



Privacy Preservation Enriched MapReduce for Hadoop Based BigData Applications

¹Chhaya S Dule, ²Dr. Girijamma H.A, ³Rajasekharaiah K M

¹Asst.Prof Dept. of Computer Science & Engineering

Jyothy Institute of Technology, Bangalore,

Visvesvaraya Technological University (VTU), Belgaum, Karnataka, India

²Professor, Department of Computer Science and Engineering,

RNS Institute of Technology, Bangalore,

Visvesvaraya Technological University (VTU), Belgaum, Karnataka, India

³Professor & HOD, Department of Computer Science and Engineering,

JnanaVikas Institute of Technology, Bangalore,

Visvesvaraya Technological University (VTU), Belgaum, Karnataka, India

Abstract: As per increase in the applications of various internet enabled services and cloud applications, the requirement of cloud infrastructure with enhanced facilities is increasing with very vast pace. Due to the increase in multiuser communication scenario on cloud infrastructure, the securities of datasets are also increasing drastically. Most of critical data on cloud is strictly required to be enriched with security and privacy preserved. Considering these requirements for huge data applications such as BigData, here in this paper an enhanced and optimized system called "Privacy preservation Enriched MapReduce framework for Hadoop based BigData applications" is proposed. In the proposed system four models to enrich overall anonymity of critical datasets has been developed. These models are privacy characterization model, anonymizer for datasets, dataset update and privacy preserved data management. In the proposed system the data owner possesses authority and interface to introduce various security levels for its data to make it privacy preserved and anonymous. The proposed model facilitates data users to retrieve datasets in its anonymized form which ultimately provides user task without publishing critical detail information about original data. This system would not only facilitate anonymity for datasets in cloud infrastructure but also optimize data recomputation by means of its partial data retaining capacity. Thus, the proposed system would bring optimization not only in terms of privacy preservation but also with enhanced resource utilization in BigData based applications.

Keywords: Privacy Preservation, MapReduce, Anonymizer, data management, Cloud Computing, Big Data

I. INTRODUCTION

In present day scenario, cloud has become an inevitable need for majority of Information Technology (IT) operational organizations. In general, the word cloud states for certain divergent IT infrastructure and environment which is specially developed for accomplishing the provision of remote access to the scalable and managed resources. Cloud computing is a technology where the data resources are shared rather than owing the personal server applications. The cloud infrastructure facilitates the allied users to access and employ the resources as per their requirement in real time applications. Hence, it can be stated that this cloud computing system facilitates users to have a very expedient and on-demand resource access. Cloud applications such as data storage, data retrieval and data portability have become some significant needs for IT organizations dealing with cloud computing. Considering the requirement, the IT development and user oriented global services can be globalized and delivered to single click by means of cloud applications such as BigData.

In present day scenario, there has been a great demand for optimum data access and resource utilization system with minimum latency and complexity and for all these, there must be a fare management optimization in processing units of Hadoop. Considering the organizational details of Hadoop systems, there are two predominant parts of Hadoop. The first component presents Hadoop Distributed File systems (HDFS) and second components states for MapReduce framework. The existing HDFS has been developed while keeping in mind the function of Google File System (GFS). In the present day scenario, the HDFS system classifies whole files into certain defined block sets which are further replicated into numerous other nodes with no specific monitoring whether the blocks are being divided evenly. Whenever there is the initialization of certain Job, the processor allied with individual node functions with its allied local disk. In HDFS system, the overall data is at first organized into certain files or directories, where the files are divided into certain blocks that are further

distributed across cluster nodes. The framework MapReduce has been popularized by open-source Hadoop model because of its efficiency in exploiting huge data sets.

