

# A Review on Major Issues of Database Systems for Improving Throughput and Leading to Research Problems

Dr. Shaik Subhani

Associate Professor Department of IT, SNIST, Hyderabad, Telangana, India

## ABSTRACT

In this paper we survey, consolidate, and present the state of the art in database system and we try to specify the most important research possibilities in the data bases. The heart of our analysis is a survey on basic to advanced concepts and their relativity which will give a conclusion of how better performance is expected by integrating those aspects. In addition we described the advantages of parallelism and distribution including the data ware housing approach. This paper reviews all the above mentioned things in a better possible way. The main motto of the paper is how the basic aspects of the data base systems are helpful to lead the systems to advanced mechanism such as parallelism, distribution and handling of the metadata base architectures.

**Keywords:** Consistency, Distributed data bases, Heterogeneous data, Normalization, Parallelism, Repository, Serializability, transactions.

## I. INTRODUCTION

Databases are incredibly prevalent; they underlie technology used by most of people every day. Databases reside behind a huge fraction of websites; they're a crucial component of telecommunications systems, banking systems, video games, and just about any other software system or electronic device that maintains some amount of persistent information. In addition to persistence, database systems provide a number of other properties that make them exceptionally useful and convenient: reliability, efficiency, scalability, concurrency control, data abstractions, and high-level query languages. The reliability of every company depends on how the data is maintained by that company; with out perfect data the organizations are not in a position to publish their strategic decisions to the market. The upcoming issues which are mentioned below guarantees the perfect maintenance of the data by the companies which leads to reliability of the data. The ultimate goal of any data base system is to provide storage for the user and to provide a

convenient environment to handle the activities of the user. The activities may involve accessing the data, creating indexex, execution of the transactions, handling concurrency, parallelism and distribution on the data base systems.

This paper is investigated as follows. In next section, we discussed about the major issues of database systems related to databases and Section 3 concludes the paper.

## II. MAJOR ISSUES OF DATABASE SYSTEMS

### A. Storage Mechanisms

The data base system is an environment but not a product, the environment is one which is having autonomous nature in the data allocation required by the user and manipulation of memory for the efficient storage with out wastage of memory. The basic storage techniques provided by the data base system are sequential allocation and random allocation, the allocation depends on the requirement of the user. In general the user is not interest to

know where the data is going to be stored and what the memory locations etc. are, but having knowledge about the format of the storing definitely gives the better storage method for the user data. Some times it is specific for applications the data should be available in sequential manner such as record by record accessing of the user ,on the other hand some applications processing is not required record- by - record processing in this case random storage is preferable. How the storage mechanism is utilized efficiently by the user to maintain the data effectively. It depends the way of storage and allocation of memory for the user data.

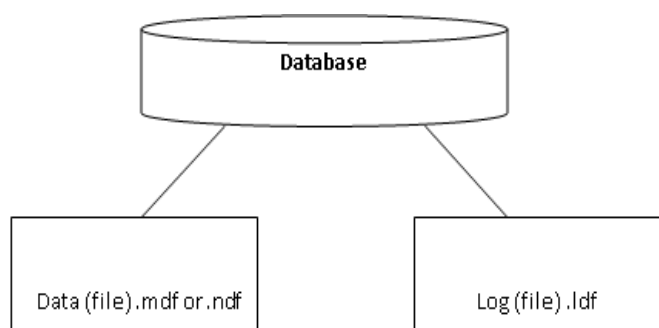


Figure 1. Storage and allocation of memory for the user data

### B. Access Techniques

The storage mechanism is abstract to the user, but the access technique should be familiar to the user, so as to process the data in efficient manner. For the faster processing of the data the most commonly used approach is usage of Indexing. Indexing is a way of accessing the data in the fastest manner by creating index file. The most commonly used indexing approaches are hash-based and tree-based indexing. If the data base user want to handle equality selection then hash-based indexing is preferable, some times the user may have the combination of search key and record-id or record-list, the tree-based indexing is better, however the range selection is possible with the usage of tree-based indexing. The tree-based indexing involves the Indexed Sequential Accessed Method proposed by IBM, to handle the

mainframe data+ -tree indexing which allows the user to handle the hierarchical searching of the data with balanced nature of the tree and controlling the wastage of memory. The main complexity in indexing is creation of the index file in dense, sparse, Clustered or Uncluttered manner. Creation of the index with less storage and selection of the index with data structure for faster accessing of the data with less complexity. The operations are insertion, deletion and searching of the data items.

