The Multidimensional Analysis of College Students’ Consumption Data
Hua-Jiang LIAO\textsuperscript{a} and Ning HUANG\textsuperscript{b,*}
\textsuperscript{1}Modern Education Technology Center, Gannan Normal University, Ganzhou, China
\textsuperscript{a}jet\_liao@qq.com, \textsuperscript{b}hngzjx@qq.com
\*Corresponding author

\textbf{Keywords:} Campus Card System, Data Warehouse, OLAP, Cub, MDX.

\textbf{Abstract.} This paper focuses on the analysis of the campus card consumption data and constructs the campus card decision support system based on data warehouse, OLAP and data mining techniques. Key steps to establish this system is presented, includes preprocess of data, building of data warehouse, construction of OLAP model and implements of system. Especially, describes the structure of the facts and the dimension tables in detail.

\textbf{Introduction}

In university, there are many evaluation works about students, such as student’s family financial status and consumption level. This evaluation is the basis of some universities subsidies, including student loans, student subsidies or poverty subsidies. At present, these evaluation works mainly rely on subjective impression from classmates, student cadres and teachers, which is difficult to be fair and objective.

At the same time, the campus card system in some universities has been put into use for several years. The system accumulates a large amount of consumption data. For instance in a university of over 20,000 students, usually more than 30 million records are created in a year. At present, such a system only provides some basic query function, such as a student’s card detail records or department’s statement of accounts. It can’t offer some analysis function, such as identifying those students whose expenses in food is significantly less than others. Because all these campus card systems are built on a traditional relational database manage system (RDBMS). The RDBMS’s function is adding, deleting, modifying and querying of records. In order to analyze data, the data warehouse and OLAP is necessary.

\textbf{Data Warehouse (DW) and Online Analysis Process (OLAP)}

A data warehouse is a topic oriented, integrated, time variant, nonvolatile data set that supports the decision making process in management\cite{1}. There is significant difference between data warehouse (DW) and database (DB). First, the database only stores the current value, while the data warehouse stores the historical data; second, the target of database is for operators, while the data warehouse’s is to provide decision support for middle and top managers; third, the data in the database changes dynamically, as long as the business is generated, the content of the database will be changed, while the content in the data warehouse is static and can only be added and refreshed at regular intervals.

OLAP, on-line analysis processing includes a series of technology to handle multidimensional data. It can categorize and aggregate data from multiple dimensions and granularity. The traditional online transaction processing (OLTP) database management system (RDBMS) cannot meet the requirement of intelligent data analysis and decision, because it needs a large number of collection and calculation of OLTP business data\cite{2}. OLAP organizes data in multidimensional dataset, and it makes up a data cube. OLAP can support slicing, cutting, drilling, rotating, rotating shafts and other analysis actions, it allows users to view data from multiple angles and levels, thus deeply understanding the information contained in the data\cite{3}.
Data Preprocessing and Building of Data Warehouse

ETL (extraction, transformation, load) is an essential step in the process of creating a data warehouse, in this process, the source data is loaded to the data warehouse [4,5]. The structure of campus card system’s database is complex. It often has large number of database tables. Furthermore, many tables have dozens of attribute columns, therefore data preprocessing is necessary and it mainly includes the following aspects.

Database Table Selection

The number of complete campus card database tables can usually reach hundreds, but the analysis task usually needs only a few of them. In fact, only 5 tables are finally selected, namely consumption detail records table, users list table, consumption category table, departmental statement of accounts table and terminal device table.

Attributes Selection

Data table may contains hundreds of attributes, most of which may not be related to the analysis task or are redundant, for example, the consumption detail records table contains over 40 attributes. Only 8 of them are selected.

Data Cleaning

Data that is not associated with the analysis task must be cleaned up before building the warehouse. The consumption detail records table contains all the card records. The purpose of the analysis is only relevant to students at school, so the other users’ records must be deleted, such as teachers’, staffs’ or other temporary persons’. The analysis focuses on the daily food consumption, so other types of records such as medical or bus should be erased beforehand.

Construction of OLAP model

Before establishing the multidimensional data set, two types of data tables should be defined. One is fact table which has the summary or detail data we are interested in, called measure. The other type is dimension table. To each dimension, there is at least one table contains relevant attributes; if the dimension is hierarchical, it also includes the hierarchical information of the dimension. The fact table is linked to each dimension table through a foreign key. There are different types of multidimensional data patterns, mainly include star schema, snowflake schema and fact constellation schema [6]. The OLAP model structure of the system is shown in figure 1.

