International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:12, No:6, 2018

Optimized Approach for Secure Data Sharing in Distributed Database

Ahmed Mateen, Zhu Qingsheng, Ahmad Bilal

***Abstract*—**In the current age of technology, information is the most precious asset of a company. Today, companies have a large amount of data. As the data become larger, access to data for some particular information is becoming slower day by day. Faster data processing to shape it in the form of information is the biggest issue. The major problems in distributed databases are the efficiency of data distribution and response time of data distribution. The security of data distribution is also a big issue. For these problems, we proposed a strategy that can maximize the efficiency of data distribution and also increase its response time. This technique gives better results for secure data distribution from multiple heterogeneous sources. The newly proposed technique facilitates the companies for secure data sharing efficiently and quickly.

***Keywords*—**ER-schema, electronic record, P2P framework, API, query formulation.

I. Introduction

THIS research will explore the improvement of the Electronic Records (ER) which means no more shuffling through paper files, waiting for faxes, or searching for paperwork to be able to provide quality and timely information with accuracy. Records management is often seen as an unnecessary or low priority administrative task that can be performed at the lowest levels within an organization. However, this perception is changing as these publicized events have demonstrated that records management is in fact the responsibility of all individuals within an organization [2]. In fact, the use of an ER can lead to records management problems, especially for government agencies with specific legal requirements. For example, an ER may improve collaboration during document development [15]. Before implementing a system, one must determine how it fits into his/her overall records management strategy. ER functionality is often integrated into Content Management (CM) systems. These systems combine additional functionality such as website management with workflow tools, standard templates, and access rights.

A. Problem Statement

Due to the decentralization, P2P systems promise an improved robustness and scalability and therefore open also a new view on data integration solutions. Thus, in the paper, by distinguishing between db-centric and P2P-centric features, we examine features common to these database systems as well as other ad-hoc features that solely characterize P2P databases. However, several design and technical challenges arise in building scalable P2P-based integration systems. We address one of them: the problem of distributed query processing. We discuss strategies of query decomposition and present results of an experimental evaluation.

Ahmed Mateen is with the College of Computer Science, Chongqing University, China (e-mail: ahmedbuttar@uaf.edu.pk).
Zhu Qingsheng is with the College of Computer Science, Chongqing University, China.
Ahmad Bilal is with the University of Agriculture Faisalabad, Pakistan.

B. Previous Work

The need for large-scale data sharing between autonomous and heterogeneous systems gave rise to the concept of P2P database systems, whereas a definition for a P2P database system can be readily provided. A comparison with the more established decentralized models, commonly referred to as distributed, federated and multi-databases, is more likely to provide a better insight to this new P2P data management technology [6]. Model and commands have been introduced, picking up the agreement of medicinal group. Its app for utilization among various healing facilities is a piece of another undertaking with respect to checking and apps that are at present behind investigating procedure of the centerline government organization. At long last, checks on agreeable estimations of execution and adaptability of the P2P are demonstrated [7].

This showed that the E-PR structure of data means to rearrange data getting and distributing in different offices, comparative cases have been dealt with and result treatments have been gathered, with movement relapse sound outcomes. Getting to comparable cases is helpful for planning suitable treatments. The medicinal area considered here is identified with growth ailment, and along these lines the meta-EPR has been outlined thinking about collaboration among oncology divisions, and has been guided by utilizing their necessities. When all these things are considered, this E-PR can be utilized for any office with suitable evolving. In the accompanying, the introduced XML composition of the E-PR is accounted [9].

Afterwards, depicted patient records are mapped into a basic XML-based meta-Electronic Patient Record. The E-PR is not a valid E-PR proposition, yet it is an efficient structure characterized to have important and total data removed from the diverse E-PRs embraced by every doctor's facility. Healing facility administrators define commands against E-PR construction; commands are then disseminated to the associated clinics facilitating E-PR occasions [2].