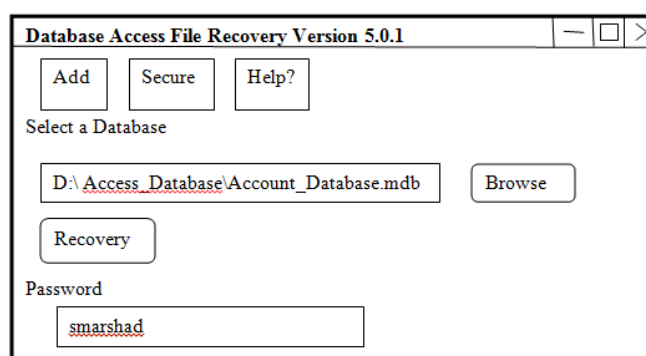


Figure 2. Access file recovery process

### C. Schema Handling

A schema is the basic element in the data base construction, because it refers to the overall design of the data base system. This schema involves both user data and as well as system data in the form of catalogues. The user schema involves the user related information in the form of data base tables or relations, constraints on relations, the amount of data required for data items etc., where as the system schema involves the system tables such as dictionaries, table space information etc.,.The success of data base system strongly depends on schema creation and its maintenance because the over all database information is represented by schema only. The logical data base design involves the abstract view of user schemas and their data items ,the most commonly used tool to handle the design of schema is ERWIN data modeler. The construction of the schema with less storage and faster accessing by avoiding unnecessary repetitions and unused data. The schemas may be accessed by a single user or else

multiple users so care should be taken for the schemas which are required by multiple users.

