ISSN No. 0976-5697

Volume 2, No. 5, Sept-Oct 2011

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A Framework of Distributed Database Management Systems in the Modern Organization and the Uncertainties removal

Prof. (Dr.) Anand K. Tripathi*
Head of Deptt. CSE &IT
SR Group of Institutions,CSE,Jhansi ,India
dr.aktripathi@gmail.com

Ms.Monika Tripathi
Asstt. Prof., Arya Kanya PG College,Jhansi ,India
monikatripathi.d@gmail.com

***Abstract* :**This research paper studies the use of distributed database management systems (DDBMSs) in the information infrastructure of modern organizations to reduce the uncertainties occurring in organization. The key purpose of the research is to determine the feasibility and applicability of DDBMSs for today's business applications. The forces which drove the selection of this topic were the improvements of distributed features in leading database management systems (DBMSs) in recent years, as well as the potential of distributed databases to provide competitive advantages for organizations for proper utilization of infrastructure to obtain the meaningful information.

***Keywords*:** Distributed Database Management Systems: DDBMS, DBMS

I. PRESENT STATE OF KNOWLEDGE

Today's business environment has an increasing need for distributed database and client/server applications as the desire for reliable, scalable and accessible information is steadily rising. Distributed database systems provide an improvement on communication and data processing due to its data distribution throughout different network sites. DDS makes not only makes data access faster, but a single-point of failure is less likely to occur, and it provides local control of data for users. However, there is some complexity when attempting to manage and control distributed database systems.

The information requirements of organizations and distributed database technology both have grown very tremendously in recent years. In fact, nearly all modern DBMSs come standard with powerful distributed features, but these features must be implemented and administered by skilled professionals. Distributed databases are much more complex than their centralized database cousins, but when properly implemented in the appropriate enterprise applications, they can provide great benefits to the organizations they support.

A. *Demands for DDBMSs:*

Prior to the popular acceptance of DDBMSs, corporations normally relied on centralized databases designed to serve very structured information requirements. These centralized databases had some characteristics in common. First, they ran on powerful and expensive hardware that could handle very large portions of a firm's data reliably. Second, they were administered by a small number of well-trained people who could manage the organization's computers to reduce the uncertainties occurring in organizations [1].Third, the dedicated data lines forming the corporate wide area network (WAN) had to be highly reliable and have a large capacity, because any downtime will preclude at least one site from operating, and every operation had to be transmitted to and from the central database in real time. These centralized databases could provide adequate performance to firms able to work around their shortcomings. These shortcomings include the lack of flexibility in the application of the firm's information and the requirement to implement a single point of failure for the entire enterprise.

This section explores the lessons learned about the limitations of centralized database systems over the thirty years they have been in general use. First the **business forces** are explored. Each of these business issues has generated information technology requirements that distributed database architectures are uniquely capable of supporting. Second, the **technology issues** are explored. These have come about from advances in information technology that have made the centralized database model **less relevant** in today's organizations.

II. BUSINESS FORCES

A. *Geographic Dispersion:*

Geographic dispersion of organizations is not an entirely new concept. Large firms have connected major regional offices to their centralized databases using dedicated lines for years. The difference now is that geographic dispersion is taken to greater extremes to provide cost savings and improved contact with the firm's customers [10]. Large regional offices are increasingly replaced with smaller locations in all of the firm's markets. This change greatly increases the number of dedicated lines which, if provided at the same service levels as in the older centralized systems, could add up to an enormous expense. Clearly, the traditional centralized database model creates a problem for firms wishing to benefit by such increased geographic dispersion.

Another aspect of the geographic dispersion problem is the growing abundance of portable computer use by mobile professionals. A common example of this is the traveling salesperson using a laptop-based database to query available inventory and take customer orders. The nature of this work prevents a full-time network connection, and the database on the mobile system must somehow be linked to the firm's master database at regular intervals to update the distributed copies of any data that has been changed. This is another case where geographic dispersion has rendered the centralized database architecture obsolete [11].

