



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5

Issue: V

Month of publication: May 2017

DOI:

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Comparative Analysis: Parallel and Distributed Database

Er. Ramandeep Kaur Brar¹, Er. Chinu², Er. Gurpreet Kaur³

^{1,3}Research Scholar, ² AP CSE, Baba Farid College of Engineering and Technology Bathinda, Punjab

Abstract: As we know that parallel and distributed databases are used to load the huge amount of data simultaneously. Distributed database is a software that provides an access mechanism that makes the distribution transparent to user whereas the Parallel database system seeks to improve the performance through parallelization of various operations such as loading data, building indexes and evaluating queries by using multiple CPU's and disks in parallel. This paper thrash out architecture, various applications, recompense, demerits which has potential to progress the performance and the comparative study of the distributed and parallel databases.

Keywords: Introduction, Advantages & Disadvantages, Architecture and comparision between distributed and parallel databases.

I. INTRODUCTION

All the databases are used to store the information and organization database in efficiently manner. There are various databases like object oriented which is used for the concept of oops, relational database tables which is used the concept of tables in the form of tuples and attribute, distributed database which is used for the data synchronously, and parallel database which is used to speed up data. In this paper we will discussed about the distributed and parallel database. It is seen that, the organizations have been attracted in the decentralization of processing while achieving the integration of the information assets within their geographically distributed systems of database, applications and users [1]. A distributed database (DDB) is a mixture of logically interrelated databases, but physically distributed larger than several computers (a network of computers)[3]. several techniques have been planned to rank distributed database systems introducing many capable algorithms. A distributed database system is a database system which is fragmented or replicated on the different configurations of hardware and software, located usually at dissimilar geographical sites within an organization [8]. A distributed database is a database in which storage devices may be stored in various computers, located in the identical physical location; or may be isolated over a network of interconnected computers [10]. A parallel database system seeks to recover presentation during parallelization of dissimilar operations, such as loading data, building indexes and calculate queries [11]. In this paper we will talk about both parallel and distributed databases in detail. Parallel databases progress processing and input/output speeds by using several CPUs and disks in parallel. Centralized and client-server database systems are not influential enough to handle such applications. In parallel processing, several operations are performed concurrently, as opposed to serial processing, in which the computational steps are performed in order[12]. A typical database resides on a dedicated computer connected to others on a data network. The server program, running on the computer, handles user requests for data. As many people try to access the database, yet, the server becomes overwhelmed. Upgrading to a quicker computer helps, but simply to a peak [13].

A. Advantages of Distributed Database[7]

- 1) **Availability:** The data is replicated at several sites. In case the local server is busy due to a few reason the data can be retrieved from the other available server.
- 2) **Robustness:** The whole system becomes more robust as multiple servers are anxious in handling data. Thus, breakdown of one system doesn't lead to work to rule of complete system.
- 3) **Sharing:** Data at multiple sites is shared throughout users at disparate sites. .
- 4) **Performance:** It leads to improved performance as various equipments are involved, the load is distributed. The database is divided into database fragments, thus local queries can be determined by local databases rather than all queries being targeted to one centralized database.
- 5) **Ease of Growth:** To attach more clients to such a system is fairly simple as Overloading is never a matter.
- 6) **Management of distributed data with different levels of transparency, Hardware, Operating System, Network and Location**

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Independence. It provides Continuous operation. No more reliance on the central site.

