Telecom Data Warehouse Construction and Application

1Qin Hai-fei and 2Hu Jian-hua
1Department of Computer Science, ChuXiong Normal University, ChuXiong, YunNan, China
2Department of Computer Science, Kunming University of Science and Technology, Kunming, YunNan, China

Abstract: With winning initiative and more commercial opportunity in furious market, business intelligence is needed to guidance business behavior and to assist decision-making. Data warehouse, as the data platform for support of DSS, the research and application have been hot issues among the theories and practice circle in china. This study come from the data warehouse construction in the relate department of XX telecom. It introduced the key technical problem as for the understanding of the data warehouse construction, the basic architecture of system, development model and method, the choice of data granularity and so on. Talk about the application of data warehouse and its orientation. However, because of the complexity of the field itself, data warehouse construction and successful application will not be easy, requires constant effort.

Keywords: Data granularity, data integration and extraction, data warehouse, OLAP, software development model

INTRODUCTION

Development of information technology have become increasingly demanding of data, as the information technology leading enterprise, telecom, has a huge amount of data. How to make data wealth, not a burden, become the target of many enterprise seek for. The operator face to the actual situation internal, introducing data warehouse technology, to integrate the data scattered, isolated from various business system internal, provided a convenient means of data access and met the needs of different users, to promote enterprise "customer-centric" business model carry out, improve the core competitiveness of enterprises (Tian, 2008). At present, in the communications industry, existing commonly the data support analysis cannot meet the actual work requirements (Huang, 2005). Popular speaking, the leader wants to see each kind of customer creation profit how in some time section? Each kind of service income what tendency is in some time? Telecom answer to this question is not easy. There are three reasons as follows. Firstly, the data is huge, dozens of G per month, using the traditional methods the response time and data storage capacity is unable to realize (is limited to low-cost PC server). Secondly and telecom billing rules are very complex and change frequently. Thirdly, the data must be accurate, cannot have a penny going out. Based on the problems encountered in the practice of telecom projects, Access many materials and experience repeated practices, have a technology idea about data warehouse (Huang, 2005).

CONSTRUCTION OF DATA WAREHOUSE

Data warehouse, OLAP, data mining is new technical as the computer development (Weiss, 2005), about the principle of data warehouse construction, many books and literatures talking many (Inmon, 2006). The data warehouse construct for the architecture different and different. As to this architecture, is involving some problem, the rule of data warehouse construction, the data model design and development and so on. Here three points experience.

Data warehouse architecture: The data warehouse architecture impact on specific technique routes. Top-down and bottom-up, these two methods are in debate in this field. The former believe, “you can reach tens of thousands of tons of fish shrimp, but these fish shrimp added up to not a big whale”. This sentence come from the father of the data warehouse W.H. Inmon; the latter think, should get the maximum return with minimal investment in the project early and then developing research, gradually perfect (Zhang and Li, 2006; Deng et al., 2002). The two views are justified, because the construction of data warehouse success or not is not only depends on the technical issues, but also on business management.

Objectively, the project scale determines the develop method. According to the top-down, this study discusses the data mart should be more accurately, of course, it is the perspective on problems. Now, somebody puts forward the combination of the two
solutions, it will be worth exploring and research. From the point of view of software engineering, the top-down analysis and the bottom-up design combination is accord with software engineering idea. From the point of view of the actual of china, the combination is more suitable for practical, from the part beginning, combining with consideration of the enterprise's whole data model. From a technical standpoint, to solve the metadata model unifying, realize the dataflow between the data warehouse and data marts.

**The application of software development model:**
Conceptual design and requirement analysis are two of the key steps within the data warehouse design process (Robert and Bernhard, 2012). Software development model is the framework that spans the entire software development process (Inmon, 2006). With the advances of software technology, this area of research is also very activity (Deng et al., 2002)