Geographically dispersed organizations require an architecture that allows the bulk of data retrieval and updates to be performed on fast and inexpensive local area networks (LANs). This architecture should reserve the more expensive WAN for data updates that are relevant to other sites. Mobile users should have a copy of the data for their local use and an efficient means to update using a part-time connection [12].

B. Information as a Resource:

Business leaders today understand the importance of information as a business resource. With centralized database systems, an organization's information is maintained and controlled by a few highly skilled individuals at one location [8]. Two major factors have led many business users to reject the centralized database model: the natural tendency for humans not to share and the introduction of personal computer (PC) -based DBMSs powerful enough to handle many concurrent users. Armed with such tools, departments and workgroups can easily build their own databases, wresting control of the information resource from the administrators of the organization's central databases and satisfying their natural tendency not to share.

The explosion of individual databases running on PC platforms can provide new opportunities to heads of departments, but may also pose problems for the organization as a whole. Information that could benefit the entire organization often becomes out of reach for users unable to access it or unaware of its existence. Additionally, because of the cheaper hardware and software used, and generally lower skills of the personnel administering these systems, reliability can be significantly less than with centralized systems. Data inconsistency is another problem that occurs in such an environment, as the same data is stored in many databases with no system for managing the multiple copies [12].

The centralized and decentralized models described above both generate major problems for large organizations. Some type of architecture that provides the advantages of both without the drawbacks would be ideal[2]. This architecture should allow decentralized use of data, while providing for database administration that can be performed by personnel with the interests of the whole firm in mind [2].

C. Corporate Rightsizing:

Modern corporations expand and contract frequently as they respond to changing competitive pressures. A study of 3,628 companies done by Cascio and Young reported in Morris, and found that one third had fired at least 15% of their employees during the period of the study . The study also concluded that in most cases, the companies had expanded to their original sizes, often within less than three years. Such activity is referred to as corporate downsizing or rightsizing, not to be confused with the same terms applied to information technology and client-server systems. It is often through the use of information technology that executives identify such business opportunities and transmit the decisions and plans to make the changes very rapidly. Ironically, it is often the information technology resource of an organization that is usually the least able to respond to such rightsizing decisions.

Centralized databases running on complex and expensive mainframes and minicomputers are usually very difficult to scale to high degrees. Adding or removing processing capacity and storage can be expensive and difficult. Many organizations require a scalable database system that can allow system administrators to handle changing demand with nothing more than the incremental purchase or removal of commodity hardware and software. Such a solution should provide a growing firm with a solution that allows rapid integration into the existing architecture and a predictable increase in capacity and performance [7].

III. TECHNOLOGY FORCES

A. Infusion of PCs and LANs in the Workplace:

The stage for distributed databases was set in the 1980s when PCs began to take hold of the corporate desktop in large numbers. The natural extension of these machines being on many desks was to connect them using local networks and servers. Office fileservers provided small organizations with decentralized server power. The culture and infrastructure of corporate computing reflected an increasingly decentralized bias, driven in many cases by end-users who began to understand and explore the power of decentralized computing

B. Increasing Demands of the Internet:

The growth in Internet use and the explosion of web pages with real-time information have dramatically increased the demands on business web sites in recent years. Web pages full of dynamic

IV. BROAD OUTLINES OF THE WORK

Distributed database systems are based upon several models and their implementations can include a number of different features. This has developed from the many varied situations and requirements organizations are faced with in putting the technology to use. The topics described below are key to understanding the capabilities and limitations of distributed database systems.

A. Fragmentation:

The fragmentation technique for distributed databases involves splitting the centralized database into portions and moving them to different locations. This distribution is accomplished by horizontal and/or vertical partitioning. No data is stored redundantly with the exception of primary keys in the case of vertical fragmentation. Using the relational model, horizontal fragmentation is accomplished by separating rows and vertical fragmentation is accomplished by separating columns. Data is normally fragmented according to the section of the organization which uses or modifies the data most frequently. For example, a firm may use a department code field to determine which department is responsible for each record and where the data should physically reside in a horizontally fragmented system.