B. Disadvantages of Distributed Database

- 1) *Complexity*: DBAs may have to do extra work to ensure that the distributed nature of the system is transparent. Extra work be required to be done to maintain multiple unrelated systems, instead of big one.
- 2) *Economics*: Increased complexity and a more extensive infrastructure means extra labour costs.
- 3) *Security*: Remote database fragments must be secured, and they are not centralized so the remote sites must be secured as well.
- 4) *Difficult to Maintain Integrity*: In a distributed database, enforcing integrity over a network may require too much of the network's resources to be feasible .
- 5) *Inexperience*: Distributed databases are difficult to work in such a young field there is not much readily available experience in "proper" practice .
- 6) *Lack of Standards*: There are no tools or methodologies yet to help users convert a centralized DBMS into a distributed DBMS.
- 7) *Database Design more Complicated*: The structure of a distributed database has to recognize fragmentation of data to particular sites and data replication .
- 8) Additional software is required
- 9) Operating

C. Distributed Database Architecture

A distributed database system allows applications to use the data from local and remote databases. Homogenous distributed database system contains each of the database is an Oracle Database. Heterogeneous distributed database system contains at least one of the databases that is not an Oracle Database. Distributed databases use a client/server architecture to process information requests. This section includes the following topic which is given below:

- 1) Homogenous Distributed Database Systems
- 2) Heterogeneous Distributed Database Systems
- 3) Client/Server Database Architecture

D. Homogenous Distributed Database Systems

A homogenous distributed database system is a network of two or more Oracle Databases that reside on one or more machines. An application can concurrently access or vary the data in various databases in a single distributed environment. For a client application, the location and platform of the databases are transparent. In this way, a distributed system gives the appearance of native data access. [2]

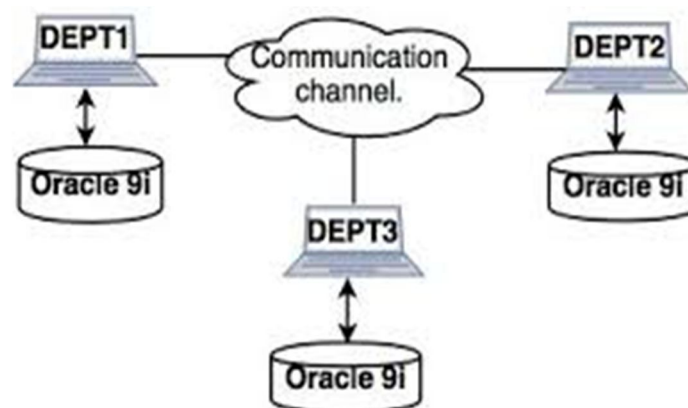
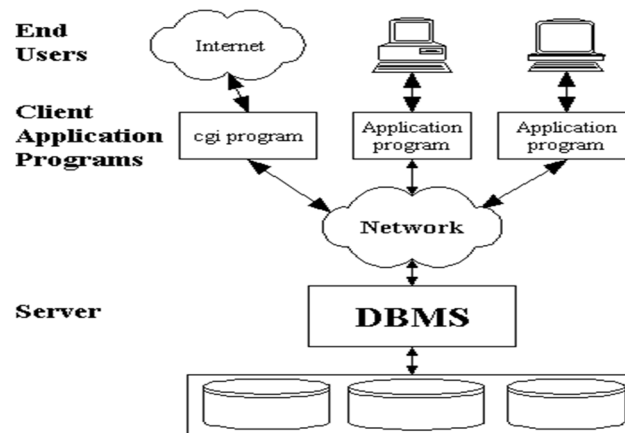


Figure-1 Homogeneous Distributed Database[22]

- 1) *Heterogeneous Distributed Database System*: In this different site run under the control of different DBMS.[16]
- 2) *Client/Server Database Architecture*: A client is an application that requests to getting information from a server.Each node in a distributed database system can shows as a client, a server, or both. A client server database displayed in the following diagram:-

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Client/Server Architecture



II. APPLICATIONS

Distributed database supports many distributed applications like businesses and services, reliable communication, i.e., communication that can allowably be trusted, is of critical importance. Generally two different communication models are used to realize communication between distributed application peers: point- to-point which is also known as unicast, and multicast which includes broadcast communication[4]. For example, E-learning systems are also used as the application of the distributed database. The information and communications technology and growth of multiple connections, anywhere, allowed a latest approach about education, by more geographic well-organized distribution of information, knowledge and practice. Hence a new concept appeared, called e-learning. E-learning courses includes the content i.e., information and instructional methods that help people learn the content[3].