The uniqueness of MapReduce framework is that this approach takes into consideration of handling of data collections across the multiple nodes and then returns back the single entity results or data sets. On the other hand this is the factor that ascertains the fault-tolerance which is visible to the developers or programmers. Thus, considering the architectural view of Hadoop for BigData applications, the performance enhancement can be obtained efficiently when focused on MapReduce optimization. In general a MapReduce application is comprised of three consecutive phases; mapping functions, data partitioning and reduce process. The first step called mapping operates on a series of key pairs and then after processing it generates output pair of key or value pairs. The generated pair (key, value) is then processed with reducer with the partitioning function. In this process the key and the complete number of reducers are taken as input which after processing generates out the index of the reducer to which the subsequent pair (key, value) must be transmitted for performing further operations. In later function the reduce function interact with allied values possessing unique key and then it generates the output. In Map phase the system introduces user-oriented logics to the input data while in reduce phase the intermediate results are processed and it results as the final reduced data in the form of key/value pair presentation. This is the matter of fact that the BigData infrastructure and its components encompass numerous data sets originating or associated with different origins or cloud sources. Since, the privacy susceptible data in mammoth BigData infrastructure can be easily retrieved because of its scattered or distributed nature, therefore in present day applications the requirement of privacy is realized predominantly. In specific need the privacy concerns are much aggravated in MapReduce of Hadoop framework for its application in BigData. This is the matter of fact the majority of such privacy issues are generic but while considering applications such as BigData these issues becomes very critical [1].

In order to accomplish the optimum privacy preservation in cloud computing frameworks, the approach of data anonymization can play a significant role. Then , in case of huge data in cloud infrastructure or BigData applications, in order to accomplish optimum function, scalability and efficiency, the system paradigms are moving towards MapReduce framework of Hadoop model. On the other hand, being a very sensitive part of Hadoop, MapReduce functions for various activities is cloud data processing and in such cases the exploration for the approaches to achieve optimum privacy preservation with higher efficiency for BigData application, there is still a long exploratory way to come with certain optimum solution. With the similar objective , in this paper, a robust technique for privacy preservation in MapReduce has been proposed that ensure to deliver a very flexible and scalable system model functional with dynamic data sets in BigData application scenario.

In this proposed system we have considered the development of privacy preservation model at the top of MapReduce in Hadoop framework, functional as an “anonymization sensitive filter” with a goal to ensure security and privacy of data sets. In this proposed system model, the developed framework at first process for privacy preservation before the preserved data is processed by MapReduce component of Hadoop model. On the other hand, in this proposed system model, the feasibility for multiple interfaces would be provided to the data owners so as to specify numerous privacy preservation requirements on the basis of varied privacy models. In this proposed system model, as soon as the data owner specifies the privacy requirements, the proposed model initiates certain anonymization paradigms of MapReduce components so as to ensure optimum anonymization of data sets for succeeding task of MapReduce framework. In order to accomplish the optimum Hadoop model, the anonymized data sets would be re-utilized so as to avoid unwanted re-computation of data sets. This feature would reduce the computational cost resulting into efficient system function. In such robust manner, the proposed system model would not only represent a privacy preservation scheme that can efficiently deals with dynamic data sets in cloud but it would also ensure maintaining of privacy needs of individual dynamic data sets. In this proposed research, in spite of privacy preservation oriented datasets anonymization a parallel scheme for encryption approach has been advocated that ensures a computational cost effective facility for multiple datasets which are anonymized for varied privacy needs individually. Ultimately, an integrated system model has been proposed that integrates developed privacy preservation model with cloud infrastructure for BigData applications. In this proposed system work, we propose to accomplish system’s functional analysis on the basis of real time data sets so as to evaluate the robustness of the proposed system.

Hadoop system is proposed to be developed on Microsoft Azure platform where it would be installed in HDInsight In this proposal the Virtual Machines (VMs) would be generated on CentOS. In this research proposal, the consideration of HD Insight has been emphasized because it encompasses numerous grids and for accessing these grids in real time applications it is required to have Hadoop system that is ultimately a cluster of Virtual Machines in cloud environment.