A data warehouse is a subject-oriented, integrated, nonvolatile and time-variant collection of data in support of management’s decisions (Inmon, 2002). In this study, from a strategic point of view, we think we should follow the spiral model development methods, because the data warehouse is subject-oriented. In other words, its construction is driven from the analysis and decision-making needs, the subject must be able to support users' analytical needs; you have to understand the users' goals intensity fully. Compared with traditional information systems, this requirement process is very different. The biggest difference is the user may not give you a clear business expression; the needs of users are fuzzy. For example, analysts makes a subject analysis as to "customer churn", the user can't give actual demand. Users may only care about "what happened and why happened in the past " and “Which customers will be churn". How answer to these questions. The data analyst is uncertain. But as constructor, we should be able to provide the complete and comprehensive data resources and provide analysis platform to the user. Because the developer is not only the architect and constructor of data warehouse, but also is the analysts of the application. Of course, this process is not a complete, because the developers and the user have a continuous interactive process, so there must be a repeated iteration. In fact, the goal of the data warehouse and the information system are different. The goal of data warehouse is data model that support analysis. The goal of transaction system is software target system. So each cycle of activity should be different.

Data-driven development approach has specific applications in the work, significant guidance. The thought is very clear, that is the data and the process logic change with data model. Its solution is not completely abandon the original data model and processing, but is on the basis of the original data model and code; explore both in common, in order to achieve the new system process. For example, the project team has completed the development of the production report of the account billing system, but also has the development tasks of the production of convergent billing system report demand, the development time compared to the account billing system is much shorter and use the developing method to complete the convergent billing must be too late. In the case of a time pressing and difficult task, the development methodology of choice is very important. The development of data-driven approach has played an important role. Find out the common between the new and the old system, become the key to solve the problem. The data model is changed, the application should be changed, but the accounting rules basically have not changed. With this, have in common. With these basic points, technical route was clear, that is, as long as solving the data mapping between the old and new. In principle, the new system processing logic has not changed; the old system can also solve the task of new system, so that the new system development cycle and system test is greatly reduced.

The nature of the two development methods are based on an iterative, just differ in level. Combining the construct of data warehouse with object-oriented approach is new method, but also of a topic of interest.

**DATA IS THE BASIS OF THE DATA WAREHOUSE CONSTRUCT**

Data quality problems are the core issue of BI, there is no high quality data, all out of the question and this is important part in the construction of data warehouse. Through practice and research found that the data preparation is the longest and the most difficult in the entire construction period of the data warehouse.

**Data integration and extraction:** Data from the database to the data warehouse is not a simple data migration. The data extracted from the database is "dirty" data, the data exist errors, missing, duplication, meaning of the same name, different names with meaning and other issues. There are several reasons about the quality of the problem appeared above, can be summed up in the followings kinds in general. Systematic error, rules errors, management and control errors, data sources error (Zhong, 2012). For the
problems of data, the null data, noise data and incorrect data, must be cleaning and integration.

The basic work of data processing is to understand the data. From the Timeline view, telecom's billing system has reconstructed many times in a few years and made the database model changes, data inconsistencies result in the data extraction difficulty. From a horizontal perspective of view, the work involves several large systems, requires careful planning the interface design of extract the object. Therefore, understanding the enterprise data model overall is important for solving extract data. Data clean-up is a detailed and onerous task. It decided the data quality of data warehouse. For example, if you want to do marketing policy to retain customers, you have to analysis the customer churn with the user. You have a coincidence rate (i.e., which customers may churn, finally is churn), must be require high quality data. So needs clean up the data (which data is incomplete, incorrect, null value, inconsistency etc.)

The data accuracy decided the warehouse construction success or not: Data is the lifeline of telecommunications enterprises, data quality is good or bad, is not only the flag in integrity, is also the key to make scientific decision. The impact is deeply, once, a user told us, “you as long as the business income calculated error is less than 5% are ok”. If that is really the case, then this project certainly failed. Needless to say, 5% error, even a penny out, users will not believe your system, because the data no accuracy at all, the analysis worthless. Enterprise is real, no matter how good technology, if you cannot solve practical problems, you only have to leave.

Data granularity: The size of the data granularity is an important issue in the data warehouse construction (Inmon, 2006), because of the data granularity directly affecting the use of resources, so the choice of data granularity is particularly critical. In fact, it is a dilemma choice. If you want to answer each customer's daily traffic, the data granularity is necessary to every day/every customer. According to the granularity size, as if currently has 2 million telephone customers, a day has 2 million event, a month will reach 60 million events. You want to follow a year, will be 21.9 billion events. So large data, will take the cost of resources is enormous. However, if you give the analyst a suggestion: Why do you track all of our customers? Using a sample of the data can also solve this problem. Of course, this is talked about the data mining. Why do you want to track every day? In fact, most people care about is how much I consume each month, so the problem becomes simple. If the problem become to solve each customer in monthly traffics, the resource is more acceptable. From above, illustrates a problem, the standard of the data granularity is relative, of course, needs of users is first, also only standard. Secondly, the co-ordination between the builders and users is particularly important.