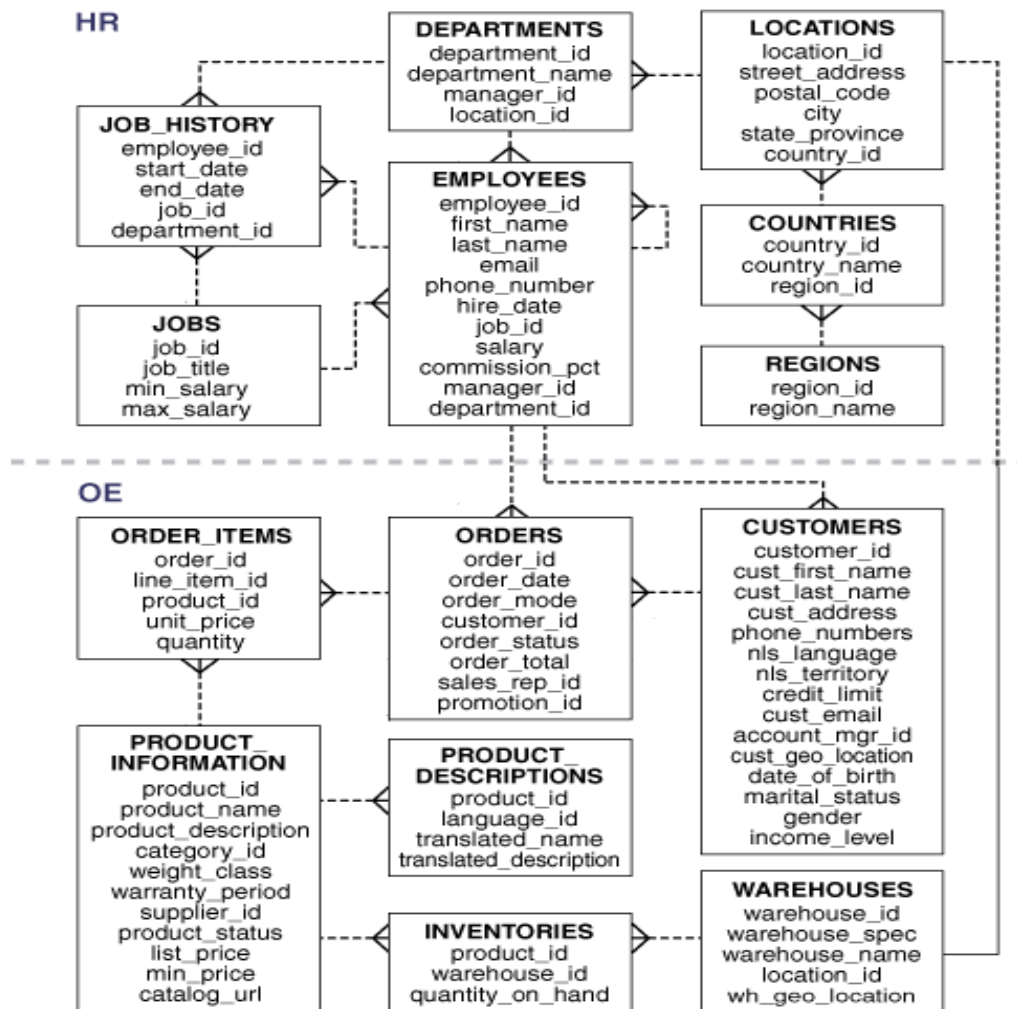


Figure 3. Schema design of the database system

#### D.Constraints/Assertions

The ultimate goal of the success full data base system is providing reliable data to the users. To provide quality data at the time of populating the data into the database the administrator should avoid the inconsistent data by putting some rules. The rule that is specified by the data base administrator on some table is known as a constraint, which guarantees some condition on the user insert/update data if the rule is satisfied by the constraint then the record is inserted/Updated otherwise it will be rejected by the constraint, so no question of ambiguity. Some times the constraint may be applicable on group of records known as Assertion. The best used constraints in relational data bases are Domain integrity constraints, Entity-Integrity constraints and referential integrity

constraints. In the form of triggers also we can impose constraints on some tables which are known as active data bases. Imposing constraint/Assertion on the data base by estimating the impact of the rule on the target users/customers which should improve the usage of the data base by the users/customers, care should be taken while handling Active data bases as they are not requiring explicit execution done by the user which may cause disturbance to other users/data.

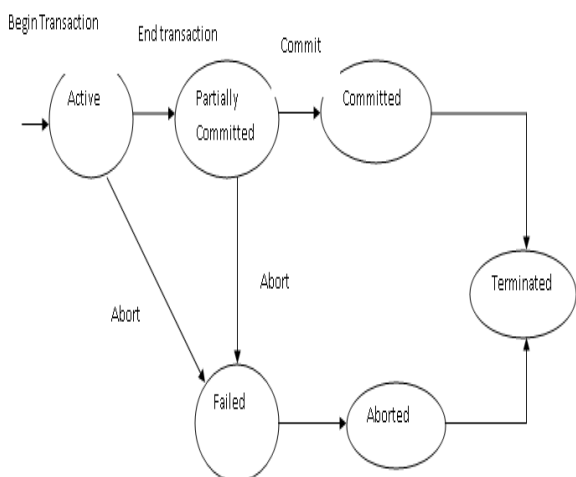
Table 1. Constraint Based Data Base System

Constraint	Column	Row	Table	External
Not Null	Y	N	N	N
Check	Y	Y	N	Y
Unique	N	N	Y	N

<b>Primary Key</b>	Y	N	Y	N
<b>Foreign Key</b>	Y	N	N	Y
<b>Index</b>	N	N	Y	N
<b>Trigger</b>	Y	Y	N	Y

**E. Transactions**

Every movement of data base system is the result of transactions. The transaction is responsible to carry the data base system from one consistent state to another consistent state. The insertion/deletion/updating the data in the data base is done by the transaction only.