II. RELATED WORKS

Literature survey plays a significant role for building foundation for any research and development. Considering these requirements, a security orientated and privacy preservation for cloud infrastructure related survey has been conducted. Some of the predominant works done till are as follows:

Sriprasadh, K. et al [2] advocated the security concerns in cloud computing infrastructure and developed an approach for cloud security with multicast key facility for individual users where users were assigned its individual dynamical session key and it is maintained as per user movements. In other work Chang Liu et al [3] advocated the significance for authenticated Key Exchange (AKE) with optimized symmetric-key encryption for cloud security and developed hierarchical key exchange scheme Cloud Background Hierarchical Key Exchange (CBHKE), a two-phase multilayered iterative key exchange approach for data security. Qiwei Lu et al [4] advocated that due to data partitioning in cloud infrastructures the collaboration of privacy and security for distributed data is must. They explored a secure and practical outsourced collaborative data mining approach for cloud and designed a simple framework for security of data sets. Xin Dong et al [5] proposed an effective data collaboration scheme with security feature called SECO and employed a two-level hierarchical identity based encryption (HIBE) for ensuring the anonymity of data sets in untrusted cloud infrastructure. Initially they explored secure cloud data collaboration service that prevents unwanted leakage of data details and then facilitates a one-to-many encryption scheme with data writing and fine-grained access control concurrently. A light weight software based agent was advocated by El Ahrache, S.I. et al [6] which functions in parallel with the cloud database on each compassing virtual machines.

In other work Taeho Jung et al [7] presented certain scheme in which they facilitated a robust anonymous license control approach, AnonyControl for addressing data privacy. Rani, A. et al [8] developed Key Insertion and Splay Tree encryption approach that implements an asynchronous key series so as to perform data sets encryption. Ming Li et al [9] developed an approach that could deliver an efficient solution for search schemes for encrypted cloud storage infrastructures employing symmetric-key encryption techniques. Kumar, A. et al [10] developed system model to perform data storage effectively in the cloud infrastructure and presented system as a model to deal with huge data collections and ensured security using encryption schemes. In [11] a system model was developed based on encryption techniques functional on the basis of block cipher. Ermakova, T. et al [12] presented an important scheme for secure framework in cloud infrastructure especially designed for health care data sharing over cloud. The author designed a multi-provider cloud infrastructure that fulfills needs for higher resource availability, data privacy and its confidentiality in cloud infrastructure. They employed secret sharing for distributing various data records across cloud services that facilitates lower redundancy and supplementary security or privacy protection and performance was analyzed in terms of key compromise, inefficient encryption schemes and at last employed secret-sharing scheme with developed multi-cloud architecture. Alvarado et al [13] explored various techniques and analyzed existing security systems in cloud infrastructures for securing or protecting user data in cloud. Similarly, Ushdia, M. et al [14] proposed a scheme with data aggregation. In this scheme data stored is assured for its authenticated and leakage free storage and second need that system considers the provision of a multi-level granularity for the results obtained after aggregation. Powers, J. et al [15] envisioned a robust and efficient cloud-based data storage model which facilitates data owners to share their details in cost effective way even in the case of BigData applications and employed an iterative approach with Paillier Encryption and a random perturbation approach. Teo, et al [16] developed a protocol for estimating certain malicious party and semi-honest party in cloud infrastructure. In this research work the authors employed Multiplication (FSMP), Secure Scalar Product (SSP), and Secure Inverse of Matrix Sum (SIMS) paradigms. Jiawei Yuan et al [17] developed one system model for optimizing efficiency of cloud models by employing scheme for power use in cloud infrastructure. In the proposed model, the authors advocated that the individual data owner in cloud performs encryption for its associated private data on local basis and then uploads encrypted cipher texts into the cloud infrastructure. Then they developed an algorithm that can process majority of functions while pertaining to the learning paradigms over encrypted cipher texts without retrieving any information about the private data. In order to facilitate a flexible processing scenario over encrypted cipher texts the author has considered and tailored the BGN "doubly homomorphic" encryption scheme that exhibits multiparty setting. Samanthula, et al [18] emphasized on range query scheme and employed simple security primitive called secure comparison of encrypted integers.