A. Techniques of Distributed Databases

- 1) *Fragmentation*: Fragmentation is a propose technique to partition a relation into two or more partitions in such that combination of partitions provides the original database without any loss of creative data. In Distributed database, record is broken into logical units called fragments which will be stored at different sites [16].

B. Advantages of Fragmentation

- 1) *Parallelism*: Due to fragmentation a transaction can be divided into several sub queries that operate on fragments. This has to increase the degree of parallelism in the system.
- 2) *Security*: Data, which is not required by restricted applications that is not stored and not accessible for unauthorized users.
- 3) *Efficiency*: Data are stored on nearest to those databases which are repeatedly used.

C. Disadvantages of Fragmentation

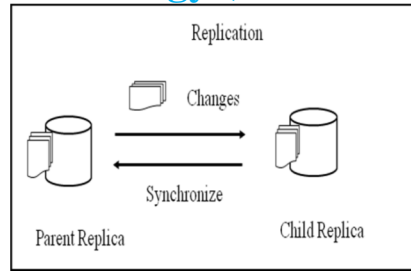
For the purpose of data stored at different sites, the overall performance becomes slow and Integrity control becomes too complex.

Types of Fragmentation

Three types of fragments are discussed below:

- 1) *Horizontal Fragmentation*: It divides the relation into the rows. It divides table horizontally by selecting the related rows and these fragments can be assigned to different sides in the distributed system. It also allows a class to be partitioned into dislodge instances. In horizontal fragmentation table remains same, only rows gets split[16].
- 2) *Hybrid Fragmentation*: Hybrid fragmentation is the mixture of both horizontal and vertical fragmentations so that hybrid fragmentation is also called mixed fragmentation [16].
- 3) *Vertical Fragmentation*: Vertical fragmentation divides the relation into attributes called columns. For the purpose to construct the original table, it is possible with the help of primary key. The main goal of vertical fragmentation is to partition relation into a set of smaller relations so that many of the applications will run on only one fragment[16].
- 4) *Data Replication*: In data replication there are multiple copies of the data which is held in different locations and different processes work with different copies[16].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)



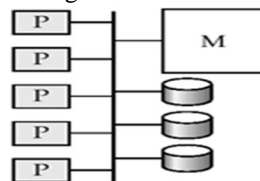
That's why there is compulsory for improving the availability of data. There can be complete replication which contains a copy of the entire database that is stored at every site. There are a number of advantages and disadvantages to replication [17].

D. Advantages of Data replication

- 1) *Expanded parallelism*: For the purpose of expanded parallelism this leads to faster query execution.
- 2) *Less Data Movement over Network*: Data replication reduces movement of data among sites and increases speed of processing.
- 3) *Availability*: If one of the sites containing relation R fails, then the relation R can be obtained from another site.

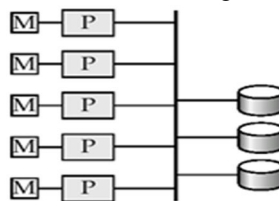
E. Disadvantages of Data Replication

- 1) *Require more Disk Space*: Storing replicas of same data at different sites consumes more disk space.
- 2) *Expanded Transparency on Update*: Whenever an updation is required, a database system must ensure that all replicas are updated.
- 3) *Expensive*: Concurrency control and recovery techniques will be more advanced and hence more expensive.
- 4) *Architecture of Parallel Database*: Parallel database can be used various architectures that are given below [14].
- 5) *Shared Memory Architecture*: In this architecture, processors shared a common memory and multiple CPUs are attached to an interconnection network and can access a common region of main memory. (See the diagram below)[9]



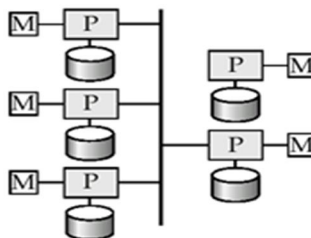
[21]