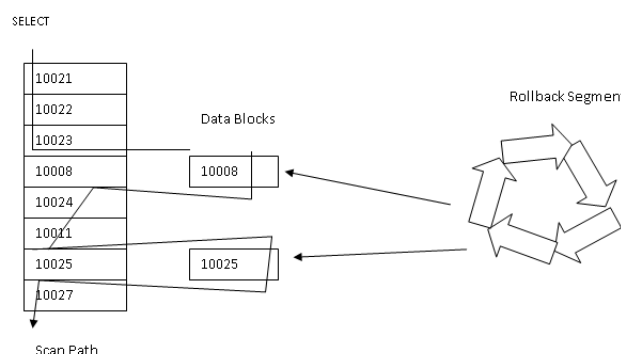
**Figure 4.** Transaction process of database system

The transactions are executed in the buffer by issuing the commit work the result of the transaction is send to the data base system ,if the user want to cancel the result of the transaction then issue rollback. The properties such as Atomicity guarantees the commit of entire work or cancel the entire work, consistency guarantees the correctness of the data before the transaction and after the transaction, Isolation is a property of disallowing other user to view/update the data while one user is processing the data, durability ensures that data remains in consistent state even after the failure. Transaction playing a vital role in data base management ,effective implementation of the transaction leads to better performance

sometimes buffer may lost its contents due to power failure or system failure which obviously cause transaction failure ,recovery of transaction is a major issue so as to handle the data base system in consistent state.

**F.Concurrency Mechanism**

Allowing multiple users and multiple transactions is required for two reasons. a)Improving Throughput b) Better Response time The above two aspects yields execution of the transactions in fastest manner and no more waiting time for small size transactions. By providing serializability in the form of conflict and view we can handle the schedules in concurrent manner. Provision of Concurrency leads to inconsistency of the data better performance is required but consistency is major in data bases for reliability. So achieving the concurrency with out compromising the consistency of the data in data base system is a major issue.



**Figure 5.** Concurrency Mechanism of database system

**G.Parallelism**

The parallel systems improve application processing and I/O speeds by using multiple cpus and disks in parallel. Adopting parallel architecture to data bases yields better performance when compared with isolated systems. The possibilities of achieving parallelism in databases are I/O parallelism, inter-query parallelism, intra-query parallelism, intra-operation parallelism,inter-operation parallelism The parallel systems provide various benefits by utilizing the resources in optimized manner, some are speedup and scale up, interconnection networks. Selection of

the architecture to achieve parallelism in data bases is playing a vital role; the main issue here is choosing better architecture according to the application requirement. The possible parallel architectures are shared memory, shared disk, shared nothing, hierarchial. Handling the above possible architectures by integrating with data bases is a critical thing.

