users (including the big users), time (year, half year, quarter, and month), industry classes, used electric levels, power prices grades. As a result, the paper forms the data set needed by the power query into three dimensions (time, power price, and user) and a measurement attribute (power sales). Time dimension makes up of month, quarter, half year, year; user dimension makes up of units, industry and other attributes; Power price dimension makes up of power price; big user dimension makes up of big users and the attributes concerned by the power companies. Moreover, user dimension distinguishes user characters, such as units' levels and industry classes. Because of the quantity and the various levels of users, it is not possible to record every basic unit exhaustively. Then, the paper selects several representative objects from the thin granularity level. For example, when analyzing the historical data of the user granularity, the paper merely records the big user. This is useful in practical applications. According to the upwards analysis, the paper builds the snowflake type logic model which is shown in figure 2. The center of the snowflake model is analysis content, corresponding to the fact table. Around it is the access angles, corresponding to the dimension table. Each dimension also can be divided into different granularity [4].

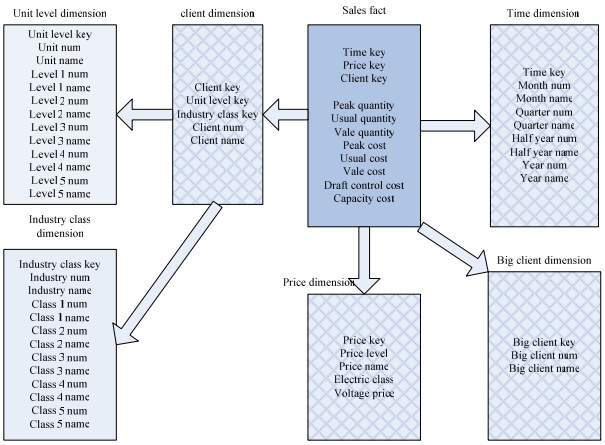

Figure 2: snowflake type logic model

D Building the data warehouse

The implement of data warehouse is completed by several supporting tools. IBM DB2 provides the supporting tools to implement the data warehouse, such as Warehouse Manager and Cube View, Alphablox and so on.

1) Data extract

Data extract is to clear and transform the data sources from different places and the data of different forms. This is complied with the definition of the subject table, the data source definition, and the data extract rules definition in the Meta database. The data extract also reconstructs and processes the data to load it into the target data warehouse. When organizing the data from different sources, the paper turns the data into a middle model and takes it into the casual workspace. Processing the data is to guarantee the integrality and the consistency of the data in the database [2]. When extracting the data, the paper cooperates intimately with the end users to achieve the unification of the data in deed. Data extract in early stage is implemented by handwork programming and the program builder. But now, we carry it out through high-effected tools, such as Infomoter of Ardent Company, SAS/WA (Warehouse Administrator) of SAS and the complete data warehouse solutions provided by several data warehouse companies [5]. In this project designing and implement, the paper adopts the DB2 Warehouse Manager of IBM to complete the data extract.

Each data extract process completes a data extract of an object whose function is comparatively absolute. For example, an extract process completes the data extract of a dimension or the fact table. A data extract process may come down to several tables, and need to apply several data extract rules, and do some data exchanges. Besides, there are restricting relations in the data exchanges and a data exchange needs to be done after one or several exchanges. For example, user dimension's data extract process needs to extract data from users' basic information table and clear the unneeded user information. At the same time, it also appends the units' levels and the industry classes of the users to the final user dimension. In the extract process, we also encounter a lot of problems. For instance, in the process of extracting the fact table data using the warehouse manager, because of the excessive data, DB2 reports problem: transaction log is already full [2]. The paper solves it by input the command below in db2 command center to augment the transaction log:

Update database configuration for SDL using logfilsiz 5120;

Update database configuration for SDL using logprimary 15;

Update database configuration for SDL using logsecond 100;

（SDL is the name of the database）

After the data extract, the system completes the original load of the power sales analysis data warehouse. Thereafter, the system periodically adds to the data warehouse according to the update rules and the load frequency.

2) Building multidimensional database

Up to now, the founded data warehouse only contains current detail level data. In order to get the comprehensive data set, the system selects OLAP presentation as the interface between data warehouse and users. This is in allusion to the data warehouse model power system that will be built. Then the paper can also satisfy the query and the report forms requirement of the decision-making and the multidimensional environment. OLAP presentation can reveal the data of the same subject in different angles. Users can also make up the angles and the mode of the presentation based on their requirement. For instance, the user wants to analyze the power sales subject. Accordingly, the paper presents the

power sales from different angles such as user (including the big users), time (year, half year, quarter, month), industry classes, unit level, and power price. When analyzing, the users can select all of the above angles and analyze in the composite effects of the angles. They can also select their interested angles to analyze. The cube model built by IBM DB2 olap center is shown in figure 3.

