

# Survey of Literature for Data Warehouse and OLAP Technology for Banking System

**Arpita Mathur**

Department of Computer Science,  
Lachoo Memorial College of Sc. & Tech. A-Sector,  
Jodhpur, India

**Abstract :-** Data warehousing is a collection of decision support technologies, aimed at enabling the knowledge worker (executive, manager, analyst) to make better and faster decisions. It serves as a physical implementation of a decision support data model and stores the information on which an enterprise needs to make strategic decisions. For the purpose of effective study the scope has been made on OLAP technology used in banking sector. The research is focused on data warehouse and OLAP technology with its special contribution in banking sector. In this paper focus is on survey of literature for datawarehouse and OLAP technologies used in banking system.

**Keywords:** Data warehouse, OLAP, Banking, Data Cube,

## 1. Introduction

Data warehousing and on-line analytical processing (OLAP) are essential elements of decision support, which has increasingly become a focus of the database industry. Many commercial products and services are now available, and all of the principal database management system vendors now have offerings in these areas. Decision support places some rather different requirements on database technology compared to traditional on-line transaction processing applications.

Over the years, researchers have proven that when it comes to data warehouse and OLAP technology it means we are storing data, reuse the data and retrieve the data for a long period of time. In banking sector the OLAP technique is basically used for accessing information at anytime and anywhere. Now days, banks are storing more information than ever before. Therefore the decision makers must have the right information at right time to help them to make more informed and intelligent decision to generate their reports.

The review of literature is subdivided into four parts: review associated with Data Warehouse, banking technology and OLAP and research gap.

### 1.1 Review Associated With Data Warehouse

**Robert M. Bruckner et al.** identified intelligent and comprehensive data warehouse systems as a powerful instrument for organizations to analyze their business. The implementation of such decision systems for an enterprise-wide management and decision support system can be very different from traditional software implementations. Because data warehouse systems are strongly data-driven entities, hence their development process is highly dependent on their underlying data. Since data warehouse systems concern many organizational units, the collection of unambiguous, complete, verifiable, consistent and usable requirements can be a very difficult task. Use cases are generally used as standard notation for object oriented requirement modeling. In this paper the team showed how use cases can enhance communication between stakeholders, domain experts, data warehouse designers and other professionals with diverse backgrounds. Three different abstraction levels namely business, user and system requirements of data warehouse requirements were introduced and discussed.

In a new study conducted by **R.G. Little and M.L. Gibson** in this study, have surveyed data warehouse implementation prospective. Project participants were consulted to record their perceptions, which may further contribute to the data warehouse implementation process. The respondents included: functional managers/staff, IS

managers/staff, and consultants. The study identified eight significant factors that according to participants may impact data warehouse implementation.

Furthermore, **Hugh J. Watson and Thilini Ariyachandra** conducted this research for analyzing the success factor of different data warehouse architectures. Following were the main objectives of this study:

- a) Understanding the factors that influence the selection of data warehouse architecture.
- b) The success of the various architectures. The academic and data warehousing literature and industry experts were used to identify architecture selection factors and success measures. Information attained was then to create questions for a Web-based survey that was used to collect data from 454 companies about the respondents, their companies, their data warehouses, the architectures they use, and the success of their architectures. The success metrics identified were information quality, system quality, individual impacts, organizational impacts, development time, and development cost.
- c) Perhaps the most interesting finding of this study is the almost equal score of the bus, hub and spoke, and centralized architectures on the information and system quality along with the individual and organizational metrics. It helps explain why these competing architectures have survived over time they are equally successful for their intended purposes. Based on these metrics, no single architecture is dominant, but still hub and spoke architecture seems to be the first choice among organizations implementing data warehouse.

Again a research was conducted by another researcher **J. Gray et al.** which stated that evaluating database system performance often requires generating synthetic databases which have certain statistical properties but filled with dummy information. When evaluating different database designs, it is often necessary to generate several databases and evaluate each design. As database sizes grow to terabytes, generation often requires longer time than evaluation. This paper presents several techniques for synthetic database generation. In particular it discusses parallelism to get generation speedup and scale. Congruential generators to get dense unique uniform distributions and Special-case discrete logarithms to generate indices concurrent to the base table generation.

In like manner **Xintao Wu, et al.** in his paper states that testing of database applications is of great importance. A significant issue in database application testing consists in the availability of representative data. This paper investigates the problem of generating a synthetic database based on a-priori knowledge about a production database. The approach followed was to fit general location model using various characteristics (e.g., constraints, statistics, rules) extracted from the production database and then generate the synthetic data using model learnt. The generated data was valid and similar to real data in terms of statistical distribution, hence it can be used for functional and performance testing. As characteristics extracted may contain information which may be used by attacker to derive some confidential information about individuals, hence the paper presents a disclosure analysis method which applies cell suppression technique for identity disclosure analysis and perturbation for value disclosure.