II. Materials & Methods

A. Developing Strategies

A vital procedure in building up another strategy is the brisk usefulness, trailed by time spent enabling the framework to

advance as more data are accumulated in short timing and efficient way. It bodes well to a suitable nature of prerequisites accumulated before starting the outline of the procedure. With an expansive number of data, the general stream of the procedure might not be quite the same as the one that is worked from a little arrangement of prerequisites. In this way, enough data must be assembled keeping in mind the end goal to have a fundamental thought of the best structure before starting to construct [10].

B. Need of Database Server in Distributed Environment

A DB server is the Oracle programming application for managing a DB, and a client is an application that requests information from a server. Every PC in a framework is a center point that can have more than one DBS. Each center point in a circled DB system can go regarding as a client and centralized system, or both, depending upon the situation. The host for the headquarter DB is going regarding as a DB server when a declaration appears against its close-by information, however, it is going as a client when it shows a declaration against remote information [3].

This examination will investigate the change in the ER which turns into a basic instrument for getting to, in a proficient way, data with respect to history and individual data required by the organization. The ER has raw material of various sorts, like numeric ones to images. Distributed Environment adds to the conventional centralized Data Base some other types of processing expenses, because of the additional design (hardware & software) to handle the distribution. These expenses present as the cost of data transfer over the network [1]. An extensive volume of data can be put away on every ER, except concerning any vast data holder, effective access strategies and conventions are required to access and read data in a productive way. A DBMS needs to support certain transaction properties to offer a reasonable access to the database. Usually, the properties are listed in four aspects of transactions: atomicity, consistency, isolation, and durability [6]. These are referred to as the ACID-properties of transactions. The distributed database technology intends to extend the concept of data independence to environments where data are distributed and replicated over several machines connected by a network [9].

C. The Experimental Model for Distributed Database Management System

These models are running with parts of the application processor: customers and neighborhood servers. A customer constitutes an interface for an end client. It is responsible for making the phonetic and semantic examination of client questions and they are debilitating into a course of action of subrequest tackling physical information things. The general improvement utilized by the customer is secured in the storeroom being a united neighboring database. Through an extra catapulting system self-decision of the locks utilized by information processors, the customer module controls the simultaneous access to the information, including general gridlock disclosure and confirmation. It likewise sends the sub inquiries to the adjoining servers, deals with the appropriated exchanges, and displays their outcomes [10].

A customer, overseeing general exchanges, cooperates with a zone server, which performs synchronous sub-exchanges with help of the contiguous DBMS. The help between a territory server and an information processor is master inferable from the inserting's SQL in C dialect and the externalization of the exchange commitment custom [8]. These procedures and modules have the capacity to still running in the structure and share resources and memory. Free procedures running on the structure go on through virtual shared memory (VSM) called tuple space (TS), passing information units called tuples. The structure, completed thusly, is used to shape a circumstance for making and testing computations of the picked DDBMS instruments: request getting ready, synchronization control, scattered trade duty, and keeping up data replication [13].

D. Transaction Management in Replicated Data for EDDBMS

The other portion, which is a replication director component, has been as of late addition to the framework, keeping in mind the end goal to keep up restructured information. What is more, the nearby customer module was made to deal with the neighborhood exchanges working on repeated information things. The joint working b/w these components is capable by portions of dynamic DBS: triggers, dB system, dB occasions and, by techniques for row space. Information about modules is expressed as primary/assistant and undefined duplicates, and is finished in the exploratory framework. In the essential scenario, the fast-working framework has been acknowledged in which tasks are running on the key duplicate, and submitted shapes are aggregated and sent to every single other duplicate as self-decision exchanges [1].

E. Locking Management in Distributed Databases

The strategy for data sharing between processors on a similar system uses in the movement for giving a data locking instrument or framework. A solitary token relocates between the processors for every unique data thing and its limitations. Just a processor having ownership of the token can play out specific operations simply like refreshing, transmitting, and refreshed the data to alternate processors [5]. Two distinctive lining courses of action are depicted for surrendering or holding ownership of the token when a processor, not possessing the token, makes a demand for the token.

