

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Performance Efficient Mining on an OLTP System Using Schema Enhancement Method

Muhammad Hasan Rasheed, Muhammad Imran Saeed and M. Sikandar Hayat Khiyal
Department of Computer Science, International Islamic University Islamabad, Pakistan

Abstract: Mining on an OLTP System is possible but the cost of running Mining Module on a Conventional OLTP Architecture is not practical. The response time decreases and the Normal Functioning of the System disturb. Schema Enhancement Method is proposed to overcome these disadvantages in OLTP System. After Modifying Conventional OLTP to Schema Enhanced OLTP Structure, the mining module runs very fast as compare to the Conventional OLTP. The idea is proposed because installation of DSS Solution is not possible for all the users. The cost and resource factors are very large in DSS Solution. Schema Enhancement Method take advantage over the Conventional OLTP is that it supports the DSS (to some extent) and the cost of the System is also low. This solution is the combination on OLTP (for running Real Time business) and DSS (for running Analysis/Mining module) together. Performance Efficient System takes advantages of storing the Summary Data inside the actual Schema of the OLTP and the Query Time for Mining the Data decrease.

Key words: Online transaction processing, online analytical processing, automated finger identification system, decision support system, data mining

INTRODUCTION

The demand of OLTP environment is increasing day by day in the developing countries like Pakistan. The Analysis Module (Data Mining) is an important part of any Software Architecture that tells the owner about the growth of the business and the future demands. OLAP is actually designed to meet the demands and needs of the Decision making or DSS (Decision Support Systems). The cost of implementing OLAP with DSS Module is not possible for a Small level business that is running on the total Human resource of 5-6 people and have limited resources of Hardware and not very Technical persons therefore there is a need for a System that is running on an OLTP architecture and supports Mining but the cost of the system is not exceeding the boundaries of the small level company or shop.

The idea of Mining on an OLTP system has been in research for many years. Han *et al.* (1994) and Chiang *et al.* (1997) have developed prototype systems for Mining in Databases System. In the Review by Rehm *et al.* (2002) some scholars have supported the idea but the implementation has not been done by them and it was made open to design such a system that should be an OLTP but have some support of Mining. There is of course not possible to have a intelligent Mining that is present in a OLAP or DSS or Warehouse Solutions

however the idea is to make such a system or to modify the existing system that will make minimum changes in the architecture of the OLTP system.

Riedel *et al.* (2000) presents a new idea of making the system intelligent and strong that helps mining on a OLTP system. The word nearly for free is an important part of the title which means that the 100% results will not be returned from this idea however by scheduling the disk requests that takes advantage of the ability of high-level functions to operate directly at individual disk drives and thus the Mining data that will be required in future by the Software will be fetch from the disk head before it is required.

SCHEMA ENHANCEMENT METHOD

We present the idea of Schema Enhancement for OLTP Systems. The basic OLTP architecture will remain the same and software will work fine with this new enhancement. This modification is only required by those Systems willing to include the Mining Module in the Systems.

Figure 1 shows the Conventional OLTP System. A Database Server and a client machine that access the Database Server. OLTP (Online Transaction Processing) runs on real time environment and Server machine accepts the requests from the clients and process the queries. If

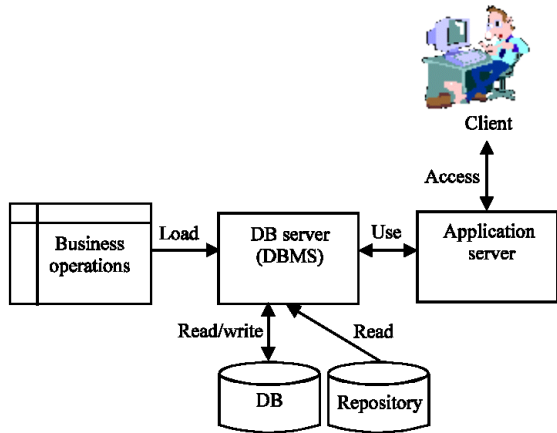


Fig. 1: Conventional OLTP system

we run complex queries the system down time will increase and response time will decrease resulting in bad image on the company.