## 1.2 Review Associated With Banking Technology

**Balwinder Singh and Pooja Malhotra (2004)** identified banks were not making full utilization of internet to disclose information to stakeholders.<sup>15</sup> Accounting organizations and standard setters are also paying attention to the rapid growth of dissemination of accounting information on the internet. The International Accounting Standards Committee (IASC) commissioned a discussion paper for business reporting on the internet. This IASC report provided a survey of web-based financial reporting practices of the 660 public corporations in 22 countries. The study concluded that a significant number of companies in many countries use the web for communication of business performance to stakeholder

**According to King (2001)** the Internet is an inflection point that has an impact that has not yet been fully determined. He also pointed out that the most important effects on cost management practices have been in the areas of information communication and transaction processing. This would imply that companies which are using the internet for accounting and financial information presentation may see additional benefits over additional costs by providing financial reports on a website.

**Costas Lapavitsas and Paulo L. Dos Santos (2008)**, argued money-dealing transactions have become cheaper, but investment costs have increased. According to the World Bank (2003) report on ICT and the Millennium

Development Goals, IT reduces transaction costs per customer and enables banks to provide small loans and services to a larger number of rural customers.

**Sivakumaran (2005)**, believes that adoption of technology has led to the following benefits: greater productivity, profitability, and efficiency; faster service and customer satisfaction; convenience and flexibility; 24x7 operations; and space and cost savings.

Also **Berger (2003)**, stated the usage of information technology (IT)xx broadly referring to computers and peripheral equipment, has seen tremendous growth in service industries in the recent past. The most obvious example is perhaps the banking industry, where through the introduction of IT related products in internet banking, electronic payments, security investments, information exchanges, banks now can provide more diverse services to customers with less manpower.

**Lapavitsas and Paulo L. Dos Santos (2008)**, argued money-dealing transactions have become cheaper, but investment costs have increased. According to the World Bank (2003) report on ICT and the Millennium Development Goals, IT reduces transaction costs per customer and enables banks to provide small loans and services to a larger number of rural customers.

Additionally, **Shirley J. Ho and Sushanta K. Mallick(2008)** believed that IT can improve bank's performance in two ways viz., IT can reduce operational cost (cost effect), and facilitate transactions among customers within the same network (network effect). On the other hand, **Baker and Berenblum (1996)**, identified investment in IT is one of the major factors determining the success or failure of organizations.

### 1.3 REVIEW ASSOCIATED WITH OLAP TECHNOLOGY

According to the study of **Mr.P.P. Karde and Dr. V.M.Thakare**. Quick response time and accuracy are important factors in the success of any database. In large databases particularly in distributed database, query response time plays an important role as timely access to information and it is the basic requirement of successful business application. According to them a data warehouse uses multiple materialized views to efficiently process a given set of queries. But the materialization of all views is not possible because of the space constraint and maintenance cost constraint. They discuss an overview of various techniques that are implemented in past recent for selection of materialized view i.e by competitive A\* algorithm, improvement algorithm and shuffled frog leaping algorithm. The issues related to maintaining the materialized view are also discussed.

According to the study of **Dr Harsh Dev and Suman Kumar Mishra** "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. 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. They discuss presents decision support in banking sector which link up the strengths of both OLAP and Data Mining. The main objective of them is to develop enhanced model for banking sector for improving the efficiency and to check the emergence & creation of innovative ways in this field. They discuss the implementation of data cube and mention how the combination of OLAP and Data mining effectively provide advanced decision support. This is not possible using OLAP or data mining alone. They 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.

According to the study of **Vipin Saxena and Ajay Pratap** in the current scenario, the size of database related to any organization is rapidly increasing and due to evolution of the object-oriented approach, many of the Software Industries are converting the old structured approach based software's into the object-oriented based software's. Therefore, for the large amount of database, it is necessary to study the faster retrieval system as On-Line Analytical Processing (OLAP) which was introduced by E.Codd in 1993. They discuss in there paper an attempt in this direction and object-oriented approach with the help of Unified Modelling Language (UML) is used to create OLAP cube which supports three dimensional data. They did a real case study of Indian Postal Services is considered to

store large amount of data in the OLAP cube and can be easily used for the analysis purpose. UML class model, sequence diagram and star schema of object-oriented database are designed along with the OLAP cube.

In the study of *Manya Sethi* essential elements of decision support i.e data warehouse and OLAP are presented. It plays an important role in database industry. She quote that “data warehouse provides an effective way for analysis and statistic to the mass data and helps to do the decision making”. Many commercial products and services are now available and all of the principal database management system vendors now have offerings in these areas. She describes backend tools for managing the warehouse and also describes the concept hierarchies which define a sequence of mappings from a set of low level concept to high level concept. Discuss the OLAP server’s i.e ROLAP, MOLAP, HOLAP and metadata repository.