Key principles of the fragmented distribution model are that only one copy of the data exists in the database, and that ownership and ability to update the database are shared. This model is similar to the centralized model in that the data is always consistent and current. The only redundancies exist with primary key fields when using vertical fragmentation. Fragmented database systems are more complex than centralized systems, but simpler than replicated systems. The issue of a single point of failure is reduced, but not eliminated. Network usage is generally lower in fragmented systems than in centralized systems.

B. Failure Recovery:

One of the advantages of replicated database systems is that they can provide a level of fault-tolerance beyond what can be achieved through more traditional means such as the use of redundant array of inexpensive disks (RAID). By replicating the database so that it is on two separate machines in different physical locations on the network, the probability that failures will cause a loss of service is significantly reduced. Two options for implementing failure recovery through database replication are available: warm standby and hot standby.

Warm standby uses asynchronous replication to maintain the standby server in a state nearly consistent with that of the primary server. Due to the lag between transactions being committed on the primary server and replication to the standby server, a small number of transactions are normally lost during a primary server failure and switchover to the standby server.

Hot standby uses synchronous replication to maintain the standby server in a state always consistent with the primary server. From an availability perspective this is the preferred solution, but the higher costs and potential lower performance of synchronous replication databases cause many organizations to select a warm standby solution. Buretta recommends a combination of local hot standby, normally RAID, and offsite warm standby server.

V. PRIMARY WORK DONE ON THE LINES

In the earliest days of centralized databases, professionals normally used a one-size fits all approach to DBMS software.

As database systems grew, increased in importance to the organization, and began operating in diverse applications, it became apparent that DBMSs should be specialized. One of the major shifts occurred when data centers running mainframes with large relational databases optimized for transaction speed began to notice poor performance during times when reports and queries were processed. It was at this point that the different requirements of online transaction processing (OLTP) and online analytical processing (OLAP) became apparent. Some organizations responded to the performance issues by restricting OLAP to late night and other off-peak times. This is less than an ideal solution as it limits the use of OLAP for the competitive advantage it should provide, and may be impossible when an organization operates round the clock or in many time zones.

A. Security:

Implementing effective security in a widely distributed database is no small task. It is observed that possible security services in a multitier architecture include authentication, authorization, nonrepudiation, confidentiality, and data integrity. Authentication is the process of having each user, host, or application server prove it self that who they are really. Authorization is the process of ensuring that each authenticated user has the necessary permission level to perform the requested tasks. Nonrepudiation is ensuring that authenticated and authorized users may not deny that they used a designated resource. Confidentiality prevents unauthorized users from accessing sensitive data. Data integrity prevents data from being modified in an unauthorized manner (11).

This makes some recommendations for implementing security in a replicated database environment. The first is that all stored and/or displayed passwords must be encrypted so that unauthorized persons and processes may not obtain them. Pseudo-user accounts, those established for systems to automatically log on to the network, are common in distributed database environments. Buretta points out that these accounts must comply with the firm's security policies and knowledge of their passwords should be limited. All file systems, raw devices, and/or database structures used to store queued data and/or messages must be secure. This item points out the many avenues in a distributed system available to unauthorized users, which must be protected. Finally, encryption techniques must be integrated within the replication service. This prevents interception of the data transmitted over the network (12).

This makes the point that distributed database systems may use either application- or data-level security. Application-level security, as its name suggests, is programmed into the application logic. Each application is responsible for governing user access to the data. Data level security is implemented in the database engine. Profiles of acceptable data items and operations are stored and checked by the database engine against the end-user's permission level on each database operation. Burleson recommends that application-level security be removed and replaced with data level security to make the distributed database more secure . The argument for this is that a skilled end-user with a workstation and commonly available development tools could easily write an application that does not follow the organization's security policy [11]. Such a security hole may be created either unintentionally by a well-meaning employee or intentionally by someone with malicious intent. When data-level security is implemented, such security holes are not possible (11)

B. The Future Prospects for DDBMSs:

The issues above demonstrate that a good DDBMS must provide security services, and that organizations must know how to properly implement them [4]. As data is distributed and end-users are given more processing power, potential for security problems increases. Organizations with distributed databases must be competent and vigilant in their execution of security.

