

Big Data Business Intelligence System Research Based On The ESL System

Yubo Liu¹

Faculty of Computer, Guangdong University of Technology, Guangzhou, 510006, China

E-mail: 1058388775@qq.com

Lei Ding²

Faculty of Computer, Guangdong University of Technology, ,Guangzhou ,510006,China

In the era of big data, the business management faces many challenges. In this paper we use related data in respect of the retail sales that the ESL(Electronic Shelf Label) system collects and imports into SQL Server 2008 R2 database. With the integration services, analysis services, reporting services integrated in SQL Server 2008 R2 database to transformation, we analyze and display the retail data. On the basis of this work, it can help the retail enterprises to improve their use value of business data and achieve efficient conversion of data into information, which can help their managers to make decisions in terms of marketing.

CENet2015

12-13 September 2015

Shanghai, China

¹Speaker

²The research work was supported by Guangzhou Liwan District Science and Technology Plan No.13ZK0032.

1. Introduction

The Retail business has become one of the world's largest industries. Not only the retail business' rise and fall will directly affect the development of the national economy, but also it is closely related to our life [1]. Nowadays, the informatization degree of retail business is high and the majority enterprises in the retail industry have established the enterprise management system to manage their daily operations such as sales, storage management, financial and other businesses. As the business data in the enterprise management system accumulates by each day, how to make sound use of these data to effectively analyze and monitor the enterprise operation has become the focus of the business' further informatization; in this sense, the big data business intelligence system will play an important role.

Electronic shelf label (referred ESL) is an electronic display device which is placed on the shelf and can replace the traditional paper price label. The electronic shelf labels are connected with the shopping malls computer database with wired or wireless network while displaying the latest commodity price. The ESL system can help collect data associated with the sale of goods and we can import these data into the SQL Server 2008 R2 database through the integrated business intelligence tools (BIDS). Thus we can transform these data into useful information which can help retail business managers to make business decisions to improve profits.

SQL Server 2008 R2 is a comprehensive database platform. The integrated business intelligence tools (BIDS) can provide data management solutions for enterprise effectively. SQL Server 2008 R2 provides data integration service (SSIS), data analysis service (SSAS), reporting service (SSRS) and other services in business intelligence. SSIS can integrate the business data demanded by BI from a variety of heterogeneous data sources, and SSAS can construct a cube, then OLAP realizes the data mining. The SSRS can make the data analysis results to meet various users' needs in the form of report or graph. SharePoint2010 provides various technologies to display the report in the Web interface. The Reporting Services Web Part of SharePoint2010 can directly integrate and show the reports in the portal and the business data catalog (BDC) can extract and display the background application data [2].

2. Solutions

2.1 The Overall Architecture

The informatization of retail business is related to the Enterprise Resource Planning (ERP), the Supply Chain Management (SCM), the Customer Relationship Management (CRM), and the Sales Management System(SMS) systems. On the basis of the business process of retail enterprise we integrate the system data of ERP, CRM, SCM and SMS as business data sources and use the SQL Server 2008 R2 to clean, extract, transform and create data warehouse. With OLAP online analytical processing technology to realize the data mining. Furthermore, we take use of the existing information resources and visual display technology to realize the display and analysis of relevant business model, which can provide scientific decision in respect of sales, finance, inventory, purchasing and other business units.

The overall architecture is shown in Fig. 1:



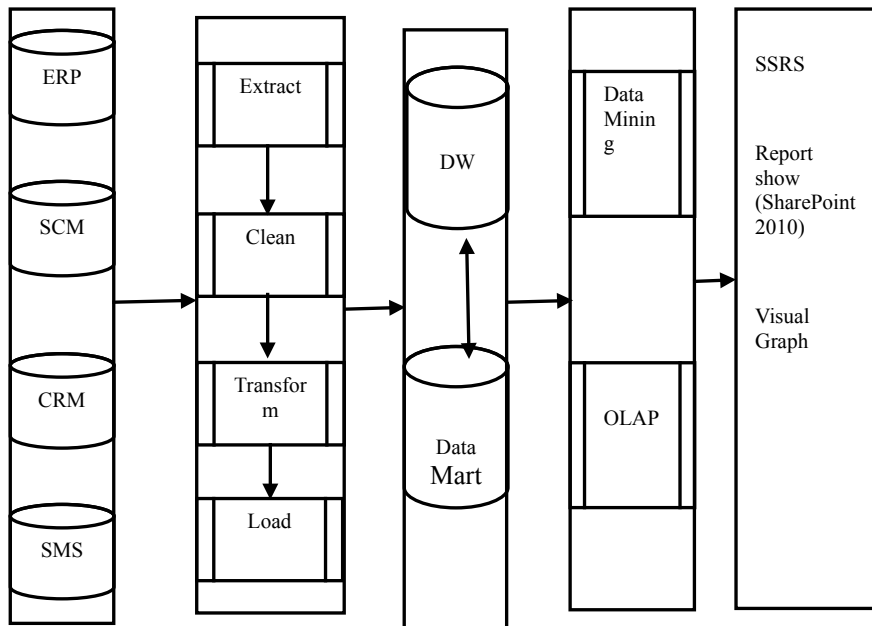


Figure 1: The Overall Architecture

2.2 Cleaning and Transformation of Data

Import the data stored in the retail enterprise database into SQL Server 2008 R2 and use the SSIS service to complete data extraction, transformation and loading. SSIS features a powerful capability of data cleaning and loading. It is the main tool of Microsoft to help enterprise realize ETL[3]. As the data obtained from SQL Server 2008 R2 are just raw data, if you want to load these data into DW, you have to solve the following two questions:

(1) Data cleaning

In order to make the analysis proceed smoothly, the data need to be cleaned before it is loaded into the data warehouse. SSIS contains some internal conversion, which can be added to the package to clean up and standardize the data; secondly, SSIS can make similar values in the data mart to divide them to one group to clean the data. For example, may duplicates can be identified by means of comparing the address of a customer in the customer record.

(2) Data transformation

In general, the data in the business database is represented by code, which can save the storage space of the database on the one hand and enhance the adapt ability of database design on the other hand, for example, the large-scale ERP system UF and SAP. Since the users may come from various parts of company, not all the users can understand the meaning of the data in the data mart; therefore it needs to use the transformation component of SSIS to transform the business data into the value that is easy to be understood before the business data is loaded to data warehouse.

In Fig. 2, we can see the process of ETL, the data source may be flat file source or OLE DB source. The SSIS offers service to transform, seek and import the data into OLE DB target.

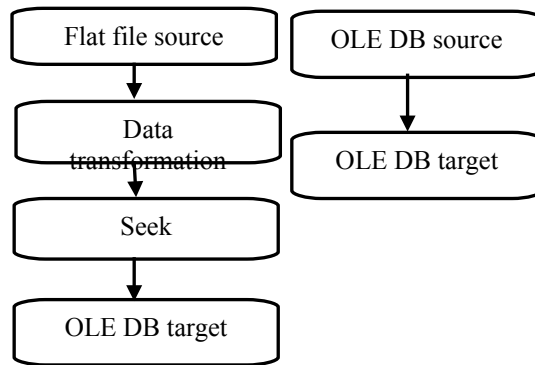


Figure 2: Process of ETL

2.3 Data Analysis

After the process of ETL, a data warehouse is established and we can analyze the data as demanded. SSAS is the tool of SQL Server 2008R2 for data analysis. In this paper, we can regard customers' consumption as the main line to analyze the customers' age, gender, consumption habits and correlated consumptions [4].

The structure of tables in the the data warehouse is shown in Fig. 3:

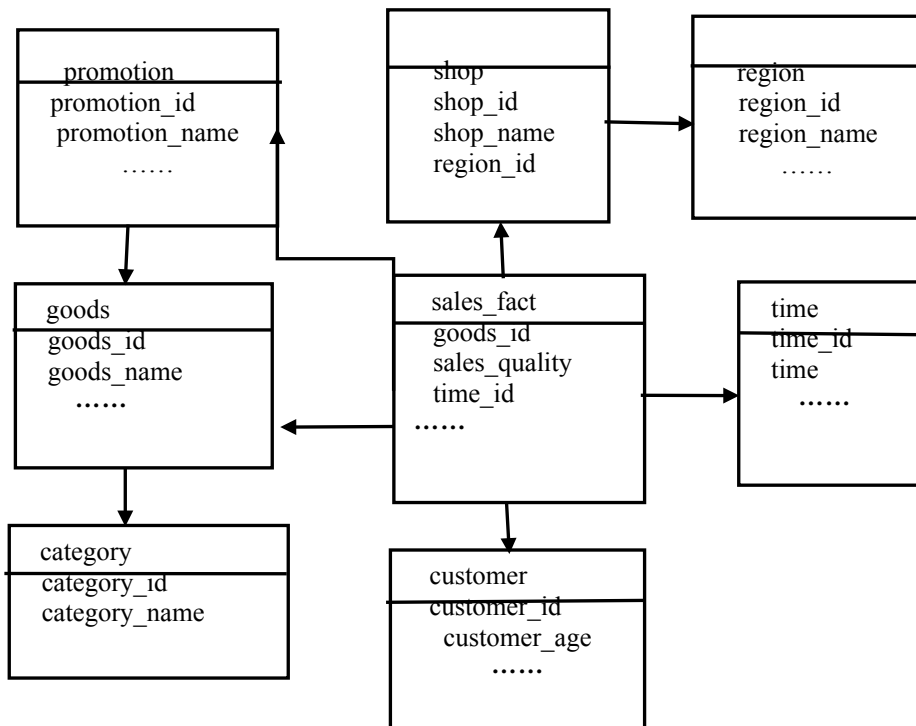


Figure 3: Structure of Tables in the Data Warehouse

In this paper, we use Apriori association rules algorithm to analyze the data in the data warehouse. Apriori is the most representative association rules algorithm. By iteration, we can find the frequent item sets in the data warehouse and take use of the frequent item sets to construct the rule of minimum confidence:

Step1: find the 1-item sets and regard it as L_1 , if it can't find the 1-item sets, it will stop the algorithm..

Step2: use the L_1 to find the candidate item sets C_2 , and then make judgment as to each component of C_2 , find 2-item sets namely L_2 .

Step3: find more k-item sets and then count k-item sets by means of loop operation.

Step4: judge whether all the rules are suitable to the minimum confidence. If it is suitable, the rule is right.

Generally, the association rules are large databases or data system. The expression of the association rule is $A_1 \wedge A_2 \wedge \dots \wedge A_m \Rightarrow B_1 \wedge B_2 \wedge \dots \wedge B_n$, and the $A_k(k=1,2,\dots, m)$, $B_j(j=1,2,\dots,n)$ are the support of data item in the database, namely, the percentage that one transaction includes A and B. The formula is $Support(A \Rightarrow B) = P(A \cup B)$. Confidence is the percentage that one transaction contains A and also contains B, that is, the conditional probability $Confidence(A \Rightarrow B) = P(A|B)$. The association between data items is called data association. According to some items in one transaction, we can deduce some other items that will appear in the same transaction. Dig out the association between the data is the purpose of the association analysis. In this way, it can help the retail enterprise look for the potential users to adjust the sales decisions.

For example, according to the original data in Table 1, the date represents the time, Goods_id represents the product ID, Goods_name represents the goods' name, the price represents each goods' price, and the amount represents the quantity that goods are sold at the same time. By using the Apriori algorithm integrated in the SSAS to analyze the data, we can know the probability of each commodity in the sales record and also the probability when a commodity appears in one sale record and another commodity appears at the same sale record; therefore we can know what categories of goods can be regarded as bundling-sale. Taking the shampoo and toilet paper for example, we can see the result in Table2:

No.	Date	Ticket_id	Goods_id	Goods_name	Price	Amount
1	2014-5-4	9900000984	06001	Shampoo	26.5	1
2	2014-5-4	9900000984	10002	Toilet paper	31	1
3	2014-5-4	9900000984	00741	Toothpaste	8.9	1
...

Table 1: Original Data Table 1

consequent	antecedent	support%	confidence
shampoo	toilet paper	8.845	21.631
toilet paper	shampoo	9.207	30.012

Table 2: Result of Association Analysis

According to these results, the following conclusions can be drawn: 8.854% of the total customers buy toilet paper, and the customers who purchased the shampoo accounted for 21.631%. 9.207% of the total customers buy shampoo and the customers who purchased the toilet paper account for 30.012%. It can be concluded that shampoo and toilet paper can be deemed as bundling-sale.

Compared with traditional analysis methods, only partial data is distinguished to a few dimension with the majority not distinguished between dimensions. This analysis method can make data distinguished between dimensions, and this method can make the data analysis more accurate; besides, in traditional analysis methods, the amount of data that can be analyzed is restrained to a small amount; however this method can analyze large amount of data to meet the need of a big department store supermarket.