At present, the two ways to deal with control the database simultaneousness first is the two-stage locking methodology and second is the timestamp approach. Both methodologies can expand the yield of the conveyed DDBMS. DDBMS will make virtual tables for that database per users. In this way, those data esteems can be gotten into perused mode by numerous clients and be accessible in compose mode by a client [5].

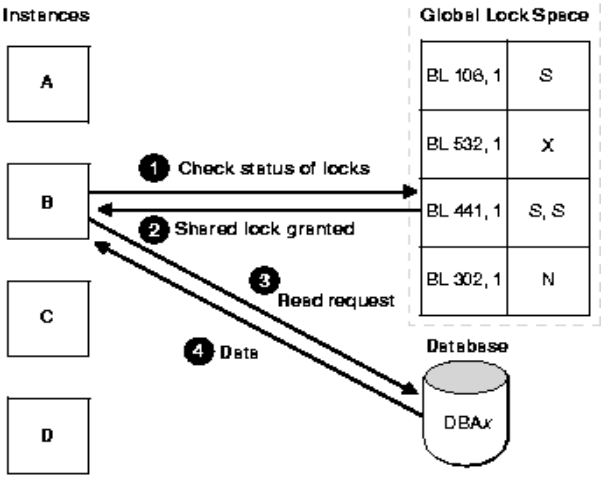

Fig. 1 Locking management in distributed databases

F. Two Phase Locking in Distributed Database

The two-stage locking approach is very important and key factor in DBMS Connection which will control all the database transactions read and write operations in term of execution, resource allocation, and scheduling the procedures to overcome the problem of deadlocks. This technique modifies and updates the resource allocation on shared data. If operation is on X row with the demand to read operation, then the read lock will be never released to any other transaction and similarly if operation S has writing request then the written lock will not be released to any other task simultaneously. After completing the read or write operation, the control will automatically release and be given to the next task stored in pool according to their demand request [5].

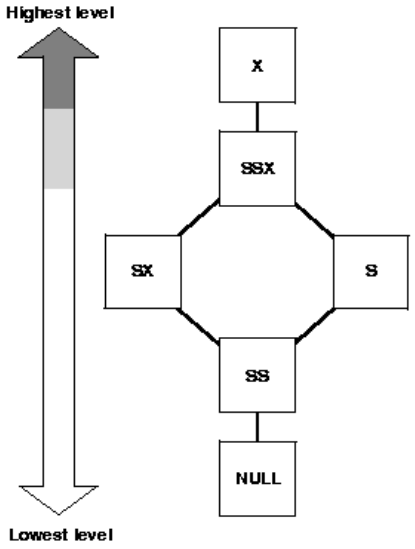

Fig. 2 Two phases locking approach

G. Timestamp Approach in Distributed Database

In the timestamp approach, a DDBMS appoints distinguishing proof number for each demand contingent on the time that each demand arrives the DDBMS. Besides doling out a Timestamp, DDBMS likewise perceives whether the demand is a perused task or composed task. The composed task has dependably a higher need than the reader task. The perused task can search up for more detail on the two-stage locking approach and the Timestamp approach in a simultaneousness control in disseminated database frameworks in Distributed System [11].

H. View and Timeout Approach

When there are more clients getting to the database, at that point there is a more noteworthy shot that at least two clients will get a similar snippet of data. In the two-stage locking approach, clients can either read or keep in touch with any data all the while.

