ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

Improved Performance of Web Based Database Management for Telemedicine by Using Three Fold Approach of Data Fragmentation,Websites Data Clustering and Data Allocation

Sneha Katale[1], Sonali Hotkar[2], Abhijeet Kashid[3], Jyoti Dange[4], Ms.Nisha Kimmatkar[5]

Student, Department of Computer Engineering, JSPM's Rajarshi Shahu College of Engineering, Tathawade, Pune Savitribai Phule Pune University. India[1234]

Assistant Professor Department of Computer Engineering, JSPM's Rajarshi Shahu College of Engineering, Tathawade, Pune, Savitribai Phule Pune University. India[5]

ABSTRACT: Many web computing systems are running real time database services where their information change continuously and expand incrementally. In this context, web data services have a major role and draw significant improvements in monitoring and controlling the information truthfulness and data propagation. Currently, web telemedicine database services are of central importance to distributed systems. However, the increasing complexity and the rapid growth of the real world healthcare challenging applications make it hard to induce the database administrative staff. In this paper, we build an integrated web data services that satisfy fast response time for large scale Tele-health database management systems. Our focus will be on database management with application scenarios in dynamic telemedicine systems to increase care admissions and decrease care difficulties such as distance, travel, and time limitations. We propose three-fold approach based on data fragmentation, database websites clustering and intelligent data distribution. This approach reduces the amount of data migrated between websites during applications' execution; achieves cost effective communications during applications' processing and improves applications' response time and throughput. The proposed approach is validated internally by measuring the impact of using our computing services' techniques on various performance features like communications cost, response time, and throughput.

KEYWORDS: Data fragmentation, database website clustering, data distribution

1. INTRODUCTION

The rapid growth and continuous change of the real world software applications have provoked researchers to propose several computing services' techniques to achieve more efficient and effective management of web telemedicine database systems (WTDS). Significant research progress has been made in the past few years to improve WTDS performance. In particular, databases as a critical component of these systems have attracted many researchers. The web plays an important role in enabling healthcare services like telemedicine to serve inaccessible areas where there are few medical resources. It offers an easy and global access to patients' data without having to interact with them in person and it provides fast channels to consult specialists in emergency situations. Different kinds of patient's information such as ECG, temperature, and heart rate need to be accessed by means of various client devices in heterogeneous communications environments. WTDS enable high quality continuous delivery of patient's information wherever and whenever needed.
Several benefits can be achieved by using web telemedicine services including: medical consultation delivery, transportation cost savings, data storage savings, and mobile applications support that overcome obstacles related to the performance (e.g., bandwidth, battery life, and storage), security (e.g., privacy, and reliability), and environment (e.g., scalability, heterogeneity, and availability). The objectives of such services are to: (i) develop large applications that scale as the scope and workload increases, (ii) achieve precise control and monitoring on medical data to generate high

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

telemedicine database system performance, (iii) provide large data archive of medical data records, accurate decision support systems, and trusted event-based notifications in typical clinical centers.
Several benefits can be achieved by using web telemedicine services including: medical consultation delivery, transportation cost savings, data storage savings, and mobile applications support that overcome obstacles related to the performance (e.g., bandwidth, battery life, and storage), security (e.g., privacy, and reliability), and environment (e.g., scalability, heterogeneity, and availability). The objectives of such services are to: (i) develop large applications that scale as the scope and workload increases,(ii)Achieve precise control and monitoring on medical data to generate high telemedicine database system performance,(iii)Provide large data archive of medical data records, accurate decision support systems, and trusted event-based notifications in typical clinical centers.
Recently, many researchers have focused on designing web medical database management systems that satisfy certain performance levels. Such performance is evaluated by measuring the amount of relevant and irrelevant data accessed and the amount of transferred medical data during transactions' processing time. Several techniques have been proposed in order to improve telemedicine database performance, optimize medical data distribution, and control medical data proliferation. These techniques believed that high performance for such systems can be achieved by improving at least one of the database web management services, namely database fragmentation, data distribution, websites clustering, distributed caching, and database scalability. However, the intractable time complexity of processing large number of medical transactions and man- aging huge number of communications make the design of such methods a non-trivial task. Moreover, none of the existing methods consider the three-fold services together which makes them impracticable in the field of web data- base systems. Additionally, using multiple medical services from different web database providers may not fit the needs for improving the telemedicine database system performance. Furthermore, the services from different web data- base providers may not be compatible or in some cases it may increase the processing time because of the constraints on the network . Finally, there has been lack in the tools that support the design, analysis and cost-effective deployments of web telemedicine database systems.
Designing and developing fast, efficient, and reliable incorporated techniques that can handle huge number of medical transactions on large number of web healthcare sites in near optimal polynomial time are key challenges in the area of WTDS. Data fragmentation, websites clustering, and data allocation are the main components of the WTDS that continue to create great research challenges as their cur- rent best near optimal solutions are all NP-Complete.