2.4 Report Generation and Display

2.4.1 The Report Generation

POS(CENet2015)018

SQL Server 2008 R2 integrates the component SQL Server Reporting Service to make it possible that the right information can be passed to the right people in any commercial environment. The Reporting Service, as a complete platform that is deployed on the server, can be adopted to create, manage and deliver traditional and interactive reports; at meanwhile, the standard modular design of the product, the API enables the software developers, the data providers and the enterprises to integrate the legacy systems or the third-party reports [5-6], for example, the original data table shown in Table 2 can be used to make Reporting Services reports shown in Fig.4. The sales results can be offered to the manager. From Fig. 4, we can see the daily sales volume of each commodity so that the manager can know how the goods are sold very day.

Product_name	2010/1/4	2010/1/5	2010/1/6	2010/1/7	2010/1/8		Total
Shampoo	¥21,900.00	¥8,250.00	¥11,700.00	¥2,550.00	¥11,100.00		¥55,500.00
Toilet paper	¥12,300.00		¥12,180.00	¥16,200.00	¥21,900.00		¥62,580.00
Toothpaste	¥2,850.00	¥3,300.00	¥13,350.00		¥8,850.00		¥28,350.00
Total	¥37,050.00	11,550.00	¥37,230.00	¥18,750.00	¥41,850.00		¥146,430.00

Figure 4: Display of Report Result

2.4.2 Report Display

SharePoint2010 provides the work flow extensions based on Microsoft SharePoint platform for business users, who can quickly and easily design task forms and business processes without writing code; therefore, it can help business users to further expand the content management of SharePoint platform. In addition, the report server integrated with the instance of SharePoint2010 can make the report items stored in the content database of SharePoint2010 so that we are able to view these reports directly in the content database of SharePoint2010 via SharePoint2010 web parts [7].

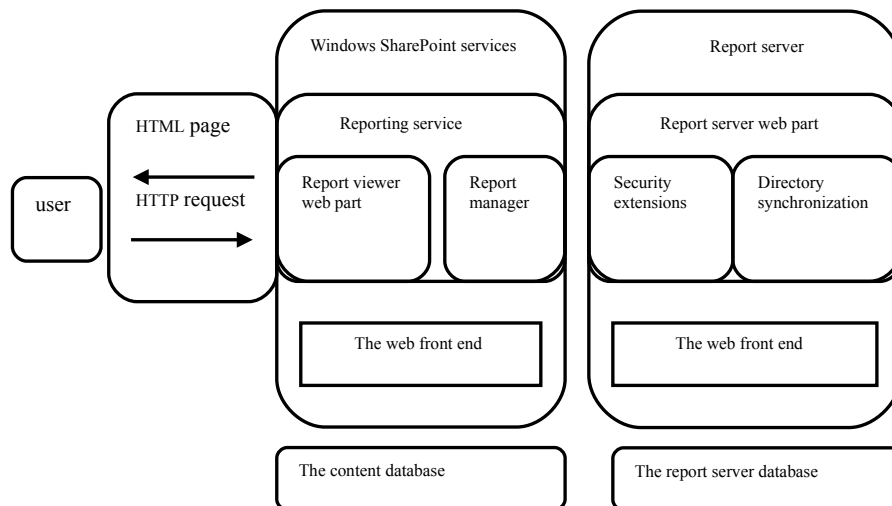


Figure 5: Related Server Components

From Fig. 5, we can see the process that a report is displayed by the SharePoint2010. When opening the report from SharePoint2010 site, the report server endpoint will perform the following actions: connect to the report server, create a session, process the report, retrieve the data and show the report in the report viewer web parts. When the report is open, the report can

POS (CENet2015) 018

be exported in different forms. The report server and SharePoint2010 operate and synchronize the data, and track the information how the report server deals with the file. If we modify the attribute of items or set of any server, it will be recorded in SharePoint2010 server, for example, the report shown in Fig. 4 is stored in SharePoint2010 database. The result that we view the report via SharePoint2010 web parts is shown in Fig.6.