I. Distributed Database Sharing System for Grid Environments

A lattice processing advancement gives frameworks that permit the consistent sharing of dispersed assets and heterogeneous assets, much the same as supercomputers, stockpiling frameworks and data sources, in an expansive and dynamic condition. The exploration works in data frameworks have for the most part centered on applications where data assets are put away in the records [4]. However, the majority of the logical applications are very reliant on database administration frameworks for arranging and putting away their data. This burden is the requirement for creating strategies that can empower the sharing and access to greatly convey heterogeneous databases, and self-ruling databases made and oversaw freely by a few research organizations. In this paper, we compose database sharing strategy that has the accompanying attractive attributes: (1) it bolsters data ordering at a better granularity than the document level, empowering content based inquiry and access (2) it permits data sharing without an internationally shared construction (3) it is recognized by its distributed flavor, making it exceptionally versatile, auto-configurable, and completely decentralized with no compelling reason to concentrated or committed servers [8].

J. Distributed Database Integrity

The distribution database integrity was the exploration exertion financed by Rome Laboratory and directed at the Computer Science Laboratory of SRI International. The principal motivation behind this exertion was to improve data uprightness and consistency administration procedures for database application conditions [13]. The target of this exertion was to look at the data trustworthiness, consistency and simultaneousness control strategies for circulated question arranged database administration frameworks. Particular zones that were examined include: (1) administer based simultaneousness control methods for dynamic preparing (2) versatile simultaneousness systems for managing consistency versus accessibility of data (3) fleeting database instruments for holding numerous renditions of database objects (4) standards, limitations, and application-particular learning bases for uprightness upkeep (5) dynamic triggers for

naturally authorizing respectability rules in view of nearby or remote updates, and (6) the impact of incrementally impermanent questions on disseminated database honesty [7].

K. Oracle Distributed Database Architecture

The appropriated database is an arrangement of databases put away on progressively the one PC that regularly appears to applications as a solitary database. Thus, an app can in the meantime get to and alter the data in numerous databases in a framework. Each Oracle DB in the framework is managed by its neighborhood Oracle server and coordinates to carry on the working in all over the world disseminated DB [19].

L. Clients and Servers

A database server is the Oracle programming managing a database, and a client is an application that requests information from a server. Every PC in a system is a center point. A center point in a scattered database system goes as a client or a server or both, depending upon the condition [14].

For an instance, the PC invokes the request to access the server database and a declaration is issued against the request to the client computer either in remote environment or local network and updates custom entity instances as if data were updated by some clients [12].

III. Results

Rewriting the query such that indexes are used properly is an essential segment of a multiple DB extracting framework. In the examination, we present the thoughts of two new examples in various databases, viz. substantial affiliation governs and extraordinary affiliation runs the show. This paper exhibits a calculation for integrating three critical examples in different databases, viz. overwhelming affiliation rules, high regular affiliation rules, and outstanding affiliation rules It also gives accurate plan for resources utilization during multiple operations which are executed in a distributed data base environment and are associated with read and write request to determine the suitable cost model module.