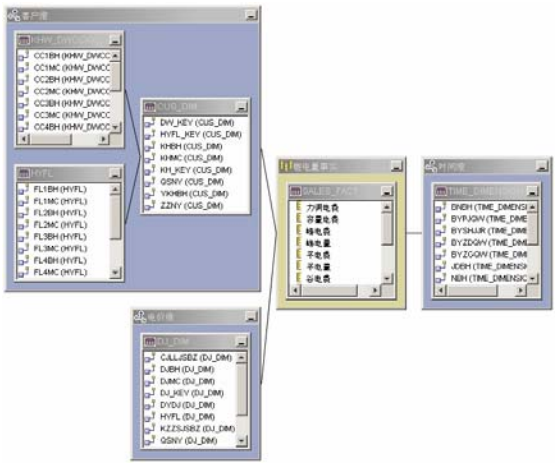

Figure 3: cube model of power sales

According to the metadata and the input of users, DB2 Cube View use the optimize adviser to make suggestions of corresponding MQT. The optimization is based on the different levels and the requirement of every dimension. MQT can improve the query performance of the GROUP BY, GROUPING, RANK and ROLLUP OLAP functions. For example, after building the cube of power sales as figure 3, the system transfers the model optimize adviser. The paper appoints the optimize slice like this: time dimension's any; user dimension's any, power price dimension's power prices. Then, in the appointed time scope and sampling data, the system gets suggested MQT table's sql sentences. If we are content to the MQT, we can execute the sql sentences and create the materialized table. In the process of analyzing mass data, db2 can automatically transfer the MQT. It is to get the optimization of time at the expense of space. So the query response time is shortened and the system is more efficient.

3) Data presentation

The data in data warehouse must be interactive with users and then become useful information. It can be final user-oriented and business subject-oriented. This is the ultimate aim of building data warehouse. After we built the cube at the background of the server, we can connect with the Alphablox's front presentation through the Alphablox's Cube Server of Alphablox tools. Query is the base of data presentation. So the Alphablox Cube Server Adapter provided by DB2 Alphablox is mostly used for carrying out the data query function of cube. Query system includes MDX (Multidimensional Expression) query, visual interface query and so forth.

I. MDX query

DB2 Alphablox queries the DB2 Cube with the MDX multidimensional query language. MDX is component of the special query language OLE DB which is created by Microsoft Company. DB2 Alphablox cube support MDX syntax and its functions. We write peak quantity query program of big user power sales using the MDX language as follows:

```
Query=" SELECT
DISTINCT( {[SDL-DKH].[price dimension]} ) ON AXIS(0),
DISTINCT( {[SDL-DKH].[big client dimension]} ) ON AXIS(1)
FROM [SDL-DKH]
WHERE
(
[SDL-DKH]. [Measures].[peak price],
[SDL-DKH]. [Time dimension]
)"
```

II. Visual interface query

Alphablox designs the database visually through the presentation template of IBM which includes three levels: edit.jsp, report.jsp, and template.xml. Data subject is power sales analysis, and the data dimensions are carried out through two dimensions or three dimensions views. The two dimensions view adopts straight square figure, histogram, graph, cake figure and so on; and the three dimensions view adopts solid straight square figure, solid cake figure, and radar figure and so on to build charts. Besides, on the basis of dimension level design, we can drill up or drill down on the graphical interfaces to achieve multidimensional presentation. Figure 4 shows the power sales analysis system presentation.

Figure 4: power sales analysis system presentation

IV. CONCLUSION

The paper develops on the base of IBM DB2 and introduces the design and the implement of power data warehouse. The data warehouse consists of a lot of main functions. For example, it provides powerful integrate analysis statistic function on the mass power data and presenting exhaustive analysis result. It also offers various report forms and charts to do decision-making analysis. And the system provides corresponding analysis result according to different operation departments' application requirements through each authorization and authentication platform. Contemporarily, it supports several databases like DB2, Oracle, Sybase and several data interface like OLE DB, ODBC. The data warehouse also supports multidimensional analysis and user's multi angles gathering data and predigesting data analysis transaction logic. The data in data warehouse always keep consistent with the up to date operation data, therefore it guarantees the data currency. DB2 offers a complete disaster recover method and establishes the base for the decision-making and the data mining in later stage.

Data warehouse is suitable for the enterprises with complex constructions and client environment. Therefore, building data warehouse in power information system is necessary. Based on the power analysis system, we can improve the data warehouse system response speed furthermore and develop the front-end presentation more satisfactory for clients. And the system also can be optimized to be base for data mining.

ACKNOWLEDGEMENTS

The work is partially supported by Jiang Hua of IBM China Company and the Dareway Company of Shandong University.

REFERENCE

[1] W. H. Inmon. Building the data warehouse (the 2nd edition) [M] . New York：Wiley Computer Pub. ，1996.

[2] Yang Qichang, Li Yulong etc. IBM DB2 high-level management guide [M]. Beijing: publishing house of electronics industry, 2004

[3] Wang Shan. Data warehouse technology and online analysis process [M]. Beijing: Science Press. 1998.

[4] Lou Weijin, Kong Fansheng, Lou Weizhong. Data warehouse and knowledge discover [J]. Computer Engineering and Applications, 2000, 36(10): 111-113

[5] Zhang Wenying, Shu Hongchun, Zhang Ye. Application of data warehouse technology in power consumption analysis of Kun Ming power network [A]. Power System Technology, 2005.

[6] Ralph Kimball Laura Reeves, Margy Ross Warren Thornthwaite. The data warehouse lifecycle toolkit: Expert methods for designing, developing and deploying data warehouse. Beijing: Publishing home of Electronics Industry, 2004.

[7] Michael. Gonzales. IBM Data Warehousing With IBM Business Intelligence Tools [M]. John Wiley & Sons. 2003.

[8] Zhang Weimin, Zhang Xiaoying, Ju Di. The application of data warehouse technology in telecom management system. Electron engineer. 2000.2.

[9] Zhan Min, Sun Yufang. The implement of OLAP in insurance system. Journal of Software. 2000.11.