Design of Data Cubes and Mining for Online Banking System

Dr. Harsh Dev
Professor, Department of CSE,
Pranveer Singh Institute of Technology
544-555, NH-2, Kalpi Road, Bhauti
Kanpur, U.P., India

Suman Kumar Mishra
Research Scholar, Department of Computer Science,
Bhagwant University, Ajmer, Rajasthan.
India – 226021

ABSTRACT

Combating with immense competition now a day's banks are focusing upon customer satisfaction rather than merely rendering their services in the Indian banking industry. Due to the tough competition in banking sector, paradigm changes are seen in this sector. Various innovations are witnessed in this field, because banking sector is adopting innovative ways to improve their services & to win the faith of their customers.

Today, banking sector is bombarded with a large number of innovative facilities, which comprises of: Centralized banking system, Internet banking, Mobile banking, SMS Alert, Smart card, RTGS, E-banking, ATM and various other such wonderful facilities, which make the work faster & easier.

This paper presents decision support in banking sector which link up the strengths of both OLAP and Data Mining. The main objective of this paper is to develop enhanced model for banking sector for improving the efficiency and to check the emergence & creation of innovative ways in this field.

Keywords

Data Mining, OLAP, Data Cubes and Decision Support System.

1. BACKGROUND

Introduction: The basic objective of the database is to provide and share information anywhere at any time. The Banking system in many countries is reorganizing the communication through Internet as the way of transaction. As technology evolved new computerized decision support applications were developed and studied. Researchers used multiple frameworks to help in building and understanding these systems. Today one can organize the history of DSS into five broad DSS categories namely: communication-driven, data-driven, document- driven, knowledge-driven and model-driven decision support systems.

Our study focuses on Banking system based on decision support applications, data-oriented systems, multidimensional data analysis, query and reporting tools. All of these technologies have been used to support Banking based decision making.

1.1 Decision Support System

A Decision Support System (DSS) is used to improve organizational performance by share and apply databases to make optimal decision in real time. Our covered three sectors: On-Line Analytical Processing (OLAP), Application of data mining in Banks and Decision Support System. It is revealed that organizations generally use OLAP rather than OLTP to build DSS [33, 11 - 13, 17, 18].

1.2 Data Warehousing

A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process [1]. The dimension represents the database applications designed according to the needs of the individual user and it may also have descriptive attributes such as name of the customers, type of account etc. In the 21st Century, the data has increased manifold which improved the performance of the database and degraded the related application. The data warehousing is that the data stored for business analysis can most effectively be accessed, separating it from the data in the operational system [2]. Today's data warehousing system support very sophisticated on – line analysis including multi- dimensional analysis. Data mining is a technology that provides sophisticated analysis for applications such as data warehouse [3, 4, 5, 6]. The aim of data warehouse is to build a systemic data storage environment and separate plentiful data needed by analysis and decision-making from traditional operation condition. It makes transfer dispersive and disaccord operation data to integrated and uniform information [10].