DATA WAREHOUSE APPLICATION AND POSITION

From the data warehouse application perspective, there are several typical front-end applications: such as reporting sever OLAP (Online Analytical Processing), ad-hoc query and data mining. As the data warehouse has not been universal, for the average user, the data warehouse and its application is mystified, seems adopting data warehouse, any problem that business analysis and decision-making can be solved. Data warehouse application is also a gradual process.

Software is service: Software is a collection of computer programs and related data that provides the instructions for telling a computer what to do and how to do it. The software's main function is to build bridges, to provide services for users to make better use of computer resources, when the data warehouse construct, application program development has become an important issue. Meet the different needs of different users, in fact, also for different users to provide different services. In this project, began to start from the supporting analysis. The main task is to provide the data to meet the various requirements for user. According to this process, the customers feel their needs are catered for well, the data provided fast and accurate, but the hardware just rely on a low-cost PC servers. Support in place, users will believe you, other needs to constantly. If you are starting directly data mining, the situation will be bad, because you do not have the field basis and successful cases, who would believe? Finally, due to the requirements of product and management Analysis, the report system, OLAP (On-Line Analytical Processing), ad-hoc query, data mining application and other applications has been gradually on-line, but each application is progressive, you can offer a good service for user needs, it may have a chance. In the process of the construction of data warehouse, summarize review what needs to be better, the support of business development in place or not, How to cooperate better. What can bring business department for benefit? Use it service for customers better.

Data warehouse applications have a valid impetus: The construction of the data warehouse and the construction of the traditional transaction processing have a big difference. Transaction processing systems, such as the accounting system of communication, you
must be having, without it, you cannot charge, businesses have to close. The construction of data warehouse is very important for competitive enterprise, but its success depends on many factors. Such as the recognition degree of enterprise leaders, enterprise culture, enterprise management, the quality of employees, in addition the personnel quality the constructor, communicative ability, domain knowledge, etc.

The level of enterprise management good or bad determines the successful of the data warehouse construction and application or not. At present, the telecom companies face fierce market competition. Face the competition in market. Managers introduce a competitive mechanism in the internal, including personal income and performance-structure, the boss also so. Everyone has specific performance indicators and every employee to feel strong competition. For example, when we develop the real-time product operation analysis tracking system, the competent leaders stare at products of traffic trends every day. Slightly moved that analyze the reasons and take measures. In the face of such enterprise culture, a developer could not afford any mistake in the data, also should consider to provide relevant analysis ability. Because it is affected the management of the enterprise directly, which is the impetus.

Face to the competition market, the department of management analysis, becomes staff officers and planners of the leader, is also close the application of the data warehouse. Data warehouse applications are driven from the business analysis department and each of the analysts. So the quality of employees high or low determines the success or failure of the analysis of the data warehouse. No one uses your systems, your technology is good, is no good. Luckily, partners give us great impetus in this respect.

Related to the specific management analysis, the business understanding is a prerequisite. This problem previously mentioned, if your business can not understand or scanty. It is difficult to exchange and communication with business personnel in business language, the others cannot believe you. From a hierarchical perspective, this exchange is broader compared with the development of a transaction processing systems, this decided by the practical problems.

CONCLUSION

The business support systems of telecom complex, the kinds of data variety, the amount of data is large, information cannot meet the requirement of the market competition on the speed, quality and scope. Mass data more and more intentional, based on the data warehouse and knowledge discovery has become the hot spots of various industrial applications. This study based on telecom business analysis system as background, the aim for introduction the construction of data warehouse and supporting the data. It main talking about the question of technology in the construction of data warehouse, the architecture of data warehouse, the data warehouse construction of model and method, the choice of data granularity and so on. Talk opinion on the application and the position of data warehouse. The customer requires are changed. The construction and application of the data warehouse will be changed. We will study constantly.

REFERENCES