In Fig. 2 OLAP or DSS system architecture is shown. This system is designed to run complex queries. High level server machines and Technical staff are present to handle high level queries. Here the response is very fast but buying such a system is not affordable and similarly running cost of these systems is very high.

In Fig. 3 Schema Enhancement Method is shown. The system is an OLTP based system. During the conversion process, which will run at the time of System Design or any other time of the System, a new Mining supported Schema will be included in the system main architecture. This Schema will be the main source of Mining and the Mining data will come from the actual Database.

The data for Mining Module will come either using the DBMS build in trigger process or user developed software that will feed the Mining Module according to the requirements of the Company. Incase of new requirements the same procedure will be repeated once for the old data and regular for the new data.

The normal Relational DBMS process will apply on this system:

- Analysis
- Requirement Gathering
- Design of ERD
- Normalization Process
- System Design
- Database Structure
- Implementation and Testing

However in Requirement Gathering and System Design, new phases will be included for supporting Mining Module.

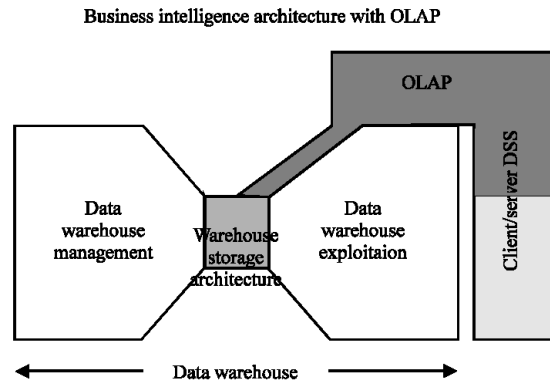


Fig. 2: OLAP or DSS system architecture

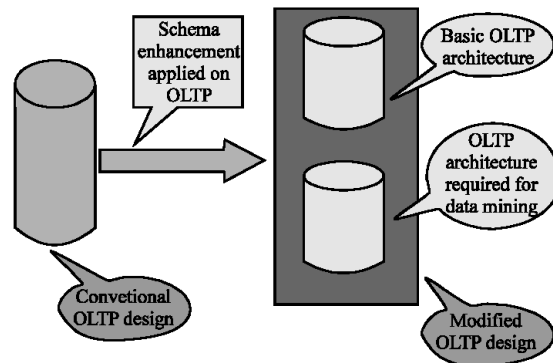


Fig. 3: Schema enhancement method

- Analysis
- Requirement Gathering of Mining
- Design of ERD
- Normalization Process
- System Design
- System Design for Mining Module
- Database Structure
- Implementation and Testing

CASE STUDY (NADRA DATA ENTRY PROJECT)

National Database and Registration Authority (NADRA) a Government Registration Organization working under the Ministry of Interior, Government of Pakistan with its Headquarters in Islamabad, Pakistan. This IT Organization is responsible for the registrations of Pakistani Citizens living in Pakistan and Abroad. It has a Dataware House of operating capacity of 20 Tera bytes of Space with latest NCR Machines. The Dataware house is designed to work in extreme environments, running heavy queries over the Data of approximately 5 crore citizens at present.

NADRA Swift Registration Center (NSRC) is the registration office working in every Tehsil of Pakistan with total of approx 300 centers in the country. The basic concept is to process the applicant record from the NSRCs, collect the data from the whole country, Consolidate and Load in the Central Dataware house after applying business rules to the Data including Facial and AFIS (Automated Finger Identification System) checks.

The working environment of the NSRC is SQL Server 2K with Windows 2K Server and a 2-Tier Software Application working in a Local Network between Clients (more than one) and Server (only one). There is a huge amount of Data present in the NSRCs after 2-3 working years making it difficult to run even small queries at Local level. The server response time decreases and the normal working of the whole system stops.

Schema Enhancement is applied on the existing System without making any changes in the Registration Software.