This is the matter of fact that a number of researches have been done for ensuring security for data sets or critical information on cloud infrastructure, but still majority of researches suffers a lot of limitations. Majority of works do employ data encryption on cloud but it should be noted here that most of computations in cloud computing functions with unencrypted datasets. Even for BigData applications, major works are found to be confined. On the other hand, anonymity approaches and privacy preservation for Hadoop model, which is a key component for BigData applications, has not been considered dominantly. Very few works have been done for privacy preservation based MapReduce framework for Hadoop based BigData applications. Therefore, in this research proposal a highly robust and efficient privacy preservation model has been proposed that not only ensures anonymity of original datasets but also provides enhanced data utilization on cloud infrastructure.

III. PROPOSED SYSTEM/METHODOLOGIES

Considering the requirement of a highly robust and secured or anonymized MapReduce framework for Hadoop Model to be employed with BigData applications, here in this paper a novel technique called Privacy Preservation Enriched MapReduce framework for Hadoop based BigData applications has been proposed.

In order to preserve the security concerns associated with datasets in cloud infrastructure, here in this paper we proposed a MapReduce optimization technique called Privacy preserved MapReduce which functions in between the genuine datasets and user specified tasks for MapReduce framework. This user specified tasks are nothing else but the retrieval of certain data sensitive details or information from distributed cloud network. Tentative system architecture has been proposed as follows.

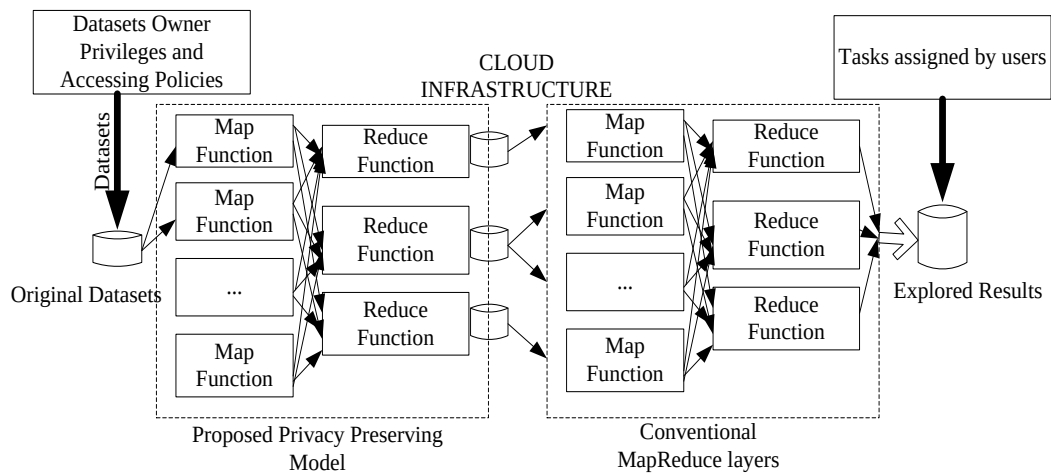


Figure: 1 Enhanced Privacy Preserved MapReduce framework for Hadoop based BigData applications

Considering the above mentioned architecture, it can be found that in the proposed model, the authenticated, critical and owner uploaded original data is uploaded with confidentiality and no data user would be able to retrieve the critical data of data owner. This proposal facilitates an interface for data owner to introduce its access policies to data users. The owner of data specifies certain policies for its privacy preservation. In spite of this in this paper, a scheme of data anonymization will be provided which functions as per privacy preserving policies it need. In order to minimize the recomputation of data sets in MapReduce function, some of the anonymized datasets are preserved. This facilitates the system to function as a data store management tool which can further update new data sets in cloud.

