Issues and Challenges in Database Research

By Manjeet Kumar
Niilm Cms, Greater Noida

Abstract - Database technology is having a major impact on the growing use of computers. Database Management is one of the most interesting areas of the research. The database research community is facing many issues and challenges in handling new and innovative approaches in the databases. Some of the database techniques on which currently research is going on are: Object Oriented Databases, Real Time Databases, Mobile Databases, Intelligent Databases, Multimedia Databases, Fuzzy Databases and some other approaches related to database like Data Mining, Data warehousing and Query Processing etc. In this paper we are going to present an overview related to research in the above discussed databases and their approaches.

GJCST-C Classification : E.0

Strictly as per the compliance and regulations of:

© 2012. Manjeet Kumar. This is a research/review paper, distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License http://creativecommons.org/licenses/by-nc/3.0/, permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Issues and Challenges in Database Research

Manjeet Kumar

Abstract - Database technology is having a major impact on the growing use of computers. Database Management is one of the most interesting areas of the research. The database research community is facing many issues and challenges in handling new and innovative approaches in the databases. Some of the database techniques on which currently research is going on are: Object Oriented Databases, Real Time Databases, Mobile Databases, Intelligent Databases, Multimedia Databases, Fuzzy Databases and some other approaches related to database like Data Mining, Data warehousing and Query Processing etc. In this paper we are going to present an overview related to research in the above discussed databases and their approaches.

I. INTRODUCTION

The data is very crucial organization resource; hence information systems have encountered a rapid progress and development [1]. Information systems, such as Database Management System (DBMS) and Information Storage and Retrieval System (ISRS) consist of collection, management, utilization and dissemination of information. Briefly, it is a means of processing information needed by an organization [2]. The superiority of information systems over the traditional filing cabinets in terms of time and space efficiency has greatly incremented the human's ability in data handling.

The full use of information systems raises new issues, especially for the access and the manipulation of information in a more intuitive, less formalized and human-friendlier way. It poses new challenges from data representation to querying and retrieval. We discuss what functions, traditionally allocated to DBMSs, still make sense in an environment, what new ones would be useful, specially investigating the querying step. Different types of databases are introducing in our real life to fulfill our all day to day requirements.

II. OBJECT ORIENTED DATABASE

Object Oriented Database Management Systems (OODBMS) have been used for handling complex data and complex relations existing among these data. These database systems are enriched with excellent modeling power. Computer Aided Design (CAD), Computer Aided Manufacturing (CAM) and Artificial Intelligence related concepts are the application areas of OODBMS. Now we are going to describe different research aspects of the object oriented systems. These aspects are: Concurrency Control, Data Clustering, Indexing in temporal databases, Trigger Scalability.

a) Concurrency control in OODBMS

Concurrency control is a process that is used to synchronize accesses to the databases to maintain its consistency. In OODBMS the concurrency control mechanism is more complicated than in a conventional database due to its complexity. Subsequently, if it is not carefully designed, the performance of the system may degrade severely. In this regard we are going to introduce a locking based scheme to increase concurrency among methods [3]. This model is based on multi-granularity locking. It uses a rich set of lock modes with different locking granularities and the concept of the commutativity among methods. In commutativity property, if two methods commute, the transaction invoking the methods can run in parallel and their execution orders do not affect their results. There are three very important features regarding the concurrency control in OODBMS with commutativity properties. First, it does not put the burden of determining commutativity for methods on application programmers. Second, it provides more concurrency by taking the fine locking granularity. Third, it reduces deadlocks due to lock escalation. Finally, it makes use of runtime information in method commutativity construction to improve the concurrency.

b) Data clustering in object oriented databases

Clustering is the technique that aims to decompose a given set of objects into subgroups or clusters based on similarity. The goal is to divide the data set in such a way that objects belonging to the same cluster are as similar as possible, where objects belonging to the different clusters are as dissimilar as possible [4, 5]. The motivation for finding and building classes in this way can be manifold.