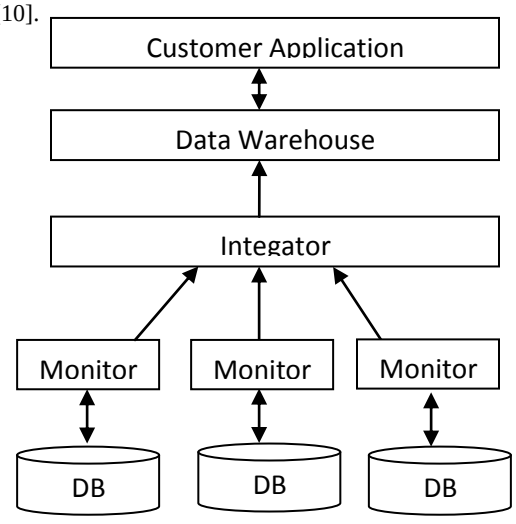


Fig 1: Basic system structure of data warehouse

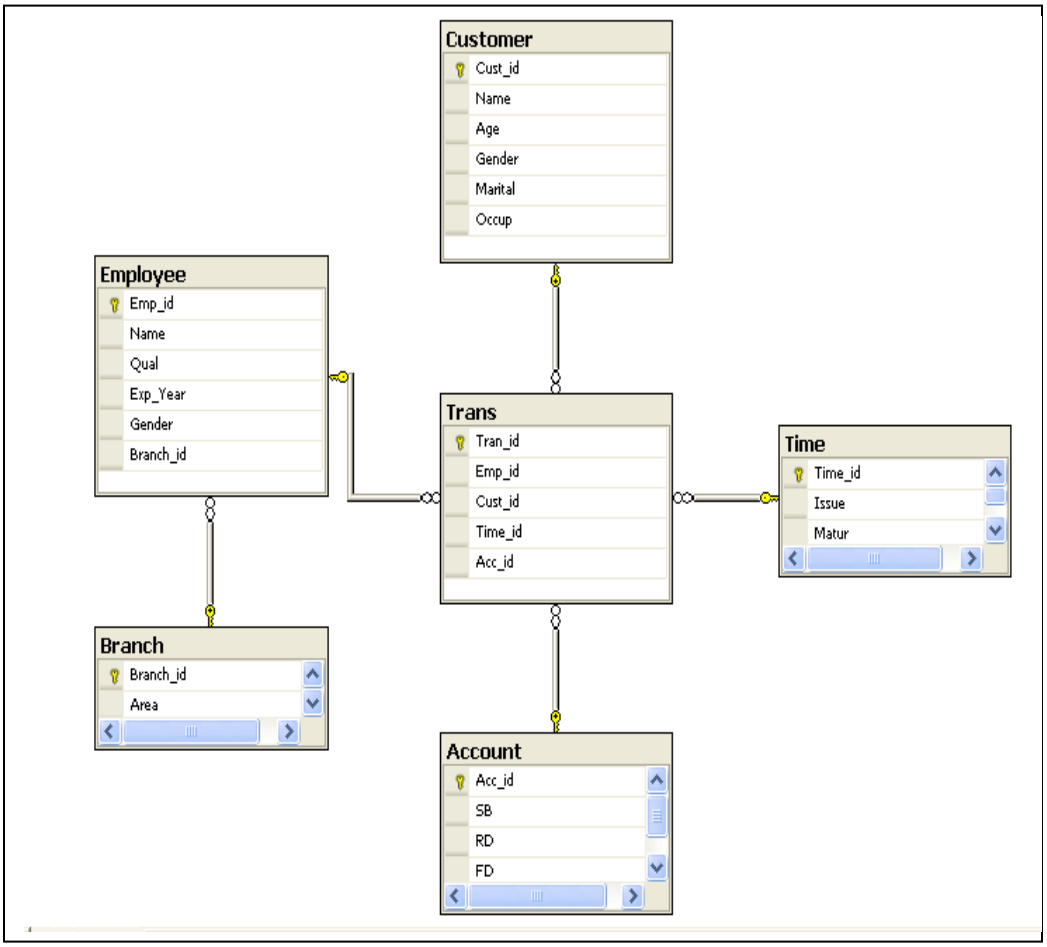


Fig. 2: ER Diagram

The trend now is to employ combined approaches, Liu-Yunfeng, Wang-Xiaohui, Zhai-Dongsheng [19] used data mining to identify Exceptional Client from the large number of bank customers to prevent abnormal customer's risks effectively. Bin Fang, Shoufeng Ma [20] used Data Mining Technology and its Application into CRM of commercial banks that can make classification, clustering and association analysis of customer behaviors according to set models to covert the data into useful information to support operating decision making. Maytham Safar and Abdullah Al-Najja [21] explained the Data Growth in Banking Sector. Zhao Li Ping, Shu Qi Liang [22] used Data Mining Application in Banking-Customer Relationship Management.

Qiuju Yin, Ke Lu [23] attempted to accomplish reduction analysis on credit evaluation index system of bank customer based on data mining method, including cluster method and decision tree method. Tony Spiteri Staines [24] Used Timed Petri Net (TPN) to Model a Bank ATM. Palaniappan, Sellappan , Ling Chu a Sook [25] analyzed Clinical data to Decision Support using OLAP with data mining. Stanley Y.W. Su, Sanjay Ranka [27] analyzed the Performance of Parallel Query Processing Algorithms for Object-Oriented Databases.