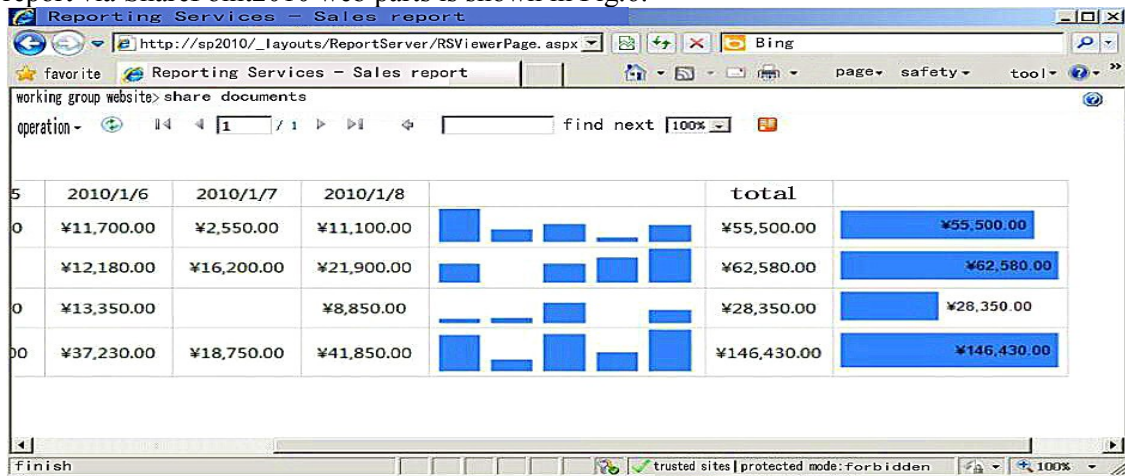


Figure 6: Report View via SharePoint2010 Web Parts

As a good collaborative work platform, the Share Point 2010 can make the report displayed to each manager of enterprise simply, which enables the management of the enterprise to be efficient.

3. Comparison with Traditional Methods

As to the traditional analysis method, only partial data is distinguished to a few dimension with the majority not distinguished between dimensions. No dimensions distinguishing being made or the dimension distinguishing being inconsistent will affect the accuracy and effectiveness of analysis of related indexes. In this paper, we put forward a solution which can distinguish the data in many dimension flexibly, in this sense, it can strengthen the accuracy of data analysis; besides, this solution can help the enterprise manager clearly know the indicators of sales, which will be helpful for the enterprise's long-term development. In addition, when compared with the traditional methods, this analysis method is more easy to be used because the traditional analysis is hard to analyze the data that is constantly changing, such as the retail enterprise's data; nevertheless, this method can fulfill this need.

4. Conclusion

SQL Server 2008 R2 is a comprehensive database platform. Its integrated business intelligence tools (BIDS) can provide the data management solutions for enterprises effectively. SharePoint2010 provides work flow extensions on the basis of Microsoft SharePoint platform for the business users, who can quickly and easily design task forms and business processes without any writing code. In this paper, we combine SQL server 2008 R2 and SharePoint 2010, and then put forward a new solution about business intelligence, which has been implemented. The solution can be adopted as a reference by retail enterprises to realize its intelligent management in terms of business.

References

- [1] Bughin J, Chui M, Manyika J. *Clouds, big data, and smart assets: Ten tech-enabled business trends to watch*[J]. McKinsey Quarterly 56(1): 75-86,2010.

- [2] D. L. Zhu. *SQL Server 2005 Data Mining and Business Intelligence complete solution scheme*[M], Electronic Industry Press: Beijing, China, 45-47,2007.(in chinese)
- [3] Michael Milton. *Head First Data Analysis*[M], Electronic Industry Press: Beijing, China,pp.78-79,2010.(in chinese)
- [4] Robert Laberge.*The Data Warehouse Mentor: Practical Data Warehouse and Business Intelligence*[M],China Machine PRESS:Beijing, China,93-94,2012.(in chinese)
- [5] Hilbert M,Lopez P.*The world's technological capacity to store,communicate,and compute information*[J],Science,332(6025): 60—65,2011.
- [6] Hai Wang, Shouhong Wang, *A knowledge management approach to data mining process for business intelligence*[J].Industrial Management & Data Systems, 108 (5): pp.622-634, 2008.
- [7] Kevin Laahs, Emer Mckenna.*Microsoft SharePoint Technologies Planing, Design and Implementation* [M].Elsevier Digital Press:Massachusetts,107-110,2005.