- 6) *Shared Disk Architecture*: In this architecture, processors share a common disk. Each CPU has a private memory and direct access to all disks through an interconnection network. (See the diagram below)[9]



[21]

- 7) *Shared Nothing Architecture*: Where each node has its own mass storage as well as main memory. (See the diagram below)[4]



[21]

F. Types of Parallelism: [14]

- 1) *Interquery Parallelism*: Execution of multiple queries in parallel.
- 2) *Interoperation Parallelism*: Execution of single queries that may consist of more than one operations to be performed. There are two forms of interoperation parallelism:-

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

- 3) *Independent Parallelism*: Execution of each operation individually in different processors only if they can be executed independent of each other.
- 4) *Pipe-Lined Parallelism*: Execution of different operations in pipe-lined fashion. For instance, if we need to join three tables, one processor may join two tables and send the result set records as and when they are produced to the other processor. In the other processor the third table can be joined with the incoming records and the final result can be produced.
- 5) *Intraoperation parallelism*: Execution of single complex operations in parallel in multiple processors. For example, ORDER BY clause of a query that tries to execute on millions of records can be parallelized on multiple processors.

F. Advantages of Parallel Databases [13]:

- 1) **Capacity**: A parallel database allows a large online trader to have thousands of users accessing information at the same time.
- 2) **Speed**: The server breaks up a user database request into parts and post each part to a separate computer. They work on the parts concurrently and combine the results, passing them back to the user. This speeds up, allowing faster access to very complex databases.
- 3) **Reliability**: A parallel database, properly configured, can continue to work in spite of the failure of any computer in the cluster.

G. Disadvantages of Parallel Database [15]

- 1) Programming to target Parallel architecture is a bit difficult but with proper understanding and practice you are good to go. Various code alteration has to be performed for different target architectures for improved performance. Communication of results might be a problem in certain cases.
- 2) Power utilization is huge by the multi core architectures.
- 3) Also, better cooling technologies are required in case of clusters.

Sr.no.	Parallel DBMS	Distributed DBMS
1.	It is a software system where multiple processors or machines are used to execute and run queries in parallel.	It is a software system that manages multiple logically interrelated databases distributed over a computer network.
2.	Shared Memory, Shared Disk, Shared nothing Architecture Hierarchical, Architecture	Homogeneous, Heterogeneous, Federated DB system, Multi database System
3.	Nodes types are compulsorily Homogeneous	In this there is no need to be homogeneous
4.	Difficult to expand	Easier to expand
5.	Operating cost is less	Operating cost is more
6.	Speed of execution is Quicker	Here speed of execution is slower
7.	The nodes are located at geographically same location.	The nodes are usually located at geographically different locations.
8.	Performance of parallel database is lower reliability & availability.	Performance of distributed database is higher reliability & availability.
9.	Backup at one site only	Backup at multiple sites
10.	Global clock control, SIMD,MIMD	No Global clock control, synchronization algorithms needed[19]
11.	Interconnections are bus ,mesh, tree, mesh of tree and hypercube network	Interconnections are Ethernet(bus), token ring and SCI(ring), switching network [19]
12.	Advantages are capacity, speed, reliability	Advantages are availability, sharing, performance, ease of growth
13.	Disadvantages are cost is increased considerably, Huge Number of resources are required	Disadvantages are complexity, security, economics, inexperience

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

III. CONCLUSION

This paper provides an overview of technology important to realize a parallel scalable and high performance database and management system. Parallel database provides quicker access to data and reduce the complexity of storing the data. On other hand, distributed database manages multiple logical inter-related database distributed over a communication network.