C. The Present Work: This paper presents decision support in banking sector which link up the strengths of both OLAP and Data Mining. The main objective of this paper is to develop enhanced model for banking sector for improving the efficiency and performance of databases.

2. PROPOSED MODEL FOR ONLINE BANKING SYSTEM

2.1 Star Schema:

On-Line Transaction Processing (OLTP) systems mainly support for recording based business transactions. They record information represent in two dimensions and automate repetitive tasks. Structured Query Language (SQL) is typically used to access information and results are presented in the form of reports which bankers use to make transactional decisions. Fig. 2 & 3 shows a simple Entity-Relationship Diagram (ERD) and SQL Command consisting six tables.

```
select c.Name,c.Age from customer c, employee e,
branch b, time t,account a, trans tr where
c.cust_id=tr.cust_id and tr.emp_id=e.emp_id and
tr.time_id=t.Time_id and tr.acc_id=a.acc_id group by
c.Name,c.Age
```

Fig 3 SQL Command

2.2 About Star Schema:

OLTP has a major drawback. Large amounts of data in normalized form require many joins even to answer simple queries. For example, to analyze relationships between Bank and Customer (Fig. 3), the query would require several table scans and multi-way table joins which can degrade performance significantly [11]. A real-life database will have many tables and the time taken to process the joins will be unacceptable.

On-Line Analytical Processing (OLAP) was introduced to overcome this problem, whereas OLTP uses two dimensional tables, OLAP uses multidimensional tables called data cubes. OLAP can analyze data in several dimensions. It also incorporates query optimization [11- 13]. Users can filter, slice-and-dice, drill-down and roll-up data to search for relevant information efficiently.

Data mining (defined as a process of nontrivial extraction of implicit, previously unknown, and potentially useful information from the data stored in a database) has now become an important research tool [13- 15]. It can discover hidden knowledge in the data and identify problems along several dimensions [15, 16].

This paper presents a Banking decision support system using OLAP and data mining that can answer complex questions which are not possible by using data mining alone.

2.3 Hierarchical Model

As OLAP uses several preprocessing operations such as data cleaning, data transformation, data integration, its output can serve as valuable data for data mining [13, 28]. OLAP operations (e.g., drilling, dicing, slicing, pivoting, filtering) enable users to navigate data flexibly, define relevant data sets, analyze data at different granularities and visualize results in different structures [29 - 31]. Applying these operations can make data mining more exploratory. The motivation for an integrated model, OLAP with data mining, is the concept hierarchy. Data in OLAP and decision tree are organized into multiple dimensions where each dimension contains multiple levels of abstraction defined by the concept hierarchy [30, 32]. analyze data at different granularities and visualize results in different structures [29 - 31]. Applying these operations can The concept hierarchy is illustrated in Fig. 4, where each member has one root and all members between roots have customers and every branch ends with a leaf member. OLAP data cubes which store concept hierarchies can be used to induce decision trees at different levels of abstraction [32, 11]. Once the decision tree mining model is built, the concept hierarchies can be used to generalize individual nodes in the tree, which can then be accessed by OLAP operations and viewed at different levels of abstraction.

