Concurrency Control in Distributed Database System

National Institute of Technology , Raipur

By

Ashish Verma(Update_4)

19111014

TERM PAPER SUBMISSION

feb 2022

Contents

1   ABSTRACT:

This paper reviews the coverage of concurrency control in Distributed networks. A distributed network becomes more popular, the need for improvement in distributed database management systems becomes even more important. The main challenges are identified as: (1)Preserving the ACID property atomicity, consistency, isolation, and durability property when concurrent transactions perform read and write operation; (2) providing recovery method when distributed data fail; (3)whatever method that is chosen they must provide feasible solutions with respect to performance. Keywords: Distributed Database, Distributed Design, Fragmentation, Replication, Allocation, Concurrency control, Transaction

1.1   INTRODUCTION:

requirement for secure, dependable, and open data in the present business climate, the requirement for conveyed information bases and customer/server applications is additionally expanding. A dispersed information base is a solitary coherent data set that is spread truly across PCs in different areas that are associated by information correspondence joins. A conveyed data set is a sort of virtual data set whose part parts are truly put away in various particular genuine data sets at various unmistakable areas. The clients at any area can get to information anyplace in the organization as though the information were completely put away at the client's own area. An appropriated information base administration framework is the product that deals with the Distributed Databases and gives an entrance component that makes this dispersion straightforward to the client. The goal of a circulated data set administration framework (DDBMS) is to control the administration of a disseminated data set (DDB) so that it appears to the client as a unified information base. This picture of a unified climate can be cultivated with the help of different sorts of straightforwardness like Location Transparency, Performance Transparency, Copy Transparency, Transaction Transparency, Transaction Transparency, Fragment Transparency, Schema Change Transparency,

and Local DBMS Transparency. Simultaneousness control is additionally a significant issue in data set frameworks. Simultaneousness control is the method involved with organizing simultaneous admittance to a data set in a multi-client information base administration framework (DBMS). There exist various strategies that give simultaneousness control. A portion of the strategies are Two-stage locking, Timestamping, Multi-adaptation timestamp, and so on.

2   MOTIVATION:

There are various business conditions that encourage the use of distributed databases: Data communications costs and reliability: If the data is geographically distributed and the applications are related to these data, it may be much more economical, in terms of communication costs, to partition the application and do the processing at each site. On the other hand, the cost of having smaller computing powers at each site is much less than the cost of having an equivalent power of a single mainframe.

Database recovery: The process of restoring data that has been lost, accidentally deleted, corrupted, or made inaccessible for any reason. In enterprise information technology (IT), data recovery typically refers to the restoration of data to a desktop, laptop, server, or external storage system from a backup.

Data sharing: the practice of making data used for scholarly research available to other investigators. Replication has a long history in science. The motto of The Royal Society is 'Nullius in verbal, translated "Take no man's word for it.

Distribution and autonomy of business units: Divisions, departments, and facilities in modern organizations are often geographically distributed, often across national boundaries. Often each unit has the authority to create

its own information systems, and often these units want local data over which they can have control. Business mergers and acquisitions often create this environment.

Improved Performance: Performance improvement, by its nature, is iterative. For this reason, removing the first bottleneck might not lead to performance improvement immediately, because another bottleneck might be revealed. Also, in some cases, if serialization points move to a more inefficient sharing mechanism,then performance could degrade. With experience, and by following a rigorous method of bottleneck elimination, applications can be debugged and made scalable.

Increased reliability and availability: When a centralized system fails, the database is unavailable to all users. In contrast to centralized systems, the distributed systems will continue to function at some reduced level, however, even when a component fails. Faster response: Well-suited for a loosely defined data structure that may evolve over time. In-memory database management system (IMDBMS) - provides faster response times and better performance.

3   DISTRIBUTED DATABASE DESIGN:

Distributed databases can be homogenous or heterogeneous. In a homogeneous distributed database system, all the physical locations have the same underlying hardware and run the same operating systems and database applications. In a heterogeneous distributed database, the hardware, operating systems, or database applications may be different at each of the locations

Data Fragmentation: In Distributed Databases, we need to define the logical unit of Database Distribution and allocation. The database may be broken up into logical units called fragments which will be stored at different sites. The simplest logical units are the tables themselves.

3.1 Three Types of Data Fragmentation are:

- Horizontal fragmentation:refers to the division of a relation into subsets (fragments) of tuples (rows).

  Example: Say we have this relation

customer_id	Name	Area	Payment Type	sex
1	Kashish	Chhattisgarh	debit card	female
2	Ragini	Goa	Credit card	female

Horizontal Fragmentation are subsets of tuples (rows)

Fragment 1

customer_id	Name	Area	Payment Type	sex
1	Kashish	Chhattisgarh	debit card	female
2	Ragini	Goa	Credit card	female

Fragment 2

customer_id	Name	Area	Payment Type	sex
3	Manan	Punjab	debit card	male

- Vertical fragmentation:refers to the division of a relation into subsets (fragments) of tuples (rows).Vertical Fragmentation of a relation R produces fragments R1, .., Rn, each of.

  Example: Say we have this relation

Vertical Fragmentation are subsets of tuples (rows)

Fragment 1

customer_id	Name	Area	sex
1	Kashish	Chhattisgarh	female
2	Ragini	Goa	female
3	Manan	Punjab	male

Fragment 2

customer_id	Payment type
1	debit card
2	credit card
3	debit card

- Hybrid fragmentation: Hybrid Fragmentation comprises the combination of characteristics of both Horizontal and Vertical Fragmentation. Each fragment can be specified by a SELECT-PROJECT combination of operations. In this case, the original table can be reconstructed by applying union and natural join operations in the appropriate order

Data Replication: the frequent electronic copying of data from a database in one computer or server to a database in another so that all users share the same level of information. The result is a distributed database in which users can access data relevant to their tasks without interfering with the work of others.

4 FUNDAMENTALS OF TRANSACTION MANAGEMENT AND CONCURRENCY CONTROL:

4.1 Transaction:

is a sequence of operations performed as a single logical unit of work. A logical unit of work must exhibit four properties, called the atomicity, consistency, isolation, and durability (ACID) properties, to qualify as a transaction.

4.2 Properties of Transaction:

A Transaction has four properties that lead to the consistency and reliability of a distributed database. These are Atomicity, Consistency, Isolation, and Durability.

Atomicity: An atomic transaction is an indivisible and irreducible series of database operations such that either all occur, or nothing occurs.

Consistency: database systems refer to the requirement that any given database transaction must change affected data only in allowed ways. Any data written to the database must be valid according to all defined rules, including constraints, cascades, triggers, and any combination thereof.

Isolation: in the context of databases, specifies when and how the changes implemented in an operation become visible to other parallel operations. Transaction isolation is an important part of any transactional system. It deals with consistency and completeness of data retrieved by queries unaffecting a user data by other user actions. A database acquires locks on data to maintain a high level of isolation.

Durability: databases are the property that ensures transactions are saved permanently and do not accidentally disappear or get erased, even during a database crash. This is usually achieved by saving all transactions to a non-volatile storage medium.

4.3   Concurrency Control:

A database management system (DBMS) concept that is used to address conflict with the simultaneous accessing or altering of data that can occur with a multiuser system.