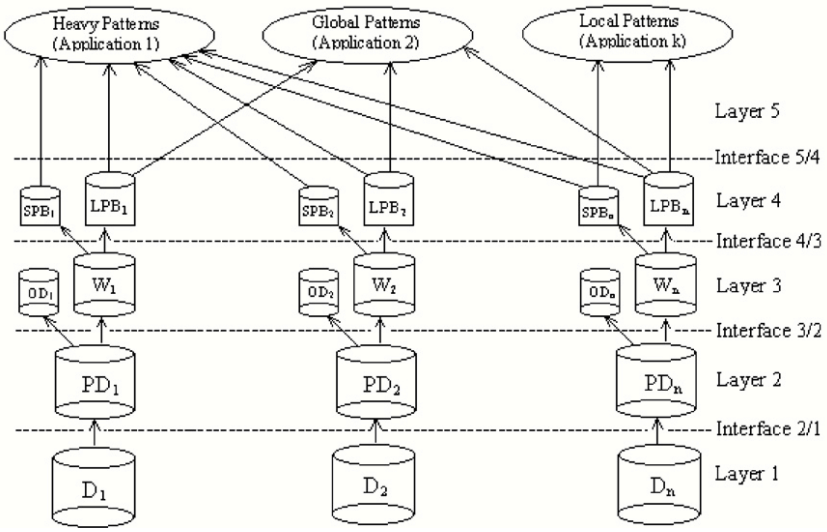

Fig. 3 Diagram of combining global DB from neighborhood designs in various DBS

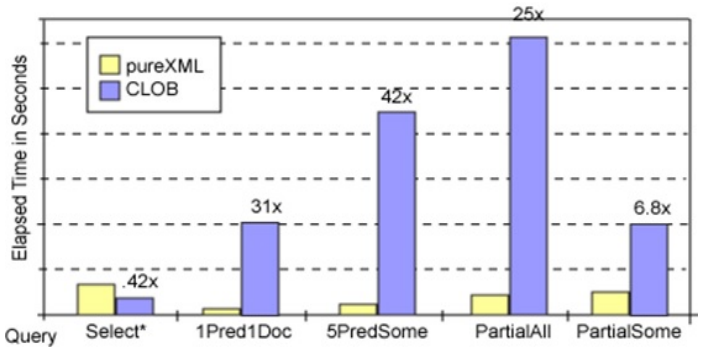

Fig. 4 Graph for Execution Time of Queries

In the distributed database, parallel execution plans of inter and intra operations are concurrently placed which has independently executed according to data fragmentation and data dependencies which are associated with multi-level operations across the parallel memory blocks [18].

A. Analysis of Execution Time ET

The main focusing point which was tuned during query optimization process is cost execution which includes I/O, CPU, memory management and assigning priority to different operations during run time. Accordingly, the ET increments with the expansion of multiple DBS are shown in Fig. 5 [16].

Numerous multi-branch organizations manage various databases. As the quantity of such organizations expands, we have to set ourselves up to create different applications in view of examples in numerous databases. There are two sorts of examples: neighborhood design and worldwide example. An example in light of a branch database is called neighborhood design. An example in view of all databases is called worldwide example. The most considerable factors during query run time response are filtering, matching, flexibility of resources and operations, syntax, parallel execution plans which fetch useful information from the database because of dynamic behavior structure of the algorithm.

B. Numeric Example with Simulation

We exhibit our procedures by an illustration and utilizing data sets of different virtual machines on network environment. We consider connectional connection T and its put away and inadequate datasets to be like those appearing in the following the figure. We created tuples of 2100 and 2150 as T and S respectively and incorporated 150 in SC. Those numbers were selected arbitrary for analyzing on different data bases on multiple number of times to determine the best execution time of query response time.

The query average time as shown in Fig. 7 was increased rapidly and in parallel when the dataset and number of records were increased while GPU has no tremendous effect. On the other hand, in relation with the query analysis, both CPU and GPU load increased in parallel as number of queries and record set increased [17].

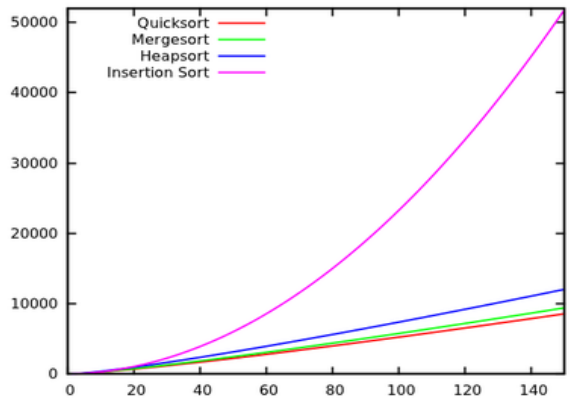

Fig. 5 Execution time versus number of databases from retail

IV. Conclusion

In the current age of technology, information is the most precious asset of a company. Today, companies have a large amount of data. As the data become larger, access to data for some particular information is becoming slower day by day. Faster data processing to shape it in the form of information is the biggest issue. Data clustering and data mining are the techniques which are used to process data fast and efficiently.