DDBMS technology has potential, but its continued growth in popularity is not guaranteed. Just as DDBMSs grew in popularity with client-server, the potential for further growth in their popularity will likely be tied to client-server. The following section examines the potential for future success or failure of DDBMS systems with a focus on client-server trends as indicators of what the future may hold.

C. Growth of Internet Computing:

Experience with distributed database client-server architectures has shown that the complexity and expense of these approaches can be overwhelming. As a result, some firms have decided to go back in the direction of centralized databases. Professional have responded to this demand by providing solutions using servers based on mainframe or minicomputer platforms and thin clients, which in most cases run only a web browser. This architecture is sometimes referred to as Internet computing.

Proponents of Internet computing claim that simplifying the distributed components of the architecture and moving data to one professionally-managed location provides higher reliability and lower operating costs. One of the original arguments for client-server was the ability to replace character-based terminals with GUI-based workstations, which are more flexible and easier to use. Internet computing retains the benefits of a central data store and GUI-based workstations. The benefits of Internet computing may generate stiff competition for widely distributed database systems.

D. Immaturity of Client-Server:

Although client-server technology has been in widespread use for over a decade, some argue that it is not yet developed to the level to provide sufficient advantages to businesses implementing new systems. A panel of industry experts speaking at the Client-Server Leadership Forum in Toronto in 1996 concluded that the client-server market is still in adolescence. The panel reiterated the popular view that three-tiered client-server systems are far superior to two-tiered systems. Three-tiered architectures require more resources to implement, but are generally more scalable and allow for thinner, easier to maintain clients than two-tiered architectures does. The panel saw the client-server industry as immature, due to the low numbers of three-tiered systems implementation. Many professionals felt that until three-tiered systems become the norm, the benefits of client-server architectures cannot be realized. Some of the members felt that this weakness may lead to client-server being replaced by Internet computing Lack of DDBMS Standards:

A distributed database needs these four issues for DDBMSs to reach at their full potential. 1. Hardware independence, 2. Operating system independence, 3. Network independence, and 4. Database independence. Today's DDBMS products are still do not meet these four standards.

DDBMS technology is relatively new, and is still suffering from professionals fighting to develop and hold on to proprietary features. Today the situation is improving, but cross-vendor connectivity is sometimes limited, especially for legacy systems that do not implement newer standards.

E. DDBMSs Currently Available:

Several years ago finding the right tool for implementing distributed databases was a challenge due to the lack of DDBMSs available. Today distributed features are common in the latest DBMS offerings from all major vendors. In fact, it is rare for DBMSs with distributed features to be referred to as "distributed" databases at all - the feature is so prevalent that it does not distinguish one product from another. The major differences between products now are the technical details of how the data distribution is performed and the special features the DBMS provides.

The DBMS market is fiercely competitive, with no one professional dominating completely. According to Dataquest figures reported in *Computer Reseller News*, the 1998 database license revenue leaders were IBM with 32.3%, Oracle with 29.4%, Microsoft with 10.2%, Informix with 4.4%, and Sybase with 3.5%. IBM's lead is due primarily to its dominance in the mainframe and AS/400 platforms; on all other platforms Oracle is the leader. This section first provides an overview of features common among today's leading distributed database product offerings. Later it examines the differences between each vendor's products as well as their diverse strategies to provide an indication of the future offerings that may become available in the future.