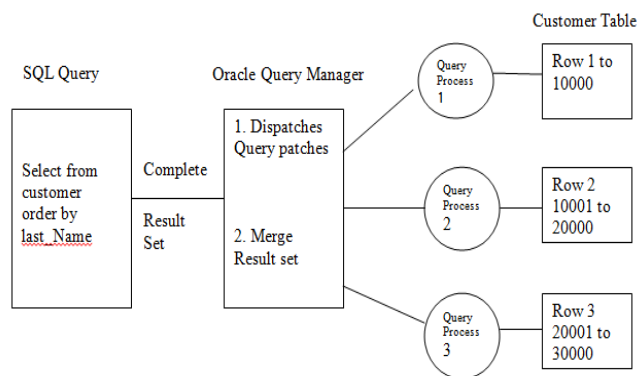


Figure 6. Parallel Process system in data bases

### H. Distribution

The geographically dispersed systems are interconnected by distributed mechanism and LAN acts as a back-bone for these systems. The distributed systems provide single system image interface to the user by providing data transparency, replication transparency, and fragmentation transparency. The important thing in the maintenance of the transaction is handling local and global transactions. Each and every system in distributed environment should participate in at least one global transaction including local transactions. The global transaction processing requires data from local system and as well as remote systems, with the help of middle-ware technologies and remote procedure calls one can achieve the result of the global transactions. The main benefits of distributed data bases include sharing of the data, autonomy, and availability of the data. The distributed data bases provides so many advantages to the data base users by improving sharing and availability of the data in spite of these advantages there are potential issues relevant to maintenance of various repositories and users.

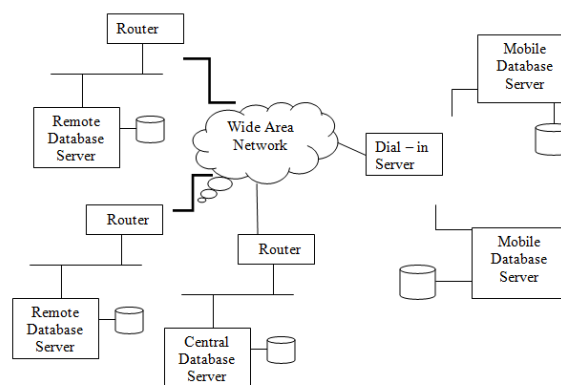


Figure 7. Distribution system representation in data bases

### I. Security

One of major area in the handling of data base systems is provision of security for the data in data base. The data base administrator with the help of authorization and authentication the security measures are improved. The general data bases with single repository it is enough to handle the security mechanism for that repository as all the users are accessing the data from that repository only. The security measure is tedious if the data base system depends on multiple repositories. In case of parallel data bases and distributed data bases the implementation security is not that much easy and it is a critical aspect there. Provision of protection to the individual user data is also important in multi-user data base systems. Providing security for multiple servers and multiple client environments, handling security measures for parallel data base and provision of security for distributed data base environments is a major issue in data bases.

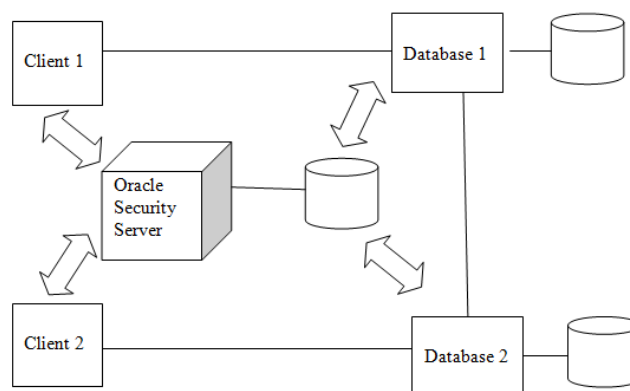


Figure 8. Provision of security in database system

### J.Data quantity

Systems research is an iterative and data intensive. Current data will give rise to new information and models and in turn will result in more new data with variations. This cycle continues, and data volume increases exponentially. Therefore management of data quantity is crucial for systems driven research. Data quality describes a set of data properties describing their ability to satisfy user's expectations or requirements concerning data usage for acquiring information in a given area of interest, learning and decision making. Databases should institute quality check measures to ensure that the data they provide to the research community is of high quality. It is much easier to enforce quality measures in a closed setup, but it is a major problem to be addressed in a social collaborative environment. Poor quality data may contain incomplete or missing fields. The data may be represented in a non-standard/legacy formats that will create problems for data and information integration. To ensure high quality, after data is received, databases should then use their own quality measures which may also include manual interest by domain experts. There should be standardized mechanism to ensure consistency and completeness of the submission. Usage of Semantic aware forms (ontology guided forms discussed later) for data procurement, on-the-spot data entry field validations using advance Web scripts may minimize proliferation of inaccurate data.