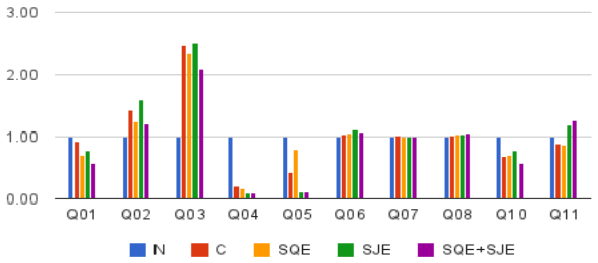

Fig. 6 Query rewrite versus recursive query evaluation

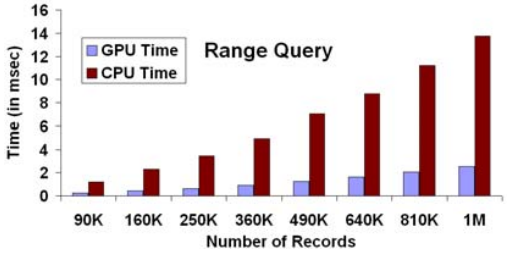

Fig. 7 Average of total

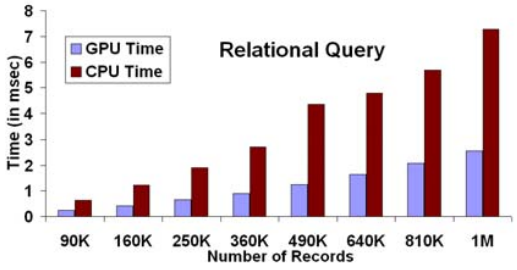

Fig. 8 Execution time of a Relational query

We portrayed two situations for outsourcing data accumulation benefits and exhibited an arrangement of decentralized distributed conventions for supporting data sharing over different private databases while limiting the data exposure among singular gatherings. Our work proceeds in a few bearings. To begin with, we are investigating elective security measurements, for example, data entropy-based measures for a broader assessment and examination of the protection attributes of the conventions. Second, we are investigating diverse topologies and other execution advancement systems for out-sourcing applications over the Internet. At long last, we are additionally inspired by researching the likelihood of building versatile conventions for outsourcing administrations crosswise over associations and nations with various protection prerequisites.

References

[1] A. Animesh, and P. R. Rao. "Synthesizing heavy association rules from different real data sources." *Pattern Recognition Letters* 29, 2008, no. 1,