F. Ability of DDBMSs to meet Business Requirements:

These section discuses organizational recently implemented distributed databases and the results they have obtained. These cases are representative of the use of distributed databases in business today [9]. Each case illustrates specific capabilities of today's DDBMSs. Later this section will analyze the capabilities of DDBMS products to meet the business and technology demands for them outlined earlier.

a. Analysis:

Recent developments in DDBMS technology, many of the business and technology requirements for distributed databases can be met. Below are the business and technology requirements discussed earlier with an analysis of the ability of today's DDBMS products to adequately meet the needs.

b. Geographic Dispersion:

DDBMS vendors have done a remarkable job of meeting the demands of businesses to support their geographically dispersed operations. Many of the advances in this area are due to the work in making efficient use of network connections. Network loads resulting from replication activity in recent versions is significantly reduced. Surridge Dawson's use of ISDN to replicate hundreds of thousands of transactions in only two hours every day illustrates this point. Replication to small client databases using occasional dial-up connections, typically with laptop computers, is also an area where recent work on lightweight replicating DBMS versions has provided organizations with a powerful tool for meeting the need to support mobile workers. With a careful analysis of business requirements and proper network design, DDBMSs can support most geographically dispersed business operations.

c. Control of the Information Resource:

DDBMSs now provide information managers with a means for centrally controlling and exploiting information scattered by server proliferation. Robust tools for handling heterogeneous server platforms and replication that is transparent to legacy applications are key factors in this. Northwest Airlines is a good example of how organizations can use the latest distributed database products to leverage existing hardware and software originally meant for use by only one element of the organization to make the entire organization more competitive [5]. Because the distributed database model works well with centralized planning and decentralized operations, database designers and administrators can maintain control of the firm's information while allowing it to be used flexibly.

d. Mergers and Acquisitions:

The distributed database products available now can facilitate mergers and acquisitions, especially if the organization is anticipating such activity and plans ahead. However, the previously cited experiences of Pricewaterhouse and other large corporations attempting to merge information resources with other firms on short notice highlights the limitations of DDBMSs and middleware to solve very complex problems[6]. Issues with different database schema, incompatible network infrastructure, and pressure to implement a solution rapidly are factors that will pose significant challenges to a smooth merger of information resources for many years to come. In these cases DDBMS software features are no substitute for high quality personnel and proper planning.

G. Corporate Rightsizing:

DBMS professionals continue to make advances in the scalability of their products - both in the capacity of individual servers and the quantity of distributed servers that may be included in a distributed database [3]. Modern database products give firms various options for growing or reducing their deployed databases. E-Plus is an example of a firm that was able to manage explosive growth while staying with one DBMS product family. Replication allows organizations to easily handle moving into new market areas. Surridge Dawson, for example, could open a new warehouse quickly by adding another slave server identical to the 21 it already has and connecting it with commercially available ISDN service. Current DDBMSs provide scalability adequate for most business applications, and the emphasis placed on this by DDBMS developers will ensure that those available in coming years will allow high degrees of scalability.

H. Client-Server Systems:

The latest versions of DDBMSs and middleware make a developer's task of implementing three-tier client-server architecture much simpler. Many of the components that formerly required a heavy programming effort are now available in off-the-shelf versions robust and flexible enough to handle most tasks. Organizations that integrate such products into their information architecture will reap the benefits of three-tier client-server. These architectures will allow for more flexibility and the ability to rapidly take advantage of business and technology opportunities that arise in the future. However, three tier client server systems will always require skilled planning and implementation to ensure the present and future needs of the firm can be met. This is another area where DDBMS features will never replace talented people.

I. Demands of the Internet:

Sea-Land Services is in a business normally considered low-tech, but it has exploited the Internet as a source of competitive advantage through its web site that allows customers to track shipments and enter bookings online. This is an example of new online services increasing loads on database servers as large numbers of firms leverage the Internet to provide improved service. By replicating the relevant data to a server dedicated for providing service to the web site, firms can reduce the impact of this increased demand on internal operations while not losing online customers as a result of excessive page generation times. DBMS developers have latched on to the Internet as a market for their products, and we will continue to see an increase in the power of these products to support e-business on the Internet.

VI. CONCLUSION

All of the major DBMS developers have made significant improvements to their newer products in the area of handling high loads of simultaneous OLTP and OLAP operations on the same server. Recent advances such as improved use of multiprocessor hardware, multithreading, and row-level locking have allowed this improved performance. However, there are still OLAP applications that generate such high system demands that they cannot function together effectively with OLTP applications on the same server. The replication features of today's major DBMSs fill this need nicely. Firms can use asynchronous replication to maintain an OLAP server separate from the OLTP server and provide high performance for both applications. Future advances in individual server capabilities to simultaneously support OLTP and OLAP plus improved replication performance will mean that IT managers will not need to compromise to provide high performance in both these areas.