REFERENCES

- [1] Anjali Saini ,” Distributed System and its Role in HealthCare System”, IJCSMC, Vol. 4, Issue. 4, April 2015, pg.302 – 308
- [2] https://docs.oracle.com/cd/B28359_01/server.111/b28310/ds_concepts001.htm#ADMIN12075
- [3] a Phd student,” The use of distributed databases in e-learning systems”
- [4] Naghmeh Ivaki, Nuno Laranjeiro *, Filipe Araujo,” A survey on reliable distributed communication”, [m5G; April 10, 2017;14:40
- [5] Yousry M. AbdulAzeem?, Ali I. ElDesouky, Hesham A. Ali,” A framework for ranking uncertain distributed database”. Received 4 September 2012
Received in revised form 25 November 2013 Accepted 24 May 2014 Available online 13 June 2014
- [6] M. Wu, J. Xu, X. Tang, W.-C. Lee, Top-k monitoring in wireless sensor networks, IEEE Trans. Knowl. Data Eng. 19 (2007) 962–976
- [7] D.S.Hiremath1, Dr.S.B.Kishor2,”Distributed Database Problem areas and Approaches”, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727 PP 15-18
- [8] Iacob (Ciobanu) Nicoleta ?? Magdalena a*,” Distributed Queries in the E-learning Environment” Procedia - Social and Behavioral Sciences 28 (2011) 241 – 24
- [9] Hiren H Darji1, BinalS Shah2, Manisha K Jaiswal3,”CONCEPTS OF DISTRIBUTED AND PARALLEL DATABASE”,IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555,Vol. 2, No.6, December 201
- [10] Parul Tomar1 and Megha2,”An Overview of Distributed Databases ”,International Journal of Information and Computation Technology.ISSN 0974-2239 Volume 4, Number 2 (2014), pp. 207-214 © International Research Publications House <http://www.irphouse.com/ijict.ht>
- [11] Dewitt, David; Gray, Jim (1992). "Parallel database systems: The future of high performance database systems" (PDF). Communications of the ACM. 35 (6): 85. doi:10.1145/129888.129894
- [12] https://en.wikipedia.org/wiki/Parallel_databases
- [13] <https://www.techwalla.com/articles/the-advantages-of-parallel-database>
- [14] Saravanakumar Kandasamy - Explore Database - <https://exploredatabase.blogspot.in/2014/03/parallel-database-intraquery-parallelism.htm>
- [15] <https://www.quora.com/What-are-the-advantages-and-disadvantages-of-parallel-processing>
- [16] Manoj Kumar Sah 1, Vinod Kumar 2, Ashish Tiwari 3,”Security and Concurrency Control in Distributed Database System”,International Journal of scientific research and management (IJSRM) ||Volume||2||Issue||12||Pages||1839-1845||2014||
- [17] Arun Kumar Yadav& Ajay Agarwal,”An Approach for Concurrency Control in Distributed Database System”, Vol. 1, No. 1, January-June (2010)
- [18] <http://ecomputernotes.com/database-system/adv-database/data-replication>
- [19] https://www.google.co.in/search?q=difference+between+distributed+and+parallel+database&source=lnms&tbm=isch&sa=X&ved=0ahUKEwisqvfa6sPTAhUTTI8KHVWwBKUQ_AUIBigB&biw=1093&bih=530#imgrc=2L_SHZ6LBouToM:
- [20] https://www.google.co.in/search?q=client+server+architecture+in+distributed+database+diagrams&noj=1&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjScKLu8HTAhUiT48KHftkDs8Q_AUICcgB&biw=1093&bih=490#imgrc=_486mjiOFgjqkM:
- [21] https://www.google.co.in/search?q=architecture+of+parallel+database&noj=1&source=lnms&tbm=isch&sa=X&ved=0ahUKEwisk9vQwcHTAhXGPo8KHSsANcQ_AUICcgB&biw=1093&bih=490#imgrc=UsisSEXSCeSztM:
- [22] <https://www.google.co.in/search?q=homogeneous+distributed+database+architecture&tbm=isch&tbo=u&source=univ&sa=X&ved=0ahUKEwj1l8eE6sHTAhXDuo8KHYYTBPQQsAQIUA&biw=1093&bih=490&dpr=1.25#imgrc=2YCXbhzhNWJmbM:>



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)