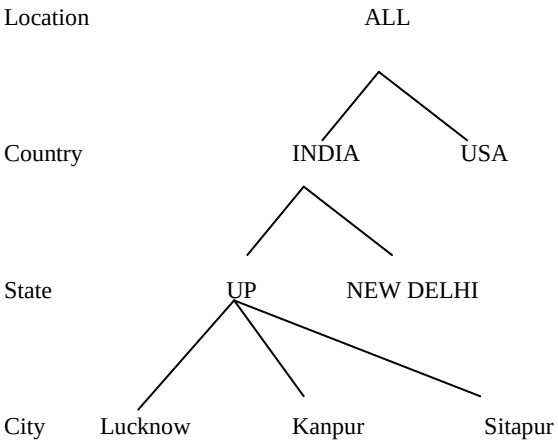


Fig 4: A concept hierarchy for different location

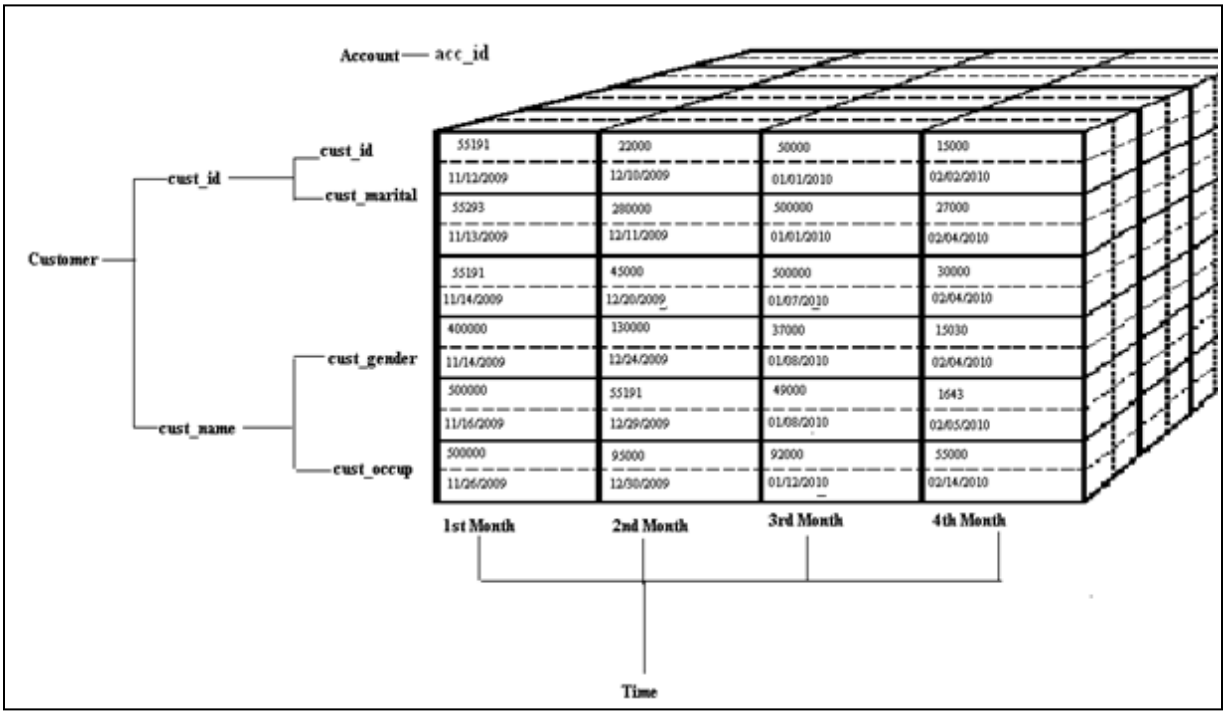


Fig 5: A logical view of OLAP cube

2.4 Data Cubes

For implementing the system a data cube is first created then the data mining process is started. The cube stores the information and allows browsing at different conceptual levels. It serves as the data source for the data mining task. Data mining can be performed on any level or dimension of the cube. After the model is built it is stored in the OLAP cube. Each dimension represents the rule corresponding to a node in the decision tree mining model (Fig.5)

OLAP operations explain the different states of the system. The data for this study is taken from Cooperative Bank Databases. The data comprise of Saving Account database, Recurring Deposit database and Fixed Deposit database. As the data is declassified, we have added several dummy attributes such as Employee, Customer and Branch Information.

A two-year sample dataset (2009 and 2010) is created to mine for knowledge discovery. Information on entities and their attributes and relationships are fed into the data warehouse.

3. IMPLEMENTATION OF DATA CUBES

This paper explains, how the combination of OLAP and Data mining effectively provide advanced decision support. Which is not possible using OLAP or data mining alone. The following sample queries as given below are support for this objective.
Fig. 6 shows the framework of the combination (OLAP with data mining) comprising of many units. The framework can be divided into two different parts: *Serverside* – for makinng the combination and *Client-side* – for fetching queries and displaying results (Fig. 6). It uses OLAP operations (slice, dice, roll-up, drill-down, pivot). The test data validates the effectiveness of the model.