SYSTEM ARCHITECTURE

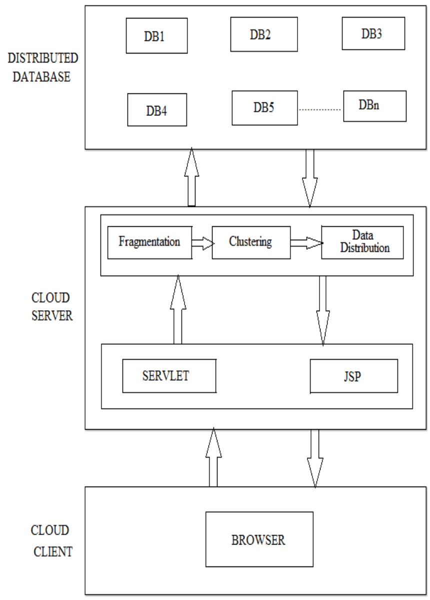

Fig 1: Telemedicine Architecture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

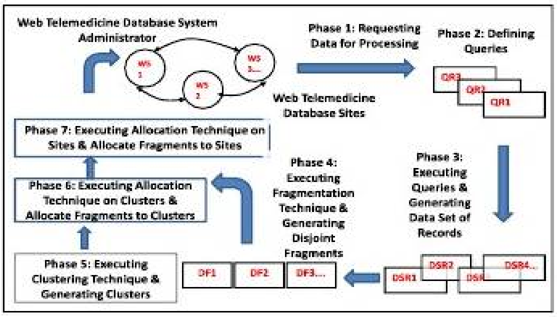

Fig 2: IFCA Computing Services Architecture

To improve the performance of medical distributed data- base systems, we incorporate data fragmentation, websites clustering, and data distribution computing services together in a new web telemedicine database system approach. This new approach intends to decrease data communication, increase system throughput, reliability, and data availability.

The decomposition of web telemedicine database relations into disjoint fragments allows database transactions to be executed concurrently and hence minimizes the total response time. Fragmentation typically increases the level of concurrency and, therefore, the system through- put. The benefits of generating telemedicine disjoint fragments cannot be deemed unless distributing these fragments over the websites, so that they reduce communication cost of database transactions. Database disjoint fragments are initially distributed over logical clusters (a group of websites that satisfy a certain physical property, e.g., communications cost). Distributing database disjoint fragments to clusters where a benefit allocation is achieved, rather than allocating the fragments to all web- sites, have an important impact on database system throughput. This type of distribution reduces the number of communications required for query processing in terms of retrieval and update transactions; it has always a significant impact on the web telemedicine database sys- tem performance. Moreover, distributing disjoint fragments among the websites where it is needed most, improves database system performance by minimizing the data transferred and accessed during the execution time, reducing the storage overheads, and increasing availability and reliability as multiple copies of the same data are allocated.