As per normal user function, the data users can specify their application needs as a job to the Hadoop and then through HDFS to MapReduce, and there runs their job on certain processed anonymized datasets. In the proposed system model, the privacy preservation scheme proposed, exhibits exploitation of jobs assigned to MapReduce framework for performing the required computations for data anonymization. This is the matter of fact that in BigData applications, there are huge data sets and even more complicated relationships among datasets; therefore it becomes significant to perform anonymized computation in MapReduce so as to preserve privacy of data sets on cloud storage.

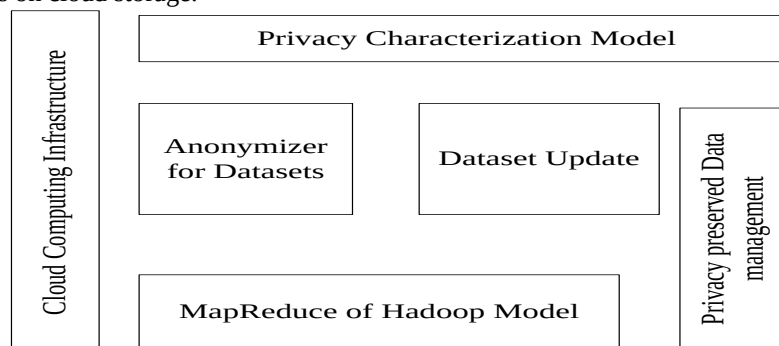


Figure: 2 proposed system sequential development and functional design

In order to accomplish the goal of privacy preserved model with enhanced system scalability and optimized performance for Hadoop framework in BigData applications here in this paper certain tentative models has been defined which can optimize overall performance of Hadoop model (especially for MapReduce) to ensure privacy preservation in cloud infrastructure. These tentative and defined components are:

1. *Privacy Characterization Model*
2. *Anonymizer for Datasets*
3. *Dataset Update*
4. *Privacy preserved Data management*

These defined system components can enhance overall system with four different issues, which can revolutionaries' to increase the efficiency of Hadoop or MapReduce framework for BigData applications. In proposed system model, the predominant components are Anonymizer for Datasets, dataset update and privacy preserved data management that function significant roles. In fact these proposed system modules would exhibit important function for ensuring privacy preservation as proposed for first module called privacy characterization scheme. Once the datasets have been processed with MapReduce, the updated data would be then processed for updates. In the last component of the proposed system, called privacy preserved data management; the emphasis would be made on ordering of the datasets so as to reduce the re-computation of datasets. In such way, this component would facilitate the utilization of cloud resources straightforwardly for achieving user specific task. The proposed research would facilitate enhancements for MapReduce component by introducing MapReduce framework at the top level of cloud that would ultimately achieve better system scalability and flexibility for real time applications, especially for BigData applications. A brief of the tentative system components to be developed has been discussed as follows:

A. PRIVACY CHARACTERIZATION MODEL

In this proposed research, it has been tried to accomplish the goal of a highly scalable and well mechanized cloud model enriched with privacy concerns while assuring datasets genuinity in BigData applications.

The first and important module is privacy characterization model that provide significant characterization of different requirement so as to facilitate an interface for data owner for introducing its required privacy issues. The need for security as characterized by data owner is characterized by certain factors, such as the privacy module. In this tentative system the privacy module being proposed could be a k-anonymity or t-closeness. The considered approach of k-anonymity based privacy preservation would facilitate multiple level of preservation. Being considered model the k-anonymity states that records or data sets being anonymized which are associated with certain quasi identifier must be higher than certain defined threshold parameter. In case of a specific scenario, when some of the quasi-identifiers becomes much specified that it is associated with very little users then in this case the associated users could be assigned authority for accessing higher confidential data and it might cause the breaching of security level. In the proposed system model, the quasi-identifiers state certain collections of anonymized datasets. On the basis of k-connectivity or t-closeness privacy model, in this tentative research paper, it has been tried to make quasi-identifiers more than the specified thresholds. The consideration of t-closeness can also ensure the distribution of certain critical datasets which are very close to the original datasets, associated with quasi-identifiers.