When the data can not fit into the main memory, they are stored in the hard disks. Without data clustering, accessing two related objects usually requires two disks I/O because they are not stored in the same page. This degrades the performance of the OODBMS because accessing a hard-disk is slow. On the other side, the main memory can perform very fast random access. Thus, related objects should be stored close to each other in order to maximize the amount of relevant information returned when a page is loaded from the disk.
c) Indexing in temporal databases

A temporal database supports the storage and querying of information that varies over time. Each data value is associated with time interval corresponding to the transaction time, valid time or in case of bitemporal database. Sometimes data are not usually removed; instead updates are made by adding new records. Such a database is expected to be much larger than its conventional counter part, so its indexing in much more complicated. Conventional indexing techniques such as B+ trees and hash based indexes are not particularly useful for indexing interval data.

Developing object oriented indexing techniques that accommodates temporal characteristics so as to capture sophisticated semantics and provide a close model of future real world applications is major issue of research in this category. In this regard first issue is opposite to static databases, temporal databases are dynamic and it is practically impossible to recognize the storage and indexing each time a new date value becomes available. The second issue in this regard is that in dynamic temporal databases is to store the data so as to minimize the disk I/O for both temporal and non-temporal queries. All issues present in OODBMS are the issues of the temporal databases.

d) Trigger scalability in active OODBMS

A scalable trigger system is the active database that can support multiple triggers on a single data element. From an application point of view scalability allows for the implementation of work flow modeling in an active database systems. Scalable trigger support allows for workflow diagram to be compiled and implemented by the active database systems. These workflow diagrams would run substantially faster inside an active database. [6]

Some of the research issues in trigger scalability are:
1. Trigger priority and selection during execution such as whether they can be checked in parallel.
2. Determining the capacity of active (running) triggers in the systems.
3. Predictability of the trigger execution.
4. Strategies for synchronous as well as asynchronous trigger support.
5. Implementation of trigger caching strategies.

III. Mobile Databases

In the future, billions of the web clients will be accessing millions of databases and the World Wide Web will be one large federated system [6]. The quick development in wireless communication technology shows that these static database systems extend their services to mobile users. Security and transaction management are the main problems to be handled in these databases. Now we are going to discuss the research aspects of these two problems.

a) Transaction management in mobile databases

There are two major issues in this field. First, the techniques do not address the isolation property of the global transactions. It is difficult to enforce the isolation property in MMDB (Mobile Multi Data Base) [7] due to large number of dispersed databases in the federation.

Second, they fail to address disconnection that represents catastrophic failures.

b) Security in mobile databases

Database federation have achieved a high degree of sophistication when implementing solutions for data protection and sharing. The security issues arises, especially in the area of access control, when mobile databases are included as part of the normal functioning of a database federation [8]. So it is required to develop a security model for mobile database federation. The main problem in security model development is how the great variety of the databases involved in the system affects the system performance and security policies implemented in the federation. Development of multilevel security for mobile databases is also good factor for research.

IV. Multimedia Databases

In many applications such as medicine, law enforcement, video game development and web design, users may create new derived multimedia objects by editing existing ones. In order to save space, a derived object can be stored as a set of editing operations used to create it along with a reference to its base [8]. So the binary format of the derived objects does not have to be physically stored in the database. When a user wants to retrieve such an object, the system accesses referenced base object, then applies the associated editing operations on it. This storage format is called a specification.

Content Based Retrieval (CBR) is the recent research criteria in multimedia databases. The feature extraction and similarity search are the major issues in the CBR research.