A. Sample Query I. How integration of OLAP with data mining enhancing real time indicator?

B. Sample Query II. How integration of OLAP with data mining used for improving visualization to expose trends which can be neglected or missed?

C. Sample Query III. How integration of OLAP with Data Mining disclose more concerned trends in data over capabilities provided by OLAP or Data Mining alone?

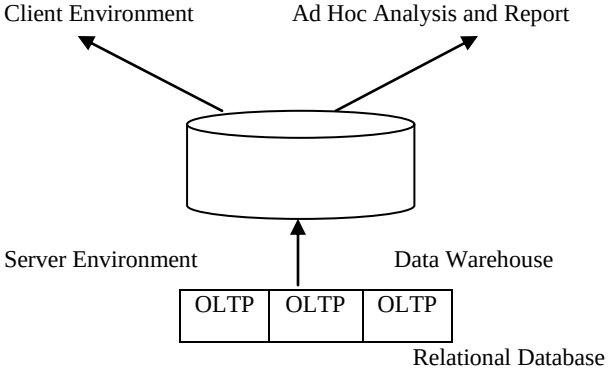


Fig 6: Integration of OLAP with data mining architecture

Results of Implementation

The system can predict the future state and generate useful information for effective decision-making. It can answer all the sample queries given above.

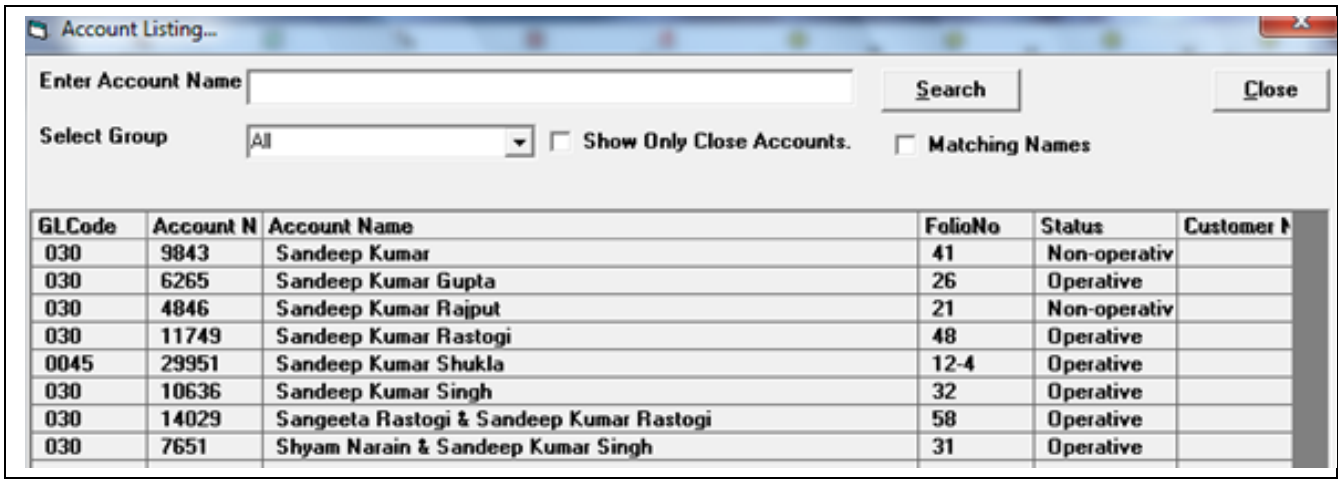
Result Query I. The integrated model enhances real time indicators by using information on bank branch utilization for saving account customers. It allows bank administrators to discover any bottlenecks that might exist. It allows them to solve problems related to bank branch utilization. . The administrator can use this information to allocate credit cash based on their characteristics. This indicator is useful for performing "what-if" analysis on credit cash limits.