Database partitioning techniques aim at improving data- base systems throughput by reducing the amount of irrelevant data packets (fragments) to be accessed and transferred among different websites. However, data fragmentation raises some difficulties; particularly when web telemedicine database applications have contradictory requirements that avert breakdown of the relation into mutually exclusive fragments. Those applications whose views are defined on more than one fragment may suffer performance ruin. In this case, it might be necessary to retrieve data from two or more fragments and take their join, which is costly. Data fragmentation technique describes how each fragment is derived from the database global relations. Three main classes of data fragmentation have been discussed in the literature; horizontal [5], vertical and hybrid. Although there are various schemes describing data partitioning, few are known for the efficiency of their algorithms and the validity of their results.

The clustering technique identifies groups of network sites in large web database systems and discovers better data distributions among them. This technique is considered to be an efficient method that has a major role in reducing the amount of transferred and accessed data during processing database transactions. Accordingly, clustering techniques help in eliminating the extra communications costs between websites and thus enhances distributed data- base systems performance. However, the assumptions on the web communications and the restrictions on the number of network sites, make clustering solutions impractical. Moreover, some constraints about network connectivity and transactions processing time bound the applicability of the proposed solutions to small number of clusters [12].

Data distribution describes the way of allocating the disjoint fragments among the web clusters and their respective sites of the database system. This process addresses the assignment of each data fragment to the distributed database websites [8]. Data distribution related techniques aim at improving distributed database systems performance. This can be accomplished by reducing the number of database fragments that are transferred and accessed during the execution time. Additionally, Data distribution techniques attempt to increase data availability, elevate database reliability, and reduce storage overhead. However, the restrictions on database retrieval and update frequencies in some data allocation methods may negatively affect the fragments distribution over the websites.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

II. RELATED WORK

Many research works have attempted to improve the performance of distributed database systems. These works have mostly investigated fragmentation, allocation and sometimes clustering problems. In this section, we present the main contributions related to these problems, discuss and compare their contributions with our proposed solutions.

1.1 Data Fragmentation

With respect to fragmentation, the unit of data distribution is a vital issue. A relation is not appropriate for distribution as application views are usually subsets of relations. Therefore, the locality of applications' accesses is defined on the derivative relations subsets. Hence it is important to divide the relation into smaller data fragments and consider it for distribution over the network sites. The authors in [5] considered each record in each database relation as a disjoint fragment that is subject for allocation in a distributed database sites. However, large number of database fragments is generated in this method, causing a high communication cost for transmitting and processing the fragments. In contrast to this approach, the authors in [11] considered the whole relation as a fragment, not all the records of the fragment have to be retrieved or updated, and a selectivity matrix that indicates the percentage of accessing a fragment by a transaction is proposed. However, this research suffers from data redundancy and fragments overlapping.

1.2 Clustering Websites

Clustering service technique identifies groups of networking sites and discovers interesting distributions among large web database systems. This technique is considered as an efficient method that has a major role in reducing transferred and accessed data during transactions processing [9]. Moreover, grouping distributed network sites into clusters helps to eliminate the extra communication costs between the sites and then enhances the distributed database system performance by minimizing the communication costs required for processing the transactions at run time. In a web database system environment where the number of sites has expanded tremendously and amount of data has increased enormously, the sites are required to manage these data and should allow data transparency to the users of the database. Moreover, to have a reliable database system, the transactions should be executed very fast in a flexible load balancing database environment. When the number of sites in a web database system increases to a large scale, the problem of supporting high system performance with consistency and availability constraints becomes crucial. Different techniques could be developed for this purpose; one of them is websites clustering. Grouping websites into clusters reduces communications cost and then enhances the performance of the web database system. However, clustering network sites is still an open problem and the optimal solution to this problem is NP-Complete . Moreover, in case of a complex network where large numbers of sites are connected to each other, a huge number of communications are required, which increases the system load and degrades its performance. The authors in [13] have proposed a hierarchical clustering algorithm that uses similarity upper approximation derived from a tolerance(similarity) relation and based on rough set theory that does not require any prior information about the data. The presented approach results in rough clusters in which an object is a member of more than one cluster. Rough clustering can help researchers to discover multiple needs and interests in a session by looking at the multiple clusters that a session belongs to. However, in order to carry out rough clustering, two additional requirements, namely, an ordered value set of each attribute and a distance measure for clustering need to be specified [12]. Clustering coefficients are needed in many approaches in order to quantify the structural network properties. The outcomes of this method declare that the average shortest distance in the node's neighborhood is smaller than all network distances. However, independent constant values and natural logarithm function are used in the shortest distance approximation function to determine the clustering mechanism, which results in generating small number of clusters.

