

# Research on university data statistics service platform based on data warehouse

He Jinlian<sup>1, a</sup>

<sup>1</sup>Jiangxi Science & Technology Normal University, Nanchang, Jiangxi, 330013, P.R. China

<sup>a</sup>jxsyhjl@163.com

**Keywords:**Data warehouse, Data statistics, Service model

**Abstract.** Through the analysis of the whole structure and data warehouse modeling, the data warehouse star structure is used to analyze the data model, and the key technology of ETL is analyzed.

## Introduction

With the application of electronic school affairs in Colleges and universities in the construction, the application of the information management system is becoming more and more, the management of colleges and universities to a good supporting role, after the operation of the information management system for a long time, also produced a mass of data, massive amounts of data are now hidden in the various information management system is the core of running colleges and universities, how effective the data of mining, statistics and analysis, become main problems faced by colleges and universities.

## Demand analysis

Through the investigation and Analysis on the actual situation of the universities, found that the use of statistical data on the demand of university is mainly divided into two categories: one is based on fixed form, report to the superior departments of the main content, is reported quarterly or annual report, it has special statistical data and complex content; two is the school each department daily required statistical data, these data are often associated with specific business sector, compared with the first data, in its content and format is relatively simple, no complicated reporting process, but the size requirements of the statistics is fine, some statistical data are to provide data to monthly, weekly and other forms and, statistical data with drilling function. Through the analysis, we can find that the data of the university is basically related to the personnel, students, teaching, scientific research, property and other fields, in various fields, there are different data classification.

In the process of building data statistics service platform, the functional requirements and business needs of the system should be integrated with the business needs. The user can obtain the corresponding data in a simple, fast and convenient way. Meanwhile, it is also able to care for the non functional requirements.

## Architecture of data statistics service platform

Through analysis, it is found that the University's data statistics service platform is composed of three parts, the source system, the data warehouse and the data statistics platform.

**The source system.** University data is huge, the main business involves the data source is also very extensive, in the University's data source system, the personnel management system, teaching management system, student management system, teaching scientific research system and other related data management system, and campus network of a few other external data, these data constitute the data base of university data statistics, because different data is the use of different database, which lead to a certain heterogeneity between the data.

**Data warehouse.** Data warehouse is to receive the data from the source system, in the process of ETL, according to the rules of data statistics service to data standardization, verification, cleaning processing, and the processing of data loading to the data mart, and then through the data market support system for data query and analysis.

(1) duplicate layer (SSA, system-of-records-staging-area). It directly to the data source of data copy, keep the data in the original, but in the process of copying, in the data source based on the added time information, in the data warehouse will be formed multiple versions of history data.

(2) atomic layer (SOR, system-of-record). SOR is based on the model, which can satisfy the data platform. It stores the data in the data warehouse. According to the data, the data is classified according to the data. SOR is the basis of the data warehouse, can support data sources to support, analysis and upgrade and update.

(3) summary layer (SMA, summary-area). SMA is the intermediate transition between SOR and DM, to facilitate the query and related work of SOR specification data, because the data granularity in DM is higher than that of SOR, it can be easy to collect data in DM to improve the data query ability of data warehouse.

(4) data DM (Mart). It can save the data for the user to access, the data is converted into users want to see and use the data, DM is the fact that all kinds of grain size data to adapt to different access methods. In the University's data statistics service platform, the main use of the star schema and the fixed report structure storage mode to realize the data construction of the market level.

**Statistical platform.** University data statistics service platform mainly uses the B/S structure of the 3 layer system structure, in order to realize the operation of data.

(1) data operation layer: the database is independent of each other, and the data processing layer does not conflict with the data of the two.

(2) logic layer. Mainly include report, report query and audit, report generation and reporting and other related is the data module summary, through the logic layer can be different permissions to users of different management authority and function, and the report of the audit and summary can only have the university management personnel to use to view, in order to facilitate the unified management of the system.

(3) representation. To achieve the function of the call to the data, through the interface for the use of the user to provide support, but also to provide some data interface services to facilitate the call other data system.

**Statistical data model.** The results of data statistics service platform data statistics are presented to the user, for the large amount of data and statistics of the University, the need for data analysis, detailed data reporting requirements for the definition, processing, planning, to effectively complete the data statistics. Specific data statistical model shown in Figure 2 below.

(1) a data star model defining the statistical platform. In the process of data statistics, it need to determine the fact that the facts of the table data, indicators and data granularity.

(2) determining the dimensions of the statistical indicators of the report. The dimension values of each row and each column in the data table are analyzed, and the values of the data in the dimension table are the sum of the data values of the multiple dimensions.

(3) determine the statistical method of a cell. The statistical analysis of the cell data in the data report, and the value of the dimension of the row and column.

(4) the storage of data from the database table. The dimension of the data report is transformed into a statement that can be executed, in the fact table of the database query and statistics.

## ETL rule design and Implementation

ETL (transformation loading extraction) is a key part of data statistics platform. It is responsible for cleaning, converting and integrating data from different data sources, and converting them into the data that is needed by the system. Finally, the data is loaded into the data warehouse and data mart. ETL is the core of the data warehouse, the design and implementation of ETL rules is an important content of the whole data warehouse.

Data extraction: the main data initialization loading and data refresh, the main function of the initialization of data loading is the existence of data in the data table, the table of facts, the main function of data refresh is to carry out data warehouse and maintenance. Data statistics service platform uses trigger mode to refresh the data. In the SSA layer, the data is inserted, modified and deleted from the data table. The data of the data in the source database is changed. The data in the SSA will follow the source data changes, so that the data is written to a temporary area (buffer), and the data is copied into the trigger. In the database layer to define the data processing operations (job) and data storage process (procedure), the operation of the data processing rules to define and adjust, and call the corresponding stored procedure to extract the refresh data storage, and delete the temporary data tables in the data will be deleted or labeled. Trigger mode can effectively extract the data, and extract the high performance, simple rules, but also to facilitate the management of staff.

Data cleaning is to deal with the data in the data warehouse. The data in the data warehouse is the same rule, which is mainly to deal with the data in the source database.

Data conversion function is to convert the data into data warehouse. The main function is to convert data into data warehouse. The data from different sources are normalized. According to the unified data dictionary and format, the data of the source system can not meet the needs of data warehouse.

## **Conclusion**

The design and implementation of university data statistics platform can improve the work efficiency and the data storage, processing, sampling, cleaning and other related functions. It can realize the function of data generation and data query, and can effectively meet the large amount of business data processing and analysis.

## **References**

- [1] Y.H. Luo, M.L. Chen. Another view of the -BI framework system of business intelligence philosophy . software world, 2007, 7:68-71.
- [2] Q.Y. Wang, L. Li, Ouyang Rongbin et al. The main program of the data integrated service management platform to explore the. experimental technology and management, 2011, 28 (5): 4-6.
- [3] N. Ma, Z. Cao, J. Li. based on data warehouse of bank supervision report system . Software guide, 2013, 12 (4): 137-141.