Result Query II. The integrated model improves information visualization. It discovers overall trends that are likely to be missed by using OLAP or data mining alone. It provides a more comprehensive analysis and facilitates decision-making by allocating bank employees to under-performing branches/geographical areas. This in turn allows the quality of employees/services in under-performing branches to be improved. It facilitates and enhances the employee performance as the various facets of employee and data interface are harmonized.

Result Query III. With data mining, bank employees can predict customers who might be analyzed with reports. OLAP provides a focused answer using historical data. However, combining by these two, we can optimize existing processes and discover more elusive patterns, for example, by analyzing customers' yearly statement of accounts. System

Experimental Results: We have implemented the Online Banking System using data cubes. The system used is a Intel core i3 – 330 M, with 3 GB DDR3 RAM, running Microsoft Windows XP Pack 3 and SQL Server 2005. The system was implemented using Microsoft Visual Basic 6.0 and Active X Data Model (ADO).

Following figure 7, shows screen shot for the query result of our experimental data which is implemented in the form of OLAP cubes.



GLCode	Account N	Account Name	FolioNo	Status	Customer N
030	9843	Sandeep Kumar	41	Non-operativ	
030	6265	Sandeep Kumar Gupta	26	Operative	
030	4846	Sandeep Kumar Rajput	21	Non-operativ	
030	11749	Sandeep Kumar Rastogi	48	Operative	
0045	29951	Sandeep Kumar Shukla	12-4	Operative	
030	10636	Sandeep Kumar Singh	32	Operative	
030	14029	Sangeeta Rastogi & Sandeep Kumar Rastogi	58	Operative	
030	7651	Shyam Narain & Sandeep Kumar Singh	31	Operative	

Fig 7: Screenshot of query result

4. CONCLUDING REMARKS

From the above work, it can be concluded that for large amount of databases, data cube design & implementation technique is suitable for faster search of data within a fraction of second. Cubes store large amount of banking data which can be used by the administrator/customer, who can search the desired record online in an efficient way. In future, the above work can be enhanced to permit employees to solve the queries with help of data cubes on banking related problems and automatically convert these problems to Multi Dimensional based queries & can also use complex queries, multimedia and spatial database.

5. ACKNOWLEDGMENTS

I especially extend my sincere and grateful thanks to *Prof.* ***Vipin Saxena****, Head of the Department of Computer Science Department,* Baba Saheb BhimRao Ambedkar University, Lucknow, Uttar Pradesh, India.

6. REFERENCES

[1] W. Inmon. Building the Data Warehouse. John Wiley & Sons, 2002.

[2] Alexis Leon, Enterprise Resource Planning (Second Edition), chapter 2, Page no. 73, TMH Publication, 2008.

[3] Berson, A, Data warehousing, Data mining and OLAP, McGrawHill, 1997

[4] Michael, L.G. and Bel, G.R, Data mining-a powerful information creating tool, OCLC systems & services, Vol.15:2, 1999, pp81-90.

[5] Robert S.C. Joseph A.V and David B., Microsoft Datawarehousing, John Wiley & Sons, 1999.

[6] Margarent, A.H. and Rod, H, Facilitating corporate knowledge: building the data warehouse, Information & computer security, Vol 5:5, 1997, pp 170-174.

[7] Daniel. J. Power: Decision Support System Concepts and Resources for managers, 2002, page 19 -24.

[8] Manel Mora, Guisseppi Forgionne, and Jatinder N.D. Gupta: Decision Making Support Systems: Achievements, Trends & Challenges for the new decade.

[9] Frada Burstein and Clyde W. Holsapple, Decision support systems in context, Published online: 29 February 2008.

[10] Feng Lei, Chen Hexin, Analysis Methods of Workflow Execution Data Based on Data Mining, Second

[11] Surajit, C. and Umeshwar, D., An Overview of Data Warehousing and OLAP Technology, ACM Sigmod Record, *26*(1), 65-74, 1997.

[12] Ralph, K. and Margy, R., The Data Warehouse Toolkit. The Complete Guide to Dimensional Modeling (2nd ed.), Canada: John Wiley & Sons, Inc, 2002.