1.3 Data Allocation (Distribution)

Data allocation describes the way of distributing the database fragments among the clusters and their respective sites in distributed database systems. This process addresses the assignment of network node(s) to each fragment. However, finding an optimal data allocation is NP-complete problem [4]. Distributing data fragments among database websites improves database system performance by minimizing the data transferred and accessed during execution, reducing the storage overhead, and increasing availability and reliability where multiple copies of the same data are allocated. Many data allocation algorithms are described in the literature. The efficiency of these algorithms is measured in term of response time. Authors in [13] proposed an approach that handles the full replication of data allocation in database systems. In this approach, a database file is fully copied to all participating nodes through the master node. This approach distributes the sequences through fragments with a round-robin strategy for sequence input set already ordered by size, where the number of sequences is about the same and number of characters at each fragment is similar.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

However, this replicated schema does not achieve any performance gain when increasing the number of nodes. When a non-previously determined number of input sequences are present, the replication model may not be the best solution and other fragmentation strategies have to be considered. In [4], the author has addressed the fragment allocation problem in web database systems. He presented an integer programming formulations for the non-redundant version of the fragment allocation problem. This formulation is extended toaddress problems, which have both storage and processing capacity constraints. In this method, the constraints essentially state that there has been exactly one copy of a fragment across all sites, which increase the risk of data inconsistency and unavailability in case of any site failure. However, the fragment size is not addressed while the storage capacity constraint is one of the major objectives of this approach. In addition, the retrieval and update frequencies are not considered in the formulations, they are assumed to be the same, which affects the fragments distribution over the sites. Moreover, this research is limited by the fact that none of the approaches presented have been implemented and tested on a real web database system. A dynamic method for data fragmentation, allocation, and replication is proposed in [25]. The objective of this approach is to minimize the cost of access, re-fragmentation, and reallocation. DYFRAM algorithm of this method examines accesses for each replica and evaluates possible re-fragmentations and reallocations based on recent history. The algorithm runs at given intervals, individually for each replica. However, data consistency and concurrency control are not considered in DYFRAM. Additionally, DYFRAM doesn't guarantee data availability and system reliability when all sites have negative utility values. In [1], the authors present a horizontal fragmentation technique that is capable of taking a fragmentation decision at the initial stage, and then allocates the fragments among the sites of DDBMS. A modified matrix MCRUD is constructed by placing predicates of attributes of a relation in rows and applications of the sites of a DDBMS in columns. Attribute locality precedence ALP; the value of importance of an attribute with respect to sites of distributed database is generated as a table from MCRUD. However, when all attributes have the same locality precedence, the same fragment has to be allocated in all sites, and a huge data redundancy occurs. Moreover, the initial values of frequencies and weights don't reflect the actual ones in real systems, and this may affect the number of fragments and their allocation accordingly. The authors in [5] presented a method for modeling the distributed database fragmentation by using UML 2.0 to improve applications performance. This method is based on a probability distribution function where the execution frequency of a transaction is estimated mainly by the most likely time. However, the most likely time is not determined to distinguish the priorities between transactions. Furthermore, no performance evaluations are performed and no significant results are generated from this method. A database tool shown in addresses the problem of designing DDBs in the context of the relational data model. Conceptual design, fragmentation issues, as well as the allocation problem are considered based on other methods in the literature. However, this tool doesn't consider the local optimization of fragment allocation problem over the distributed network sites. In addition, many design parameters need to be estimated and entered by designers where different results may be generated for the same application case.