In the proposed model, the data owners would be facilitated with their own specification for data anonymization while exhibiting data classification and its clustering. Even without knowing the application types, the second component of the proposed model, Anonymizing of data can generate anonymized datasets for general uses expected from users. The characterization must be significant because, in case of our proposed BigData application, in fact all the original datasets are used directly these days, but for data anonymization only certain segments of data attributes are employed. In spite of this the data owner can also be provided a facility to assign abstraction for data usefulness or exposure to users. Abstraction based facility would enhance our proposed system with the flexibility of data anonymization.

B. ANONYMIZER FOR DATASETS

Considering an approach of anonymization, a well justified approach comes into mind and it is nothing else but the "Generalization". This approach states that a parent domain parameter or entity can be substituted with its descendant's or child domain entities so as to preserve privacy in its domain taxonomy tree. The data suppression in this approach can effectively substitute the original owner data with certain pre-specified symbol that would ultimately hide all the details of that specific data entity. This model would play a vital role in distinguishing the attribute parameters and certain critical parameters without causing any changes or manipulation in those parameters or attributes.

In the proposed system model, certain generalization approaches can be taken into consideration such as full-domain [19], sub-tree [20], multi-dimensional [21] and cell [22]. This is the matter of fact that these generalization schemes have exhibited better but when considering applications such as BigData, the generalization approach like multi-dimensional and cell generalization are forced to suffer a lot due to data exploration issues and inconsistency. Most of the existing schemes are insufficient for BigData applications because of data placements in varied computation nodes. Therefore in this research the MapReduce based mechanism is considered. In order to achieve better results it would be tried to implement Map and Reduce in iterative fashion, because of iterative anonymization in proposed modeling function. In the proposed system model, algorithms for privacy preserved MapReduce would be developed for Hadoop in BigData applications. It would play a vital role in assuring data anonymity.

C. DATASET UPDATE

In order to facilitate data access to various data users as per their application needs, it is needed to provide privileges in certain secure way and therefore in cloud infrastructure, it is always required to generate datasets and update anonymized data sets for users. Then while, the critical data sets on cloud infrastructure might vary over time and this is very common in BigData applications. Therefore, in this proposed system work, certain model would be developed for updating original datasets and anonymized data sets as per privacy concerns. Here, systems would be developed in such a way that it might anonymize the datasets of updated parts. In this phase varied anonymization layer would be employed for accomplishing privacy preservation and it would be tried to regulate the level of anonymization across the anonymized data sets. In this system we would try to bring optimum data utilization by users.

D. PRIVACY PRESERVED DATA MANAGEMENT

In order to optimize overall performance and computational efficiency a scheme for data retaining can be advocated for reutilizing the datasets. A data attribution scheme [23] can be considered for managing datasets. The regenerative nature of data attribution might facilitate regeneration of dataset from its antecedent datasets. It avoids regeneration of datasets. In the proposed model, the implementation of privacy characterization module may facilitate easy characterization of various privacy concerns and a dataset would be anonymized into various anonymous data. This would make system potential for recovering privacy sensitive details from numerous anonymous datasets. To achieve this issue for BigData applications, a system model can be advocated using encryption to ensure privacy preservation for all anonymous datasets that can be further shared among data users. Since, the encryption on all datasets can cause the huge overheads because of recurrent uses by data users, in this research, we would try to encrypt a definite part of datasets that might enhance privacy preservation with optimized computation cost. A paradigm [24] can be explored for enhancement by our proposed approach where certain selective data would be encrypted.