Note: All the values are taken in real time environment (Working Environment)

- [1]: Time taken on conventional OLTP to get Daily Statistics
- [2]: Time taken on conventional OLTP to get Monthly Statistics
- [3]: Time taken on conventional OLTP to get Yearly Statistics
- [11]: Time taken on Schema Enhancement Method to get Daily Statistics
- [12]: Time taken on Schema Enhancement Method to get Monthly Statistics
- [13]: Time taken on Schema Enhancement Method to get Yearly Statistics

- [111]: Time taken to Shift Daily Data to OLTP Analysis DB
- [112]: Time taken to Query Daily Data from OLTP Analysis DB
- [113]: Time taken to shift Monthly Data from OLTP Analysis Daily Table to OLTP Analysis Monthly Table
- [114]: Time taken to query Monthly Data from OLTP Analysis DB
- [115]: Time taken to shift Yearly data from OLTP Analysis Monthly Table to OLTP Analysis Yearly Table
- [116]: Time taken to query Yearly Data from OLTP Analysis DB

PERFORMANCE EFFICIENT SCHEMA ENHANCEMENT METHOD

Performance Efficient System is the demand of every business of the 21st century. Schema Enhancement method (Table 1) as compared to the conventional OLTP (Table 2) is Performance Effective Solution. The Query Time Decrease and the efficiency of the System increase by this method. As from the Table 3-B, it is clear that the Query Time for running Daily Report is approx 6 minutes however in Schema Enhancement Method the Total time is approximately 1 min 15 sec (for transferring the Data from Live DB to the Summary Tables) and 4 sec (for Query the Summary Table to get the final results). The Transfer times decrease for the Monthly and Yearly Data because when the Summary Data is present in the Tables then the Monthly and Yearly Summary Data is generated from the Daily Summary Data instead of querying the Full Tables.

Table 1: Sample size

Record set	Live database	19968
	Backup database	4043648
	Schema enhancement DB	4319

Table 2: Conventional OLTP

System name	Daily [1]	Monthly [2]	Yearly [3]
1.6 GHz Centrino Mobile	Full Scan 6 min 30 sec	Full Scan 5 min 3 sec	Full Scan 4 min 42 sec
800 MHz Intel Original	2 min 24 sec	2 min 12 sec	2 min 17 sec
Xeon 3.20 GHz	37 sec	21 sec	22 sec
1.70 GHz Intel Original	3 min 55 sec	4 min 26 sec	4 min 40 sec

Table 3: Schema enhancement method

System name	Daily [11]		Monthly [12]		Yearly [13]	
	One time [111]	Table scan [112]	One time [113]	Table scan [114]	One time [115]	Table scan [116]
1.6 GHz Centrino Mobile	1 min 15 sec	4 sec	3 sec	5 sec	3 sec	4 sec
800 MHz Intel Original	18 sec	Less than 1 sec	2 sec	Less than 1 sec	2 sec	1 sec
Xeon 3.20 GHz	5 sec	Less than 1 sec	2 sec	Less than 1 sec	2 sec	Less than 1 sec
1.70 GHz Intel Original	16 sec	3 sec	3 sec	3 sec	3 sec	3 sec

CONCLUSIONS

Schema Enhancement Method has shown very positive results on this Implementation. We are hopeful that the results on other systems will also be very practical. This process is very easy to implement and the normal structure of the system is not affected by the change.

ACKNOWLEDGMENT

We are thankful to National Database and Registration Authority for providing us the Information we have mentioned in our paper. We are also thankful to Dean and Head of Computer Science Department, International Islamic University Islamabad, Pakistan for providing us the working environment and Resources for completing our research.

REFERENCES

- Chiang, J.Y., J. Han, S. Chee, J. Chen and Q. Chen, 1997. DBMiner: A system for data mining in relational databases and data warehouses. In: Proc. CASCON'97: Meeting of Minds, Toronto, Canada, November.
- Han, J., Y. Fu, Y. Huang, Y. Cai and N. Cercone, 1994. DBLearn: A system prototype for knowledge discovery in relational databases. ACM SIGMOD Record.
- Rehm, C., J. Oates and D. Marco, 2002. One Database Model for OLAP and OLTP. Published in DM Review Online, DMReview.com.
- Riedel, E., C. Faloutsos, G.R. Ganger and D.F. Nagle, 2000. Data mining on an OLTP system (nearly) for free. ACM SIGMOD international conference on Management of data archive. Dallas, Texas, United States ISSN: 0163-5808.