Improving Resource Management And Solving Scheduling Problem In Dataware House Using OLAP AND OLTP

Authors
Seenu Kohar¹, Surender Singh²
¹M.tech Computer Engineering OITM Hisar, GJU Univesity Hisar
²Assit. Professor CSE deptt. OITM Hisar, GJU University Hisar
Email-sheenukoh@gmail.com

ABSTRACT
A data warehouses provides us generalized and consolidated data in multidimensional view. Along with generalized and consolidated view of data, a data warehouses also provides us Online Analytical Processing (OLAP) tools. These tools help us in interactive and effective analysis of data in a multidimensional space. This analysis results in data generalization and data mining. Data mining functions such as association, clustering, classification, prediction can be integrated with OLAP operations to enhance the interactive mining of knowledge at multiple level of abstraction. That's why data warehouse has now become an important platform for data analysis and online analytical processing. Data warehousing is the process of constructing and using a data warehouse. A data warehouse is constructed by integrating data from multiple heterogeneous sources that support analytical reporting, structured and/or ad hoc queries, and decision making. Data warehousing involves data cleaning, data integration, and data consolidations.

1. INTRODUCTION
The term "Data Warehouse" was first coined by Bill Inmon in 1990. According to Inmon, a data warehouse is a subject oriented, integrated, time-variant, and non-volatile collection of data. This data helps analysts to take informed decisions in an organization. An operational database undergoes frequent changes on a daily basis on account of the transactions that take place. Suppose a business executive wants to analyze previous feedback on any data such as a product, a supplier, or any consumer data, then the executive will have no data available to analyze because the previous data has been updated due to transactions. Using Data Warehouse Information
There are decision support technologies that help utilize the data available in a data warehouse. These technologies help executives to use the warehouse quickly and effectively. They can gather data, analyze it, and take decisions based on the information present in the warehouse. The information gathered in a warehouse can be used in any of the following domains: Tuning Production Strategies - The product strategies can be well tuned by repositioning the products and managing the product portfolios by comparing the sales quarterly or yearly. Customer Analysis - Customer analysis is done by analyzing the customer's buying preferences, buying time, budget cycles, etc. Operations Analysis - Data warehousing also helps in customer relationship management, and making environmental corrections. The information also allows us to analyze business operation Integrating Heterogeneous Databases To integrate heterogeneous databases, we have two approaches: Query-driven Approach Update-driven Approach Query-Driven Approach. This is the traditional approach to integrate heterogeneous databases. This approach was used to build wrappers and integrators on top of multiple heterogeneous databases. These integrators are also known as mediators. Process of Query-Driven Approach When a query is issued to a client side, a metadata dictionary
translates the query into an appropriate form for individual heterogeneous sites involved. Now these queries are mapped and sent to the local query processor. The results from heterogeneous sites are integrated into a global answer set. Disadvantages Query-driven approach needs complex integration and filtering processes. This approach is very inefficient. It is very expensive for frequent queries. This approach is also very expensive for queries that require aggregations.

**Update-Driven Approach**

This is an alternative to the traditional approach. Today's data warehouse systems follow update-driven approach rather than the traditional approach discussed earlier. In update-driven approach, the information from multiple heterogeneous sources are integrated in advance and are stored in a warehouse. This information is available for direct querying and analysis. It includes tools for extracting data from multiple operational databases and external sources; for cleaning, transforming and integrating this data; for loading data into the data warehouse; and for periodically refreshing the warehouse to reflect updates at the sources and to purge data from the warehouse, perhaps onto slower archival storage. In addition to the main warehouse, there may be several departmental data marts. Data in the warehouse and data marts is stored and managed by one or more warehouse servers, which present multidimensional views of data to a variety of front end tools: query tools, report writers, analysis tools, and data mining tools. Finally, there is a repository for storing and managing metadata, and tools for monitoring and administering the warehouse system. The warehouse may be distributed for load balancing, scalability, and higher availability.

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. This paper provides an overview of data warehousing and OLAP technologies, with an emphasis on their new requirements. We describe back end tools for extracting, cleaning and loading data into a data warehouse; multidimensional data models typical of OLAP; front end client tools for querying and data analysis; server extensions for efficient query processing; and tools for metadata management and for managing the warehouse. In addition to surveying the state of the art, this paper also identifies some promising research issues, some of which are related to problems that the database research community has worked on for years, but others are only just beginning to be addressed. This overview is based on a tutorial that the authors presented at the VLDB Conference, 1996.

2. Manjunath T. N. wrote on Realistic Analysis of Data Warehousing and Data Mining Application in Education Domain

Data-driven decision support systems, such as data warehouses can serve the requirement of extraction of information from more than one subject area. Data warehouses standardize the data across the organization so as to have a single view of information. Data warehouses can provide the information required by the decision makers. Developing a data warehouse for educational institute is the less focused area since educational institutes are non-profit and service oriented organizations. In present day scenario where education has been privatized and cut throat competition is prevailing, institutes needs to be more organized and need to take better decisions.

2. LITERATURE REVIEW

1. Surajit Chaudhuri wrote on An Overview of Data Warehousing and OLAP Technology (Appears in ACM Sigmod Record, March 1997).
Institute’s enrollments are increasing as a result of increase in the number of branches and intake. Now a day, any reputed Institute’s enrollments count in to thousands. In view of these factors the challenges for the management are meeting the diverse needs of students and facing increased complexity in academic processes. The complexity of these challenges requires continual improvements in operational strategies based on accurate, timely and consistent information. The cost of building a data warehouse is expensive for any educational institution as it requires data warehouse tools for building data warehouse and extracting data using data mining tools from data warehouse. The present study provides an option to build data warehouse and extract useful information using data warehousing and data mining open source tools. In this paper they have explored the need of data warehouse / business intelligence for an educational institute, the operational data of an educational institution has been used for experimentation. The study may help decision makers of educational institutes across the globe for better decisions.