III. ACKNOWLEDGMENTS

It gives us great pleasure in presenting the preliminary project report on "Improved performance of web based database management for telemedicine by using three fold approach of data fragmentation,websites data clustering and data allocation".I would like to take this opportunity to thank my internal guide Ms.N.V.Kimmatkar for giving me all the help and guidance i needed. We are really grateful to them for their kind support. Their valuable suggestions were very helpful. We are also grateful to Prof. Bagvan, Head of Computer Engineering Department, JSPM'S Rajarshi Shahu College Of Engineering Tathawade, Pune-33 for his indispensable support, suggestions.

IV. CONCLUSION

We proposed a new approach to promote telemedicine system performance. Our approach integrates three enhanced computing services' techniques namely, database fragmentation, network sites clustering and fragments allocation. We develop these techniques to solve technical challenges, like distributing data fragments among multiple web servers, handling failures, and making tradeoff between data availability and consistency. We propose an estimation model to compute communications cost which helps in finding cost-effective data allocation solutions. The novelty of our approach lies in the integration of web database sites clustering as a new component of the process of WTDS design in order to improve performance and satisfy a certain level of quality in web services.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

REFERENCES

[1] Shahidul Islam Khan,Dr.A.S.M.Latiful Hoque, "A New Technique for Database Fragmentation in Distributed Systems," Int'l J.Computer Applications, vol. 5, no. 9, 2010,pp. 20-24.

[2] B. H. Guze, R. Estep, and C. Fisher, "Telemedicine: A review of its use and a proposal for application in psychiatric consultation," Med. Inform., vol. 20, no. 1, pp. 1–18, 1995.

[3] C. Danilowicz and N. Nguyen, "Consensus Methods for Solving Inconsistency of Replicated Data in Distributed Systems," Distributed and Parallel Databases, vol. 14, 2003,pp. 53-69.

[4] Syam Menon, Member, IEEE," Allocating Fragments in Distributed Databases" IEEE transactions on parallel and distributed systems, Vol. 16, No. 7, JULY 2005

[5] A. Tamhanka and S. Ram, "Database Fragmentation and Allocation: An Integrated Methodology and Case study," IEEE Trans. Systems, Man and Cybernetics, Part A: systems and Humans, May,vol. 28, no. 3, 1998 pp. 288-305.

[6] T. Paakkala, J. Aalto, V. K¨ah¨ar¨a, and S. Sepp¨anen "Diagnostic performance of a Tele-radiology system in primary health care."

[7] H. K. Huang, "Tele radiology technologies and some service models," Computer. Med. Image. Graph. vol. 20, no. 2, pp. 59–68, 1996.

[8] J. E. Cabral, Jr. and Y. Kim Multimedia systems for telemedicine and their communications requirements.

[9] O. Ratib, Y. Ligier, and J. R. Scherrer, "Digital image management and communication in medicine," Comput. Med. Imag. Graph. vol. 18, no.2, pp. 73–84, 1994.

[10] H. K. Huang et al., "Implementation of a large-scale picture archiving and communication system," Computer. Med. Image. Graph. vol. 17, no. 1, pp. 1–11, 1993.

[11] D. F. Leotta and Y. Kim, "Requirements for picture archiving and communications," IEEE Eng. Med. Biol. Mag. , pp. 62–69, Mar. 1993.

[12] A. Jain, M. Murty, and P. Flynn "Data Clustering: A Review," ACM Computing Surveys, vol. 31, no. 3, 1999, pp. 264-323.

[13] Zahid Pervaiz, Walid G. Aref, Senior Member, IEEE, Arif Ghafoor, Fellow, IEEE, and Nagabhushana Prabhu," Accuracy-Constrained Privacy-Preserving Access Control Mechanism for Relational Data". IEEE Commune VOL. 26, NO. 4, pp. 1-13, APRIL 2014.