REFERENCES

- [1] Chaudhuri S (2012) What next?: A half-dozen data management research goals for big data and the cloud. In: Proceedings of the 31st symposium on principles of database systems, pp 1–4
- [2] Sriprasad, K.; Saicharansrinivasan; Pandithurai, O.; Saravanan, A., "A novel method to secure cloud computing through multicast key management," *Information Communication and Embedded Systems (ICICES)*, 2013 International Conference on 21-22 Feb. 2013, pp.305-311.
- [3] Chang Liu; Xuyun Zhang; Chengfei Liu; Yun Yang; Ranjan, R.; Georgakopoulos, D.; Jinjun Chen, "An Iterative Hierarchical Key Exchange Scheme for Secure Scheduling of Big Data Applications in Cloud Computing," *Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2013 12th IEEE International Conference on 16-18 July 2013, pp.9-16,
- [4] Qiwei Lu; Yan Xiong; Xudong Gong; Wenchao Huang, "Secure Collaborative Outsourced Data Mining with Multi-owner in Cloud Computing," *Trust, Security and Privacy in Computing and Communications (TrustCom)*, IEEE 11th International Conference on 25-27 June 2012, pp.100-108.
- [5] Xin Dong; Jiadi Yu; Yuan Luo; Yingying Chen; GuangtaoXue; Minglu Li, "Achieving secure and efficient data collaboration in cloud computing," *Quality of Service (IWQoS)*, 2013 IEEE/ACM 21st International Symposium on 3-4 June 2013, pp.1-6.
- [6] E Ahrache, S.I.; Ghodous, P.; Badir, H.; Sbihi, A., "Secure processing of large scale databases in a cloud computing environment," *Computer Systems and Applications (AICCSA)*, 2013 ACS International Conference on 27-30 May 2013, pp.1-4.
- [7] Taeho Jung; Xiang-Yang Li; Zhiguo Wan; Meng Wan, "Privacy preserving cloud data access with multi-authorities," *INFOCOM, 2013 Proceedings IEEE* 14-19 April 2013, pp.2625-2633
- [8] Rani, A.MercyGnana; Marimuthu, A., "Key Insertion and Splay Tree Encryption Algorithm for Secure Data Outsourcing in Cloud," *Computing and Communication Technologies (WCCCT)*, 2014 World Congress on Feb. 27 2014-March 1 2014, pp.92-97.
- [9] Ming Li; Shucheng Yu; KuiRen; Wenjing Lou; Hou, Y.T., "Toward privacy-assured and searchable cloud data storage services," *Network, IEEE* July-August 2013, vol.27 pp.56-62.
- [10] Kumar, A.; HoonJae Lee; Singh, R.P., "Efficient and secure Cloud storage for handling big data," *Information Science and Service Science and Data Mining (ISSDM)*, 2012 6th International Conference on New Trends in 23-25 Oct. 2012, pp.162-166.
- [11] Sugumaran, M.; Murugan, B.Bala; Kamalraj, D., "An Architecture for Data Security in Cloud Computing," *Computing and Communication Technologies (WCCCT)*, 2014 World Congress on Feb. 27 2014-March 1 2014, pp.252-255.
- [12] Ermakova, T.; Fabian, B., "Secret Sharing for Health Data in Multi-provider Clouds," *Business Informatics (CBI)*, 2013 IEEE 15th Conference on 15-18 July 2013, pp.93-100.
- [13] Alvarado, M.D.; Agrawal, R.; Baker, Y., "Security mechanisms utilized in a secured cloud infrastructure," *Southeastcon, 2013 Proceedings of IEEE* 4-7 April 2013, pp.1-5.
- [14] Ushdia, M.; Itoh, K.; Katayama, Y.; Kozakura, F.; Tsuda, H., "A Proposal of Privacy-Preserving Data Aggregation on the Cloud Computing," *Network-Based Information Systems (NBIS)*, 2013 16th International Conference on 4-6 Sept. 2013, pp.141-148.
- [15] Powers, J.; Chen, K., "Secure Computation of Top-K Eigenvectors for Shared Matrices in the Cloud," *Cloud Computing (CLOUD)*, 2013 IEEE Sixth International Conference on June 28 2013-July 3 2013, pp.155-162.
- [16] Teo, S.G.; Lee, V.; Shuguo Han, "A Study of Efficiency and Accuracy of Secure Multiparty Protocol in Privacy-Preserving Data Mining," *Advanced Information Networking and Applications Workshops (WAINA)*, 2012 26th International Conference on 26-29 March 2012, pp.85-90.
- [17] Jiawei Yuan; Shucheng Yu, "Privacy Preserving Back-Propagation Neural Network Learning Made Practical with Cloud Computing," *Parallel and Distributed Systems, IEEE Transactions on* Jan. 2014, vol.25, pp.212-221.
- [18] Samanthula, B.K.; Wei Jiang, "Efficient Privacy-Preserving Range Queries over Encrypted Data in Cloud Computing," *Cloud Computing (CLOUD)*, 2013 IEEE Sixth International Conference on June 28 2013-July 3 2013, pp.51-58.