[13] Torben, B.P. and Christian, S.J., Multidimensional Database Technology, IEEE Computer, *34*(12), 40-46, 2001, December.

[14] Usama, M. F., Data Mining and Knowledge Discovery: Making Sense Out of Data, IEEE Expert, 20-25, 1996, October.

[15] Ming-Syan, C., Jiawei, H. and Philip, S.Y., Data Mining: An Overview From a Database Perspective, IEEE Transactions on Knowledge and Data Engineering, *8*(6), 866-883, 1996, December.

[16] Usama F., Data Mining and Knowledge Discovery in Databases: Implications for Scientific Databases. Proceedings of the 9th International Conference on Scientific and Statistical Database Management (SSDBM '97), Olympia, WA., 2-11, 1997.

[17] Surajit, C., Umeshwar, D., and Ganti, V., Database Technology for Decision Support Systems, IEEE Computer, *34*(12), 48-55, Dec. 2001.

[18] Sarwagi, S., Explaining Differences in Multidimensional Aggregate, Proceedings of the 25th International

Conference on Very Large Data Bases, Scotland, United Kingdom, 42-53, September, 1999.

[19] Liu-Yunfeng, Wang-Xiaohui,Zhai-Dongsheng, Commercial Banks Exceptional Client Distinguish Based on Data Mining, 2010 International Forum on Information Technology and Applications.

[20] Bin Fang, Shoufeng Ma [20], Data Mining Technology and its Application in CRM of Commercial Banks, First International Workshop on Database Technology and Applications, 2009.

[21] Maytham Safar and Abdullah Al-Najjar, Data Growth in Banking Sector, 17th International Conference on Database and Expert Systems Applications (DEXA'06), 2006.

[22] Zhao Li Ping, Shu Qi Liang [22], Data Mining Application in Banking-Customer Relationship Management, International Conference on Computer Application and System Modeling (ICCASM 2010), 2010.

[23] Qiuju Yin, Ke Lu [22], Data Mining Based Reduction on Credit Evaluation Index of Bank Personal Customer, International Conference on Future Information Technology and Management Engineering, 2010.

[24 Tony Spiteri Staines, Using a Timed Petri Net (TPN) to Model a Bank ATM. 13th Annual IEEE International Symposium and Workshop on Engineering of Computer Based Systems (ECBS'06), 2006.

[25] Palaniappan Sellappan, Ling Chu a Sook, Clinical Decision Support Using OLAP With Data Mining, IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.9, September 2008

[26] Arie van Deursen, Tobias Kuipers, Identifying Objects using Cluster and Concept Analysis, ICSE '99 Los Angeles, ACM, 1999.

[27] Stanley Y.W. Su, Sanjay Ranka, Performance Analysis of Parallel Query Processing Algorithms for Object-Oriented Databases, IEEE Transactions On Knowledge And Data Engineering, Vol. 12, No. 6, November/December 2000.

[28] Fayyad, U., Gregory, P.-S. and Smyth, P., From Data Mining to Knowledge Discovery in Databases, AI Magazine, *37*(3), 37-54, 1996.

[29] Parseye, K., OLAP and Data Mining: Bridging the Gap.Database Programming and Design, *10*, 30-37, 1998.

[30] Han, J., OLAP Mining: An Integration of OLAP with Data Mining, Proceedings of 1997 IFIP Conference on Data Semantics (DS-7), Leysin, Switzerland, 1-11, October, 1997.

[31] Han, J., Chiang, J.Y., Chee, S., Chen, J., Chen, Q., Cheng, S. & et al., DBMiner: A System for Data Mining in Relational Databases and Data Warehouses, Proceedings of the 1997 Conference of the Centre for Advanced Studies on Collaborative research, Ontario, Canada, 1-12, November, 1997.

[32] Han, J., Kamber, M., Data Mining Concepts and Techniques, San Diego, USA: Morgan Kaufmann Publishers, pp. 294- 296.

[33] Helen, H. and Peter, H., Using OLAP and Multidimensional Data for Decision Making, IEEE IT Professional, 44-50, 2001, October.