pp. 59-71.
[2] A. Serge, B. Alexe et al., "An electronic patient record" on steroids": distributed, peer-to-peer, secure and privacy-conscious." In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, VLDB Endowment, 2004, pp. 1273-1276.
[3] B. Angela, "HePToX: marrying XML and heterogeneity in your P2P databases." In *Proceedings of the 31st international conference on Very large data bases*, VLDB Endowment, 2005, pp. 1267-1270.
[4] B. Ignacio, V. Hernández, and F. Mas. "A p2p platform for sharing radiological images and diagnoses." In *Proc DiDaMIC Workshop*. 2004.
[5] C. Mario, et al., "SIGMCC: A system for sharing meta patient records in a Peer-to-Peer environment." *Future Generation Computer Systems* 24, 2008, no. 3, pp. 222-234.
[6] F. Karen A, and K. Cochran, "Using data mining and data warehousing techniques." *Industrial Management & Data Systems* 99, 1999, no.5, pp.189-196.
[7] P. Karthik, G. T. Reddy and Kaari V "Tuning the SQL Query in order to Reduce Time Consumption" International Journal of Computer Science Issues, July 2012, Volume 9, 4(3), 418-423.
[8] J. Benjamin and J. Grimson, "Synapses/SynEx goes XML." *Studies in health technology and informatics*, 1999, pp. 906-911.
[9] M. Meghdad et al., "Sharing Clinical Information in P2P Environment with RBAC Mechanism." *IJCSNS International Journal of Computer Science and Network Security* 9, 2009, no. 1, pp.7-14.
[10] M. Gerome and D. Suciu,"Cryptographically Enforced Conditional Access for XML." In *WebDB*, 2002, pp. 101-106.
[11] P. Amir,"Managerial decision support with knowledge of accuracy and completeness of the relational aggregate functions." *Decision Support Systems* 42, 2006, no. 3, pp.1494-1502.
[12] S. Benjawan, and Dia L. Ali. "Multiple reads, one write simultaneously in distributed database." *Computers & industrial engineering* 31, 1996, no. 1-2, pp.319-322.
[13] T. M.A Lynda, "GRIDB: A scalable distributed database sharing system for grid environments." *organization* 431, no. 098, pp.26.
[14] Tavares, Carlos M., and Durval M. Tavares. "Real-time distributed data base locking manager." U.S. Patent, 1996, no. 7, pp.515-537.
[15] U. Manzoor and S. Nefti, "Decentralized trust management" In *Proceedings of the 17th IEEE Symposium on Security and Privacy*. Oakland, CA, 2009, pp.164–173.
[16] Z. Hongwei et al., "Reliable bursty convergecast in wireless sensor networks." *Computer Communications* 30, 2007, no. 13 pp. 2560-2576.
[17] N. K. Govindaraju, B. Lloyd, W. Wang, M. Lin, D. Manocha, "Fast Computation of Database Operations using Graphics processors. In: Proceedings of the 2004" *ACM SIGMOD International Conference on Management of Data, SIGMOD,* 2004, pp. 215-226.
[18] F. Priyatna, O. Corcho,´ J. Sequeda, "Formalisation and experiences of R2RMLbased SPARQL to SQL query translation using Morph" *In: 23rd International World Wide Web Conference,* 2014, Seoul, 7–11 April 2014, pp. 479–490.
[19] K. K. Hercule, M. M. Eugene, B. B. Paulin and Lilongo. B. J, "Study of the Master-Slave replication in a distributed database" International Journal of Computer Science Issues, September 2011, Volume 8, 5(3), 319-329.

Ahmed Mateen

Academic Background

- 2000/08 Graduated from Punjab University, Lahore, Pakistan. (Undergraduate)
- 2003/05 Graduated from The University of Lahore, Pakistan. (Master)
- 2007/04 Graduated from University of Agriculture Faisalabad, Pakistan. (M.Phil.)
- 2004/05 ~Present Lecturer in Computer Science Department at University of Agriculture Faisalabad, Pakistan
- 2017/09 ~ Doctoral Student in College of Computer Science, Chongqing University, China

Academic/ Professional awards (Govt./Institution/Society)

- Member of Core Committee (Diploma in Agribusiness by LUMS)
- System Operating Engineer in Direct rate of Distance Learning (UAF)
- Member of Senate (University of Agriculture Faisalabad
- Member of Departmental Board of Studies
- Member of Internet Society (USA)
- Member of Editorial Board of Journal of Computing and Applications
- Member of Turnitin Society (USA)
- Executive Member of Academic Staff Association (UAF 2012)
- Joint Sectary of Academic Staff Association 2010 UAF
- Chief Councilor ASA 2009 UAF
- Member of Computer Science Teacher Association (USA)
- Member of Social Science Research Network (United State)
- Association Computing Machinery (ACM) USA.
- Australian Computer Society Inc. (ACT)
- Member of Editorial Board of Journal of Computing and Applications. (JCA)
- Member of Editorial Board of Central Asian Journal of Information Technology (CAJIT)
- Member of Software Engineering Society UAF
- Member of the Seminar/Special problem Commatie CS Department

Academic Interest

Agriculture Modeling and Simulation, Wireless Networks, Real Time Systems, Distributed Cloud and Cluster Computing, Big Data, Software Engineering, Machine Learning.