Design of Fact Table

Fact table usually contains large amounts of data. It has several measure fields, and a series of keys which link with dimension tables. There are two fact tables in this system. One is the consumption detail records table which contains a large number of card records in all kinds of terminals. The measure fields are card consumption amount and time; the foreign keys are customer key, date key, type key and terminal key. The second fact table is meal consumption table. In the consumption detail records table there are a large number of scattered single card records. In general, there are several records in one meal. When analyzing the students’ consumption level, the important measure is the amount of meal consumption. Therefore, the consumption detail records should be collected by the card time period, which creates a new fact table, it’s measure is the meal consumption amount, it also includes primary key, user key, meal key, time key and grade key.
**Design of Dimension Tables**

**Time Dimension.** The dimension of time is frequently used in OLAP, it provides time-based granularity levels for analysis. In SQL Server 2012 Analysis Services (SSAS), the time dimension is those attributes which are relevant to time, such as year, half year, quarter, month and day, it is organized hierarchically, the granularity of the time dimension is determined by the reporting requirements of historical data. This system selects three granularities about time, namely year, month and day.

**Meal Dimension.** In this system, the period of card consumption, meal is an important dimension. Just as time dimension, meal consumption is also a time relevant attribute of consumption; but in this system, the statistics of meals in a particular day has no real meaning. It is more important to investigate the meal as an independent dimension. There are 4 kinds of meal types, breakfast, lunch, dinner and supper in this system. By using meal dimension, the card records could be subtotaled by different time period. This is of great significance to the analysis of students’ consumption habits and consumption levels. For example, the students who often don’t eat breakfast will get higher scores in assessing poor students than those often have supper records. The meal dimension table includes 4 fields: primary key, begin time, end time and meal name, as shown in table 1.

**Table 1. The Meal Dimension Table.**

<table>
<thead>
<tr>
<th>Primary key</th>
<th>Begin time</th>
<th>End time</th>
<th>Meal name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>05:00:00</td>
<td>09:59:59</td>
<td>breakfast</td>
</tr>
<tr>
<td>2</td>
<td>10:00:00</td>
<td>15:59:59</td>
<td>lunch</td>
</tr>
<tr>
<td>3</td>
<td>16:00:00</td>
<td>20:59:59</td>
<td>dinner</td>
</tr>
<tr>
<td>4</td>
<td>21:00:00</td>
<td>04:59:59</td>
<td>supper</td>
</tr>
</tbody>
</table>
**User Dimension.** Many important functions of the system are related to the user dimension. The user dimension table also has the most abundant information in all dimension tables. Many attributes of the user or student are important dimensions to multidimensional analysis. The main attributes are as follows:

- Gender, used to analyze the consumption characteristics of students from a gender perspective.
- Native place, used to analyze the consumption characteristics of students from different sources.
- Department and classes, used to analyze the consumption characteristics of students from different departments or majors.

**Grade Dimension.** College students of different grades tend to have different characteristics in their consumption, the grade dimension provides the function for analysis at this angle. Grade does not belong to the users table. It needs to be computed based on the year of enrollment. Add grade field to the consumption detail records table, the grade value can be calculated from card time and the year of enrollment.

Some of these dimension tables can be applied to two fact tables, such as time dimension, user dimension or grade dimension. Others can only be applied to particular fact table, for example, the meal dimension can only be applied to the meal consumption fact table.

**System Implementation**

There are 3 steps to realize this system. First, building up the data warehouse and OLAP database using Microsoft SQL Server 2012 analysis service; Second, establishing the multi-dimensional data set; finally, the development of foreground application. The last two steps are realized in Microsoft Visual Studio 2010. The application connects the Analysis Service by ADOMA.NET. It visits the cube in the data warehouse, which executes the query order expressed in MDX language and returns result.

The users’ selection and input information is converted into MDX query language, it is passed to the Analysis Service server, which will execute the query order and return the result. The result is displayed in the DataGridView control. When user selects a cube and dimensions, a MDX query order will be generated, the typical format is like the following:

```
SELECT {[Dim Grade].[Grade Name].Children } on columns,
{[Dim Meal].[Meal Name].Children } on rows
FROM  [MealCube]
WHERE ([Measures].[Amount],[Dim Date].[The year].[2016 year])
```

Using ADOMD.NET component in Visual Studio 2010, the connection to Analysis Service will be established, through which the MDX query order is passed to the AdomdCommand object which executes the order and get the target data from the multi-dimensional data set.

**Conclusion**

OLAP is a powerful tool for multidimensional analysis of large data. This paper describes the establishing of the data warehouse base on the mass consumption data in the campus card system, analyzes how to put up the OLAP model, especially focuses on the design of fact tables and relevant dimensions tables. This paper uses Analysis Services, Visual Studio2010 and C# to realize the multidimensional data query system, it provides a useful tool for analysis and decision support for management of college students.

**Acknowledgement**

This research was financially supported by the Science and Technology Research Project of Jiangxi Province Office of Education in 2016 (GJJ161676).
References


