

Application and research on business intelligence in audit business

Lei Che^{1,2,a}, Xiaoping Yang^{1,b*} and Fan Jiang^{3,c}

¹ Renmin University of China, China

² Beijing Information Science & Technology University, China

³ University of Pittsburgh, America

^achelei@bistu.edu.cn, ^byang@ruc.edu.cn, ^cramily_jf@outlook.com

Abstract. With audit data requirements, taking for example the audit data of construction industry, the audit data integration requirements are analyzed and application architecture of business intelligence technology in the audit business is put forward. The modeling process and the populating process of data warehouse are detailed, and the scheme of data pre-processing is set out. The report display methods are proposed, such as proportion chart, tendency chart, index plots and etc. The application model and value of business intelligence technology in audit business are analyzed, which is based on data mining technology such as outlier detection. This research has been applied to National Twelfth Five-Year Plan Science and Technology Support Program and has achieved better analysis and prediction of audit.

Keywords: Business intelligence, Audit business, Data warehouse, Multidimensional data analysis.

1 Introduction

With the development of information technology, audit data is becoming increasingly large and complex, and the original data management pattern has formed large and small islands of data. Audit trails have been overshadowed by the growing complexity of information systems and mass business data, so that variety storage form of data sources need to be analyzed by auditors, which is very difficult to be realized with the traditional data query and statistical analysis methods. Beyond that, many frauds have also noted the importance to maintain the balance of some financial ratios, the role of audit gradually weakening due to artificial control important ratio. With the development of the computer, the auditor can use computer to write audit plan, audit analysis, and carry out information systems audit. But the traditional data analysis method cannot deal with the huge database system, outdated technology tools is bound to affect the breadth and depth of audit. As far as application research of computers concerned, with the advent of big data, business intelligence and other analysis-oriented and decision-oriented techniques have emerged, which will undoubtedly not only provide new ideas and methods for modern audit, but also provide a new way for the development of auditing information. Only if business intelligence technology is applied to audit business, a structured data warehouse and multidimensional data set being built, auditors can organize data and analysis data from different dimensions, even exploit valuable experience and rules, to provide decision support for the audit work.

2 Business intelligence technology

2.1 Basic concepts

Business intelligence refers to using data warehouse and data mining techniques to store and manage the customer data, and using a variety of data statistical analysis tools to analyse customer data and provide all kinds of analysis report, such as customer evaluation, service evaluation, marketing evaluation, the future market demand and so on, which provide decision making information for the enterprise²³. Business intelligence technology includes data warehouse technology, multidimensional data analysis technology and data mining technology.

In computing, a data warehouse (DW, DWH), or an enterprise data warehouse (EDW), is a system used for reporting and data analysis⁴. Integrating data from one or more disparate sources creates a central repository of data, a data warehouse (DW). Data warehouses store current and historical data and are used for creating trending reports for senior management reporting such as annual and quarterly comparisons⁵. Multidimensional analysis is a data analysis process that groups data into two or more categories: data dimensions and measurements. Data mining (sometimes called data or knowledge discovery) is the process of analysing data from different perspectives and summarizing it into useful information that can be used to increase revenue, cuts costs, or both⁶. Data mining software is one of a number of analytical tools for analysing data. It allows users to analyse data from many different dimensions or angles, categorize it,

and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

2.2 Software selection

There are a lot of BI products. However, according to the size, market awareness and development momentum, the products provided by SAP, Oracle, IBM, Microsoft, SAS, the five large multinational BI provider, are dominating the market, and are also the pacesetter for BI. As one of the major platform providers in the industry, Microsoft has provided an integrated suite of business intelligence products in the SQL Server, including related data warehouse, data analysis, data mining, data integration, development, and management tools. The main three tools for BI architecture are: SQL Server Integration Services and SQL Server Analysis Services and SQL Server Reporting Services. At the same time, Microsoft's BI products relate closely to other products, and provide a complete solution from the data store files to the deployment server display, in which some commonly used Office software are also. Therefore, this study selects Microsoft's BI products.

3 Data warehouse model design

3.1 Audit data integration requirements

There are more than 20 industries in audit business. And there are financial data, business data and etc., which is related to audit business.

In audit business, auditors can understand cash flow condition, operation capacity, debt paying ability through analyzing enterprise financial situation, so that auditors can evaluate lasting management ability of enterprises and find doubtful points of audit. Audit financial data mining and analysis are the decision-making tools to provide correct information for auditors. Therefore, auditors can understand the enter-prise profit and debt paying ability, so that they can discover the financial problems of enterprises and take corresponding measures in time. There are some typical case of the computer audit work in big data environment, such as quality analysis of credit assets, financing platform risk analysis, two sets of mort-gage analysis, capital abnormal flow analysis and so on. Decision support scheme based on audit business data can not only provide important decision support for the enterprise, but also brings the convenience for the auditors. Audit business has strong demand for decision support solutions, because auditors need to do a lot of analysis work on financial statements. The construction of data warehouse is the key step to realize decision support.