3. Mr. Dishek Mankad wrote on “The Study on Data Warehouse Design and Usage”

Data warehousing is a booming industry with many interesting research problem. The data warehouse is concentrated on only few aspects. Here we are discussing about the data warehouse design and usage. Let’s look at various approaches to the data ware house design and usage process and the steps involved. Data warehouse can be built using a top-down approach, bottom –down approach or a combination of both. In this research paper we are discussing about the data warehouse design process.

3. NEED AND SCOPE OF RESEARCH

A data warehouse is a database, which is kept separate from the organization's operational database. There is no frequent updating done in a data warehouse. It possesses consolidated historical data, which helps the organization to analyze its business. A data warehouse helps executives to organize, understand, and use their data to take strategic decisions. Data warehouse systems help in the integration of diversity of application systems. A data warehouse system helps in consolidated historical data analysis. Data Warehouse is separated from Operational Databases. A data warehouses is kept separate from operational databases due to the following reasons: An operational database is constructed for well-known tasks and workloads such as searching particular records, indexing, etc. In contract, data warehouse queries are often complex and they present a general form of data. Operational databases support concurrent processing of multiple transactions. Concurrency control and recovery mechanisms are required for operational databases to ensure robustness and consistency of the database. An operational database query allows to read and modify operations, while an OLAP query needs only read only access of stored data. An operational database maintains current data. On the other hand, a data warehouse maintains historical data.

4. PROPOSED WORK

- Investigation of new Challenges in data warehouse management.
- Study of runaway queries problems.
- Taking corrective measurement to manage Resource.
- Taking corrective steps in order to solve scheduling problem.

5. CHALLENGES

We have described the substantial technical challenges in developing and deploying decision support systems. While many commercial products and services exist, there are still several interesting avenues for research. We will only touch on a few of these here. Data cleaning is a
problem that is reminiscent of heterogeneous data integration, a problem that has been studied for many years. But here the emphasis is on data inconsistencies instead of schema inconsistencies. Data cleaning, as we indicated, is also closely related to data mining, with the objective of suggesting possible inconsistencies.

6. SCOPE OF RESEARCH
The problem of physical design of data warehouses should rekindle interest in the well-known problems of index selection, data partitioning and the selection of materialized views. However, while revisiting these problems, it is important to recognize the special role played by aggregation. Decision support systems already provide the field of query optimization with increasing challenges in the traditional questions of selectivity estimation and cost-based algorithms that can exploit transformations without exploding the search space (there are plenty of transformations, but few reliable cost estimation techniques and few smart cost-based algorithms/search strategies to exploit them). Partitioning the functionality of the query engine between the middleware (e.g., ROLAP layer) and the back end server is also an interesting problem. The management of data warehouses also presents new challenges. Detecting runaway queries, and managing and scheduling resources are problems that are important but have not been well solved. Some work has been done on the logical correctness of incrementally updating materialized views, but the performance, scalability, and recoverability properties of these techniques have not been investigated. In particular, failure and check pointing issues in load and refresh in the presence of many indices and materialized views needs further research. The adaptation and use of workflow technology might help, but this needs further investigation.

7. REFERENCE
1. Mr. Dishek Mankad “The Study on Data Warehouse Design and Usage” International Journal of Scientific and Research Publications , Volume 3, Issue 3, March 2013 ISSN 2250- 3153
2. Surajit Chaudhuri wrote on An Overview of Data Warehousing and OLAP Technology (Appears in ACM Sigmod Record, March 1997).
3. Manjunath T. N. wrote on Realistic Analysis of Data Warehousing and Data Mining Application in Education Domain
4. Kimball, R.The Data Warehouse Toolkit. John Wiley, 1996.
5. Barclay, T., R. Barnes, J. Gray, P. Sundaesan, “Loading Databases using Dataflow Parallelism.” SIGMOD Record, Vol.23, No. 4, Dec.1994.
6. Blakeley, J.A., N. Coburn, P. Larson. “Updating Derived Relations: Detecting Irrelevant and Autonomously ComputableUpdates.” ACM TODS, Vol.4, No. 3, 1989.
7. Gupta, A., I.S. Mumick, “Maintenance of Materialized Views: Problems, Techniques, and Applications.” Data Eng. Bulletin, Vol. 18, No. 2, June 1995. 9 Zhuge, Y., H. Garcia-Molina, J. Hammer, J. Widom, “View Maintenance in a Warehousing Environment, Proc. Of SIGMOD Conf., 1995.
8. Roussopoulos, N., et al., “The Maryland ADMS Project: Views R Us.” Data Eng. Bulletin, Vol. 18, No.2, June 1995.[11]
9. O’Neil P., Quass D. “Improved Query Performance withVariant Indices”, To appear in Proc. of SIGMOD Conf., 1997.
10. O’Neil P., Graefe G. “Multi-Table Joins through BitmappedJoin Indices” SIGMOD Record, Sep 1995.
11. Harinarayan V., Rajaraman A., Ullman J.D. “ Implementing Data Cubes
Efficiently” Proc. of SIGMOD Conf., 1996.
12. Chaudhuri S., Krishnamurthy R., Potamianos S., Shim K. “Optimizing Queries with Materialized Views” Intl.Conference on Data Engineering, 1995.
13. Levy A., Mendelzon A., Sagiv Y. “Answering Queries Using Views” Proc. of PODS, 1995. 16 Yang H.Z., Larson P.A. “Query Transformations for PSJ Queries”, Proc. of VLDB, 1987