VII. REFERENCES

[1]. Prof. (Dr.) Anand K. Tripathi and Mrs. Monica Tripathi: “MIS Uncertainties and the Computer Support”: An Analytical Approach”, Proceeding of the National Conference on Next Generation Computing Technology, at ITS Ghaziabad publication wisdom publication, Delhi-16-17 Oct-2008.,edition-3,vol-1

[2]. Prof. (Dr.) Anand K. Tripathi and Mrs. Monica Tripathi**:** “A Frame: Why Information System Fails”, Proceeding of the National Conference on “Managing Complexity in Informative World. ‘’ publication Bhilai Institute of Technology, Durg, Bhilai (CG) dated 7-8 Nov.-2008. edition-1,vol-1,pg.112-115

[3]. Prof. (Dr.) Anand K. Tripathi and Mrs. Monica Tripathi **,** AICTE Sponsored, National Conference on “Next generation

computing & Information systems" at Model Institute of Engineering & Technology, Jammu, Feb.14-15, 2009. edition-2,vol-1,pg. 724-731

[4]. Prof. (Dr.) Anand K. Tripathi and Mrs. Monica Tripathi **,**Paper "A Framework :A Role of Information Technology for Decision Support System", CSI Sponsored National level Conference on "Cutting Edge Computer and Electronics Technologies" under TEQIP organized by College of Technology, GB Pant University, Pant Nagar,14-15 Feb.2009, edition-4,vol-1, ,pg. 915-919

[5]. (Dr.) Anand K. Tripathi and Mrs. Monica Tripathi **,** National level Conference on **"**Computer Intelligence**"** at Pioneer Institute, Indore, May-2009,Ediotion-1,Vol.-1, ,pg. 724-729

[6]. (Dr.) Anand K. Tripathi and Mrs. Monica Tripathi Paper "Information Technology to Transform the Business "Proceeding of National Conference on Next Generation Technologies for Information Management (NGTIM-2009) ,7th Nov.-2009 ,organized by ITS-Management & IT Institute, Ghaziabad (UP),ISBN :978-81-89547-69-1, edition-1,vol-1, ,pg. 1724-1731

[7]. "Information Technology and Structuring Of Organization: Information System Research" Published in IMSEC International Journal of Research IMS College Engineering, Ghaziabad, Vol.-1 July-Dec.-2009 (Referred Based**)** edition-1,vol-,1, ,pg. 925-928

[8]. "Impact of High Professional Teaching: Information Technology to Globalize the Business" Published in International Journal on Computer Engineering and Information Technology, Bhilai, ISSN 0974-2034 (Referred Based), edition-1,vol-1, ,pg. 815-818

[9]. Knowledge, Skills and Abilities: Recruitment IS/IT Professionals in Fast Changing Business Landscape Published in International Journal on Computer Engineering and Information Technology, Bhilai, edition-1,vol-1, ,pg. 606-609

[10]. Cost Benefit Analysis against Computerization of Oranization."Journal Prestige Institute of Management, Dewas (MP) Vol.-1 No. 1,MPENG00969/12/1/2009-TC,Jan. 2010 , edition-1,vol-1,,pg. 724-731

[11]. "Data Security levels for Computerized MIS" Artificial Intelligence & its Applications (AIA – 2009) , Department of Computer Science & Engineering, Faculty of Engineering & Technology, RBS College, Bichpuri, Agra (UP), 12-13 September 2009, edition-1,vol-1, ,pg. 555-559

[12]. Role of IT Infrastructure for Business Transformation Proficient-An International Journal of Management, Malviya Nagar, Jaipur,Rajastan ,ISSN No.0975-475X,Vol.-1,July-Dec.-2009 (Referred Based), edition-4,vol-1, ,pg. 789—793