3.2 Business intelligence technology in the application of audit services architecture

As Figure 1 shows, we can see that the application architecture of business intelligence technology in audit business includes Data Sources, ODS, ETL, DW, DM, OLAP server, Data Analysis, Data Mining and so on.

The data of Data Sources come from the data of industries involved by audit business, such as financial data, enterprise data and so on. These data format can be structured, unstructured, or semi-structured. Taking for example the audit data of the construction industry, this paper illustrates the building of data warehouse model, data analysis based on data warehouse and data mining methods. The public data come from the official website, whose data format is an EXCEL file. The industry data includes cash flow table, profit table, asset and liability table and financial indicator table. In order to make up for the data synchronization different between the business system and Data Warehouse, Operational Data Store (ODS) is inserted between the Data source and Data warehouse. ETL is the tool to extract data, transfer data and load data from Data source to ODS and from ODS to DW.

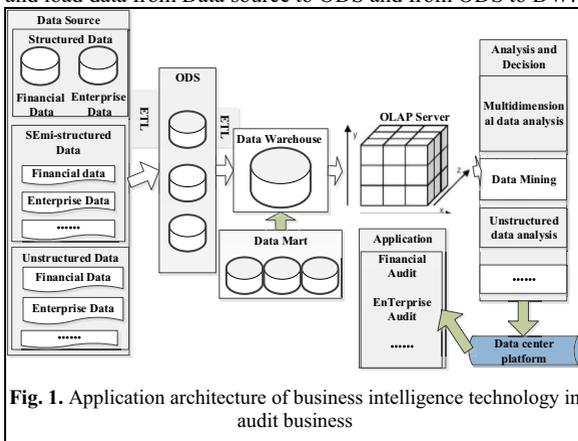


Fig. 1. Application architecture of business intelligence technology in audit business

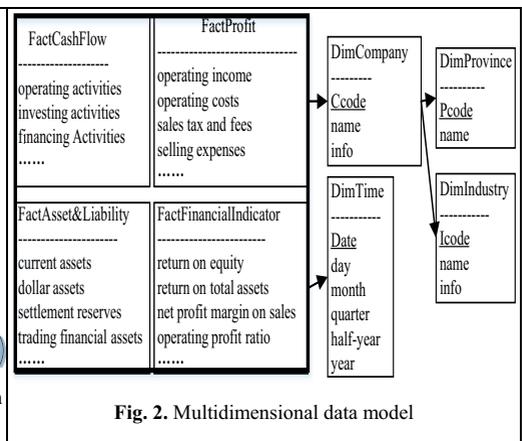


Fig. 2. Multidimensional data model

4 Data warehouse construction

4.1 Multidimensional data model construction

Data warehouse includes construction determining topics, dimensions tables, fact tables and populating the data into data warehouse or data market.

Determining the topic is to determine the topic of data analysis or reports showing. In specific audit process of the construction industry, the following results should be gotten based on existing industry data:

(1) Enterprises of every industry, the amount of the major issues money of the industry, the proportion of assets of the enterprise in its industry

(2) The trend diagram comparing the asset-liability ratio with the average of the enterprise;

(3) The information about the enterprise profits, cash flow indicators, operating income and etc.

In view of the financial audit data gathered from the construction industry, the topics which need to be studied and analyzed include the following 4 categories: "financial indicators of the construction industry", "the balance sheets of the construction industry" and "the cash flow of the construction".

Dimension refers to the different analysis angles. Taking for example the topic of construction industry, we should study financial indicators through time dimensions, enterprise dimensions, industry dimensions, province dimensions and etc.

This model of data warehouse is based on the Fact Groups Model, in which dimensions Star Models or Snowflakes Models connect dimensions Fact Tables through the same dimension tables such as time dimension, enterprise dimension and so on. From Figure 2 we can see the local conceptual model of the data warehouse according the collected data, among which there are 4 dimension tables and 4 fact tables. The dimension tables are DimTime, DimProvince, DimCompany and DimIndustry. The fact tables are FactCashFlow, FactProfit, FactAsset&Liability and FactFinancialIndicator.

When the dimension table is used to achieve dimension, the dimension level need to be represented, by the following three methods: by the different fields of dimension table, by association between the tables and by the associated fields. In those representations, the representation based on the fields is the simplest, and has higher query efficiency, because the queries are single-table queries. Representing the dimension levels by different tables can extract many common dimension tables, improve the maintainability of dimension tables, which has low query efficiency because the query requires multiple correlation between tables. The third method can represent uneven and irregular dimensions well. The "DIMTime" table is the application of the first method.