- [19] LeFevre K, DeWitt DJ, Ramakrishnan R (2005) Incognito: efficient full-domain k-anonymity. In: Proceedings of 2005 ACM SIGMOD international conference on management of data (SIGMOD '05), pp 49–60
- [20] Fung BCM, Wang K, Yu PS (2007) Anonymizing classification data for privacy preservation. IEEE Trans Knowledge Data Eng 19(5):711–725
- [21] LeFevre K, DeWitt DJ, Ramakrishnan R (2006) Mondrian multidimensional k-anonymity. In: Proceedings of 22nd international conference on data engineering (ICDE '06), article 25
- [22] Xu J, Wang W, Pei J, Wang X, Shi B, Fu AWC (2006) Utility-based anonymization using local recoding. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data (KDD'06), pp 785–790
- [23] Davidson SB, Khanna S, Milo T, Panigrahi D, Roy S (2011) Provenance views for module privacy. In: Proceedings of the thirtieth ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems (PODS'11), pp 175–186
- [24] Zhang X, Liu C, Nepal S, Pandey S, Chen J (2012) A privacy leakage upper-bound constraint based approach for cost-effective privacy preserving of intermediate datasets in cloud. IEEE Trans Parallel Distrib System 24(4): 1192–1202

AUTHORS:

Presently **Ms. Chhaya S Dule** is working as Asst. Professor, Department of Computer Science & Engineering, Jyothy Institute of Technology, Bangalore. She has done B.E and M.Tech. in Computer Science & Engineering from reputed Universities, India .She is having 14+ years of teaching experience. She has research publications in reputed national and international journals. Her research area is security issues in cloud computing. Her other area of interests are DBMS, Software Engg., Software Testing ,Data Mining and Data warehouse, Neural Network Fuzzy logic. She has published 6 international research papers and 11 national papers in various conferences and journals. She has attended 4 faculty development programs organized by AICTE and TEQIP and 8 workshops on various topics



Presently **Dr. Girijamma H A** is working as Professor In The Department of Computer Science and Engineering, RNS Institute Of Technology, Bangalore .She has done B.E ,M.Tech. Aad Ph.D from reputed Universities. India. She is having total 18+ years of experiences (16+ years of Teaching experience and 2 years of Industrial Experience) and 9 years of research experience. She has published 3 international research papers .She is Life Member of MISTE and she is guiding research students under VTU, Belgaum.



Presently Mr. **Rajasekharaiah K.M.** is working as **Professor & HOD** Department of Computer Science & Engineering, Jnana Vikas Institute of Technology, Bangalore. He has done M.Tech. in Computer Science & Engg. M.Sc. Information Technology, M.Phil. in Computer Science, and PGDIT from reputed Universities, India. He is having 20+ years of total experience including 16 years of Industrial experiences. He is a Life fellow Member of Indian Society for Technical Education (ISTE), New Delhi. He is presently pursuing the doctoral degree in the Branch of Computer Science & Engineering, in the domain area of Data Mining & Warehousing. He has research publications in reputed national and international journals. His other area of interests are DBMS, Software Engg., Software Architecture, Computer Networks, Programming Languages, Data Structures and Mobile Computing. He is also a resource scholar for other Engineering Colleges/University.