V. Real Time Databases

The most important concept of the real time database system is that it should maintain the consistency constraints as well as timing constraints. In RTDBMS, processing of both temporal and persistence data is performed. Stock Trading, Air Traffic Control, telecommunications and Aircraft Flight Programs are the applications areas of the Real Time Database System. The main objective of these databases is to meet the timing constraints of data transactions regardless of a system or transaction failures. Recovery in Databases...
and processing nested transactions in RTDBMS are best research topics.

a) Recovery in Real Time Database

Recovery prepares a RTDBMS for coping with failures [9]. There are so many techniques available for recovery. There are logging, check pointing and reloading. The main objectives of these techniques are to reduce recovery time, minimizing the percentage of transactions missing their deadlines and minimizing the percentage of temporal data becoming invalid.

b) Processing nested transactions in RTDBMS

Extensive work has been performed in the area of real time transaction scheduling on single level transaction model, but limited work done at the nested models. For most of the advanced applications, transactions are long, complicated and access data items at various network sites. In a conventional real time single level transaction model, a transition is considered as a flat single unit of tasks that consists of a sequence of primitive actions with a given deadlines of individual transactions. Priorities in such systems are usually assigned according to the given deadlines of individual transactions. Schedulers may use these priorities to determine how to allocate resources. Due to the atomicity requirements of transactions, if there is failure during one’s execution, it is rolled back and restarted. Therefore, even if all transactions are initially schedulable, those rollbacks and restarts may cause them to miss their deadlines. Thus, transaction model beyond the flat model is desired to maximize system performance.

VI. Intelligent and Fuzzy Databases

Next generation information system applications require powerful and intelligent information management that necessitates an efficient interaction between databases and knowledgebase technologies [10, 11]. It is also important for these applications to incorporate uncertainty in data objects, integrity constraints and applications [12].

The relational database management systems having capability to handle the uncertain and imprecise data has been categorized under the Fuzzy Relational Database Management Systems (FRDBMS). In FRDBMS, we introduce fuzzy extensions to the relational database models. Within this framework of fuzzy data representation, similarity, conformance of tuples, concept of fuzzy functional dependencies and partial fuzzy functional dependencies are utilized to define the fuzzy key notions, transitive closures and fuzzy normal forms [13, 14, 15].

Fuzzy Object Oriented Data Model is also the extension of object oriented data model with fuzzy data handling permitting explicit representation of fuzzy data.

VII. Query Processing

MQO (Multiple Query Optimization) is the major issue in query processing research [16, 17]. Actually, MQO tries to reduce the execution cost of the group of queries by performing common tasks only once. In MQO, at the beginning of the optimization, all promising alternative plans have been generated and shared tasks are identified. This problem has been identified as an NP-Complete optimization problem where different heuristic functions are used to guide an A* search.

VIII. Data Mining and Warehousing

The big majority of the requests from a data warehouse involve dynamic ad-hoc queries [18]. The skill to answer these queries in a very quick manner is a critical issue in a data warehousing environment. Proper indexing is very important to avoid I/O intensive scans against the large data warehouse tables. The index has to be scalable to support the dynamic nature of the ad-hoc queries. The cost of building indexes using all the important attributes is prohibitive. So, it is challenging to find a subset of the indexes that would improve the ad-hoc queries’ performance automatically.

The second major issue with data mining and warehousing is to determine the workload for selecting the indexes, since it is unknown users start using data warehouse.

a) Multi Relational Data Mining and Query Flocks

Multi Relational Data Mining is the recent research area in data mining. One of the earliest work in multi relational data mining is query flocks techniques. It extends the concept of traditional association rule mining with a generate and test model for different kinds of patterns [19]. One possible extension of a query flock technique is the addition of view definitions including recursive views. Although the query flock techniques can be applied to a database schema including both the Intensional Database (IDB) or rules and Extensible Database (EDB) or table relations [20]. We can compile query flocks from data log into SQL in order to be able to use commercially available DBMS as our underlying engine of our system. We can also extend on multi relational data mining using inductive logic programming for discovering rules.

b) Sequential pattern matching

The main drawback of the sequential pattern matching is the large number of sequential patterns [21, 22] discovered which makes it harder for the decision maker to interpret them. Hence, repetition of sequences is the main problem added to the sequential pattern matching.

IX. Conclusion

The intended purpose of this paper is to provide a preliminary investigation of potential applications and
research aspects in databases. Different problems regarding various databases are briefly, but clearly discussed, for example Sequential Pattern Matching and Multiple Query Optimization in Data Mining. In future, the outcome of this paper would provide an innovative path for database researchers.

REFERENCES Références Referencias