### K.Other Advanced Areas

The advanced areas of data bases are maintenance of online Analytical processing and online transaction processing. The data base systems are basically follows the relational model in the handling of the data; there may be the possibility of handling the data in the form of object-oriented manner, object-relational manner according to the application demand. The other important consideration is getting the data from various heterogeneous sources such as Flat files, relational model,XML source etc.,the various source information may be put into a

common repository which may termed as Data ware housing. Here handling of Extraction, Transformation and Loading is one task of populating data into data ware house repository after that getting required reports with the help of Analytical processing, which is techno functional module The other advanced areas include metadata base maintenance, handling of efficient security mechanisms, execution speedup of the transactions, provision of common query language to all different types of the data and different kinds of the users. Maintenance of heterogeneous sources, handling of multi data bases, improving security measures for various kinds of data bases, handling of object oriented and object relational data bases are major research issues in the data bases.

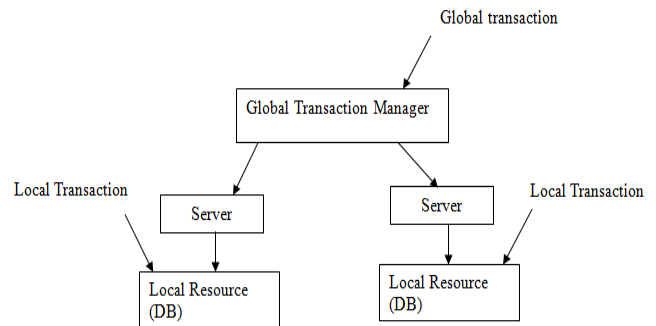


Figure 9. Transaction issue related databases

### III. CONCLUSION

The presentation of basic aspects of the data bases and leading to advanced aspects involves various technical things, because the base aspects such as integrity constraints and normalizations are automatically covered by the advanced areas such as parallel data bases, distributed data bases and data ware housing. One more thing is the basic security mechanism used in single server is improved up to multiple servers and multiple clients. The inference specifies the internal concept should be same such as consistency,t ransactions, normal forms, security, concurrency for all the advanced areas so at the atomic level of data base construction the designer should take care the about the basic aspects so as to convert the basic data bases to advanced ones.

#### **IV. REFERENCES**

- [1]. Recent Progress on Selected Topics in Database Research Report from Nine Young Chinese Researchers Working in the United States.
- [2]. Database System Concepts by Korth, third edition.
- [3]. Performance of Short-Commit in Extreme Database Environment, International Journal of Database Management Systems (IJDMS), Vol.3, No.2, May 2011.
- [4]. A survey of user-centric data warehouses: from personalization to recommendation, International Journal of Database Management Systems (IJDMS), Vol.3, No.2, May 2011.
- [5]. Performance analysis of distributed data base systems Michael Stonebraker, John Woodfill, Jeff Ranstrom.
- [6]. A Study on Challenges and Opportunities in Master Data Management., International Journal of Database Management Systems (IJDMS), Vol.3, No.2, May 2011.
- [7]. Comprehensive online Atomic Database Management System (DBMS) with Highly Qualified Computing Capabilities, International Journal of Database Management Systems ( IJDMS ), Vol.3, No.2, May 2011.
- [8]. A comparative study between the performance of relational & Object oriented database in data warehousing, International Journal of Database Management Systems (IJDMS), Vol.3, No.2, May 2011.
- [9]. Data Mining - An Industrial Research PerspectiveC. ApteIEEE Computational Science and Engineering April-June 1997.
- [10]. A Comprehensive Study on Data Mining Process with Distribution K Uma Pavan Kumar, Snist, hyd, national conference, 2010.
- [11]. Parallelism in the processing of Clusters in mining with distribution, K.Uma Pavan kumar, Interntional conference in Coimbatore, 2011.