## Conclusion

Lot of studies on data warehousing and OLAP are found during the review of existing literature that Data cleaning is a problem that is a reminiscent of heterogeneous data integration, a problem that has been studied for many years. But here the emphasis is on data inconsistencies. Data cleaning is closely related to data mining, with the objective of suggesting possible inconsistencies. The management of data warehouses also presents a new challenge. Detecting the queries, Managing and scheduling resources are the problems that are important but that have not been well solved. thus there is a need of study of data warehouse and OLAP technology for banking sectors.

## References

1. S. Saagari, P.Devi Anusha, Ch.Lakshmi Priyanka, V.S.S.N. Sailaja. (2009). Designing efficient sorting algorithms for many core GPUs. In Proc. Int’l Parallel and Distributed Processing Symposium (IPDPS).
2. J. Gray and Dehne, T. Eavis. (2003). Distributed multi-dimensional ROLAP indexing for the data cube. Proc. IEEE/ACM Int. Symposium on Cluster Computing and the Grid.
3. Jarunee Wonglimpiyarat (2006). R-trees: a dynamic index structure for spatial searching. ACM SIGMOD, pages 47–57.
4. Jeffrey A. Clark; Thomas F. Siems (2007). A fault-tolerant environment for large-scale query processing. In High Performance Computing (HiPC), 19th International Conference on, pages 1–10.
5. Xintao Wu. (2004). Top-down computation of partial ROLAP data cubes. Proc. Hawaii Int. Conference on System Sciences.
6. Costas Lapavitsas And Paulo L. Dos Santos (2008). Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers.
7. Costas Lapavitsas and Paulo L. Dos Santos (2008). Efficient Concurrent Operations in Spatial Databases. PhD Thesis, Virginia Polytechnic.
8. Sivakumaran (2005). GPGPU.org. Cudpp library. <http://gpgpu.org/developer/cudpp>.
9. Suja .P.Mathews and Sunu George. (1997). Cubetree: Organization of and bulk incremental updates on the data cube. In Proceedings of the 1997 ACM SIGMOD Conference, pages 89–99.
10. Robert M. Bruckner and A. Rau-Chaplin. (2001). A cluster architecture for parallel data warehousing. Proc. Int. Conference on Cluster Computing and the Grid (CCGRID 2001).
11. Suraj Chaudari and Umeshwar Dayal. (2002). Hyper-Threading Technology Architecture and Microarchitecture. Intel Technology Journal, 6(1):4–15.
12. Thomas Vetterli and A. S. G. Duane G. (2010). Merrill. Revisiting sorting for gpgpu stream architectures. In Proceedings of the 19th international conference on Parallel architectures and compilation techniques.
13. Vadlamani Ravi.(2007). Private communication. nVIDIA Corporation.
14. Vaibhav R. Bhedi, Shrinivas P. Deshpande, Ujwal A. Lanjewar. (1996). On computing the data cube. Technical Report RJ10026, IBM Almaden Research Center, San Jose, California.
15. Vincent Rainardi, Dementiev, L. Kettner, and P. Sanders. (2008). Stxxl: standard template library for xxl data sets. Software: Practice and Experience, 38 (6).
16. Mr.P.P. Karde and Dr. V.M.Thakare. (2010). Dremel: interactive analysis of web-scale datasets. Proc. VLDB Endow., 3(1-2):330–339.
17. P.Kavitha and Dr.G.N.K.Suresh Babu. (2009) Optimizing data warehouse loading procedures for enabling useful-time data warehousing. IDEAS, pages 292–299.
18. Vipin Saxena and Ajay Pratap. (2008). Real-time data warehouse loading methodology. IDEAS, pages 49–58.
19. Manya Sethi.( 2011). In-Memory Data Management. Springer Verlag.
20. Marshon (2003). Real-Time Index Concurrency Control. IEEE Transactions on Knowledge and Data Engineering, 12(3):429–447.
21. Michael H. Brackett and N. Koziris.(2011). Brown dwarf: A fully-distributed, faulttolerant data warehousing system. Journal of Parallel and Distributed Computing, 71(11):1434 – 1446.



**Arpita Mathur** is BE(CS), M Tech (CS & E) and PhD (CS & E) from Jodhpur National University, Jodhpur. She has 13 years of experience in teaching various technical computer papers. She joined as a faculty member in Lachoo memorial college of Science and technology, Jodhpur in 2003 and currently is working as Assistant Professor in the Department of Computer Science in Lachoo Memorial Collegfe of Sc & Tech, Jodhpur. Her research interests include Database Management System, Data minning, Content Based Image Retrieval.