4.2 Populating the data warehouse

Populating data warehouse refers to data extracting, data transformation and data loading, namely the ETL (Extract - Transform - Load). SSIS is Microsoft SQL Server Integration Services, which is an integration solution generating high performance data, and the ETL platform provided by Microsoft. The ETL process will be discussed based on the collected data of the construction industry. The data source is from Excel. Every indicator corresponds to an excel file, at the same time, every sheet in every excel file corresponds to an enterprise. ETL steps are as follows:

(1) Accessing Excel file, and deleting meaningless data by the data cleansing work. Normalizing every fields' name of Excel files according to the corresponding mapping relationship, and designing corresponding relationship between the excel files and the database.

(2) Importing all sheet of all excel files into the local temporary tables by SSIS graphical operation. That's to say, if there are 4 indicator tables and 50 enterprises in every indicator table, there are 200 local temporary tables would be imported data into.

(3) Emerging the same indicator information of all enterprises into one permanent indicator tables using Procedure of the database.

4.3 Data preprocessing

Multidimensional data analysis needs to analyse data with different comprehensive degrees, therefore the details data should be integrated. Data integration is a very time-consuming operation. In order to shorten query response time, OLAP uses precomputed technology, by which the whole or part data will be calculated and stored in advance. During the financial analysis of the audit business in the, the auditor should not come to a conclusion not just by the information of cash flow, but should combine with assets & liabilities and profit. Only in such way, the audit can make more comprehensive and correct evaluation on business activities. Therefore, the following data models are the example ones which should be precomputed.

Data Model 1: $\frac{FactCashFlow.[operating\ activities\ net\ cash\ flows]}{FactAsset\&Liability.[enterprises\ equity]}$

Function Information: Representing the ability of share capital generating cash and the ability of the largest corporate dividend distributions.

Data Model 2: $\frac{FactCashFlow.[operating\ activities\ net\ cash\ flows]}{FactAsset\&Liability.[total\ assets]}$

Function Information: Representing the ability of the total assets generating cash.

Data Model 3: $\frac{FactCashFlow.[operating\ activities\ net\ cash\ flows]}{FactAsset\&Liability.[current\ liabilities]}$

Function Information: Representing the ability of cash acquired in business activities repaying short-term debt.

Data Model 4: $\frac{FactCashFlow.[operating\ activities\ net\ cash\ flows]}{FactAsset\&Liability.[full\ liabilities]}$

Function Information: Representing the ability of cash acquired in business activities repaying full debt.

Data Model 5: $\frac{FactCashFlow.[operating\ activities\ net\ cash\ flows]}{FactProfit.[net\ profit]}$

Function Information: Representing how much cash will support when enterprises achieve the profit of one yuan book and the superior and inferior quality of corporate profits to some extent.

Data Model 6: $\frac{FactCashFlow.[cash\ received\ from\ sales\ of\ goods\ or\ rendering\ of\ Services]}{FactProfit.[major\ business\ income]}$

Function Information: Representing the corporate sales recovering cash situation and the quality of corporate sales.

5 Showing report

Once the data warehouse is built, the user can access and query the information and data by data analysis tools, real-time query and reporting tools, and the user can see variety of reports of data analysis results, as shown in Figure 3. We can make the visual solutions according to business requirements. There are 3 type reports designed for the current commands: proportion charts, trend charts and indicator charts. The pie charts are suitable for showing the proportion charts, the line charts are suitable for showing the trend charts, and the bar charts are suitable for showing the indicator charts. In SQL Server BI solutions, the SSRS can be used for showing reports to the users.



Fig. 3. Multidimensional reports of one corporate assets & liabilities information



Fig. 4. Reports showing the total operating income

If the report shown in the Figure 4 need to be created, the corresponding procedure in SQL Server and the reporting service should be created and executed or deployed. The procedure is created as follows:

--Procedure: Showing the total operating income

```
CREATE PROCEDURE TOI_Pro
@Year int, @CompanyName nvarchar(255)
AS
```

```
SELECT SUM(FactProfit.opratingincome) AS TotalOpratingIncome, Max(DimCompany.name) AS Company,
Max(DimProvince.name) AS Province, YEAR(FactProfit.Date) AS Year, MONTH(FactProfit.Date) AS Month
FROM FactProfit JOIN DimCompany ON FactProfit. Ccode = DimCompany. Ccode JOIN DimProvince ON
DimCompany. Pcode = DimProvince. Pcode
WHERE YEAR(FactProfit.Date)=@Year and DimCompany.name=@ CompanyName
GROUP BY YEAR(FactProfit.Date), DimCompany.name, MONTH(FactProfit.Date)
```

The parameters of the report model are showed as follows:

Connection Character String: Data source=localhost;Initial Catalog=AuditDemo

```
DataSet1: TotalOperatingIncome DataSet2: YearList
DataSourceName: AuditDB DataSourceName: AuditDB
Query Type: Procedure Query Type: Text
Procedure Name: TOI_Pro Query: SELECT DISTINCT YEAR(Date) AS Y FROM DimProfit ORDER BY Y DESC
DataSet3: CompanyList
DataSourceName: AuditDB
Query Type: Text
Query: SELECT DISTINCT name FROM DimCompany
```

6 Data mining

Data Mining refers to gaining the knowledge in which people are interested from the data of large databases. These information is implicit, previously unknown and potentially useful. Data mining technology is a static computer data analysis method, which obtain implication modes in the data internally by analysis of the static data, such as classification, Clustering, Association, etc. And these modes can help auditors to analyse vast amounts of electronic data of the audited entities, and find the audit trail to determine audit doubts. Data Mining techniques and methods would enhance the ability of auditors to analyse audit data, and increase efficiency and quality of the audit work. Therefore, Data Mining technology will get more applications and development in the field of audit. The Data Mining models SQL Server supporting are: Decision Tree, Bayesian Classification, Association Rules, Clustering Analysis and Sequential Cluster and so on.

Some problems in Computer Audit can be solved based on these Data Mining models 910.

(1) Based on clustering technology, auditors can carry out data grouping operation, and then master the rules among data of each group and differences among data of different groups, so that auditors will find the characteristics of the checked problem.

(2) Based on the classification and prediction technology, auditors can find the inner link between input data and output data from the massive amounts data of audited units, then use the found rules to analyse the historical data. Finally, audits can compare the predicted value and actual value, and lock the auditing emphasis more quickly and easier, so as to find doubtful points of audit more correctly.

(3) Based on association analysis, auditors can extract frequent pattern of data, so as to sum up the two events or items rules hidden in the data. But these rules can only be precondition of the next judgement.

(4) Based on outlier detection technology, auditors can mine the isolated points from the data, so as to improve the probability of finding abnormal data, find possible regulation violating phenomenon.

(5) The audit corpus of every industry audit can be built based on Text Mining technology, which is used for processing of the field of every industry auditing language, management information, thereby supporting the optimization of auditing methods.

7 Conclusions

With the rapid development of economy and computer technology, the operation, management and accounting mode of audited units are experiencing outstanding changes, more and more dependent on efficient, complicated and huge information system. In the face of huge amounts of data, it is necessary for auditors to introduce BI technology into auditor process. It's the new requirements, which help auditors to carry out auditor work in big data environment, and provide auditors auditor idea and audit key identification. This research has been applied to National Twelfth Five-Year Plan Science and Technology Support Program ("The research and development and application demonstration of audit simulation and simulation platform to national audit immune system") and has achieved better analysis and prediction of audit.

Acknowledgments

This work was financially supported by these projects as following: Beijing Higher Education Young Elite Teacher Project (NO. YETP1503), The Scientific Research Project of Beijing Educational Committee (NO. KM201511232016 and SM201511232004), Teaching Reform Project of Beijing Information Science & technology University (NO. 2015JGYB30), National Natural Science Foundation (NO. 61272513).

References

1. Isabel Pedrosa, Carlos J. Costa, Computer Assisted Audit Tools and Techniques in Real World: CAATT's Application and Approaches in Context, *Computer Information Systems and Industrial Management Applications*, 4(2012)
2. Elbashir M.Z., Collier P.A., Sutton S.G., The role of organizational absorptive capacity in strategic use of business intelligence to support integrated management control systems, *The Accounting Review*, **86**, 1(2011)
3. Chaudhuri S., Dayal U., Narasayya V., An overview of business intelligence technology. *communications of the ACM*, **54**(2011)
4. Jindal R., Taneja S., Comparative study of data warehouse design approaches: a survey. *International of Database Management Systems*, **4**,1(2012)
5. QU X. , Multi-granularity Temporal and Spatial Approximate Aggregate Query on RFID Data Warehouse, *Computer Science*, **39**, 6(2012)
6. Han Jiawei, Micheline Kamber, Jian Pei, DATA MINING: Concepts and Techniques, Third Edition.Beijing: Machinery Industry Press, **84**, 92-99(2012)

7. R. Uday Kiran, P. Krishna Reddy, Improved approaches to mine rare association rules in transactional databases. Proceedings of the Fourth SIGMOD PhD Workshop on Innovative Database Research, (2010)
8. Gyorgy J. Simon, Vipin Kumar, Peter W. Li, *A simple statistical model and association rule filtering for classification*, *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2011)
9. Lalit Wadhwa, Dr. Virender pal, Forensic Accounting and Fraud Examination in India, *International Journal of Applied Engineering Research*, **11**(2012)
10. Roger S. Debreceeny, Glen L. Gray, Data mining journal entries for fraud detection: An exploratory study, *International Journal of Accounting Information Systems*, **11** (2011)