

**ATTAINING OF PRIVACY PRESERVING FOR LARGE-SCALE DATA  
ON CLOUD SYSTEM****K.Srilata<sup>1</sup>, Ch.Venkatesh<sup>2</sup>**<sup>1</sup>M.Tech Student, Dept of CSE, Arjun College of Technology & Sciences, Hyderabad, T.S, India<sup>2</sup>Assistant Professor, Dept of CSE, Arjun College of Technology & Sciences, Hyderabad, T.S, India**ABSTRACT:**

Several researchers researched on the problem of scalability regarding extensive data anonymization and a variety of algorithms with several anonymization operations were projected. Data anonymization was to a great extent considered for managing of data privacy in non-interactive data publishing as well as sharing situation. We put forward an exceptionally well-organized two-phase top-down specialization method for anonymization of data on the basis of Map Reduce on cloud. In our research, we manage Map Reduce, which is an extensive adopted framework of parallel data processing, to tackle scalability difficulty of top-down specialization method for extensive data anonymization. Top-down specialization technique, offers an advanced trade-off among data utility as well as data consistency, which is extensively functional for data anonymization.

***Keywords: Data anonymization, Two-phase top-down specialization, MapReduce, Data privacy, Data consistency.***

**1. INTRODUCTION:**

Data sets were turned out to be outsized so that anonymization of such huge data sets

has become a huge demand for conventional algorithms. Extent of data sets that require anonymizing in several applications of cloud

system enhance extremely in conformity with cloud computing as well as trends of Big Data. Extensive data frameworks of processing such as MapReduce were integrated with cloud for provision of dominant computation potential for applications [1]. As a result, it is capable to adopt such frameworks to deal with scalability difficulty of anonymizing important data for preserving of privacy. Data privacy can be discloses by fewer attempt by malevolent cloud users due to failures of several conventional privacy protection ways on cloud that brings severe social reputation destruction to data owners. In our work we suggest an extremely efficient two-phase top-down specialization method for anonymization of data on the basis of MapReduce on cloud. Top-down specialization method, offers a superior trade-off among data utility as well as data consistency, which is extensively functional for data anonymization. In our research, we control MapReduce, which is an extensive adopted framework of parallel data processing, to tackle scalability difficulty of top-down specialization method for extensive data anonymization.

## 2. METHODOLOGY:

Technology of cloud computing make available immense computation power as well as storage capability by means of utilization of huge number of commodity computers mutually, facilitating users to organize applications inexpensively devoid of heavy investment. Issues regarding data privacy have to be required earlier than data sets are analyzed on cloud [2][3]. Various distributed algorithms are set up for preserving of privacy regarding of numerous data sets that are retained by numerous parties. Our work mostly focus on scalability issue of top-down specialization method anonymization, and make use of MapReduce to anonymize extensive data sets earlier than data are processed by various other MapReduce jobs, that arrive at privacy preservation. Top-down specialization technique is an iterative procedure that starts from topmost domain values in taxonomy trees of attributes and each round of iteration includes three most important steps such as finding best possible specialization, performing of specialization process and updating of the values of search metric for subsequent round. We initiate two-phase top-down specialization technique for anonymization of data on basis of Map

Reduce on cloud to carry out computation necessary in top-down specialization in an extremely scalable as well as efficient approach. To make complete usage of parallel capability of MapReduce on cloud, specializations necessary in the process of anonymization are split into two phases. The two phases of our approach are on the basis of two levels of parallelization that are provisioned by MapReduce on cloud. In first phase, actual data sets are divided as a group of minute data sets, and these are anonymized in parallel, constructing intermediate results and in second one, intermediate results are included to one, and later anonymized to attain consistent k-anonymous datasets. In both phases of proposed approach, we design a group of pioneering MapReduce jobs to achieve specialization computation within an extremely scalable means. We control MapReduce to achieve concrete computation in both phases. MapReduce jobs are purposely considered and synchronized to carry out specialization on data sets collaboratively.

### **3. AN OVERVIEW OF PROPOSED SYSTEM:**

Cloud users decrease vast investment of infrastructure, and focus on their individual core business on the other hand; several possible customers are uncertain to take benefit of cloud because of privacy as well as security concerns. Data anonymization signifies hiding of identity or sensitive data for data record owners. In the recent times, data in numerous cloud applications increases extremely in conformity with Big Data, thus making it a challenge for software tools to manage, and practice such extensive data within a satisfactory elapsed time. Therefore it is a challenge for traditional methods of anonymization to attain privacy preservation on privacy-sensitive extensive data sets because of their shortage of scalability. We put forward an extremely efficient two-phase top-down specialization method for anonymization of data on the basis of Map Reduce on cloud. Top-down specialization method, offers a superior trade-off among data utility as well as data consistency, which is extensively functional for data anonymization [4]. We control MapReduce, which is an extensive adopted framework of parallel data processing, to tackle scalability difficulty of top-down

specialization method for extensive data anonymization. The majority of algorithms of top-down specialization are centralized; results in failure of handling extensive data sets. While some of distributed algorithms were projected that mainly spotlights on secure anonymization of data sets from numerous parties, to a certain extent than scalability issue. Since MapReduce computation concept is comparatively easy, it is a challenge for designing of appropriate Map Reduce jobs for top-down specialization. Top-down specialization method, offers a superior trade-off among data utility as well as data consistency, which is extensively functional for data anonymization. In the top-down specialization method, a data set is anonymized by means of performing specialization process. Top-down specialization method utilizes anonymization level to manage specialization procedure. Top-down specialization method is an iterative procedure that starts from topmost domain values in taxonomy trees of attributes and each round of iteration includes three most important steps such as finding best possible specialization, performing of specialization process and updating of the values of search

metric for subsequent round. We introduce two-phase top-down specialization technique for anonymization of data on basis of Map Reduce on cloud to carry out computation necessary in -down specialization in an extremely scalable as well as efficient approach [5]. The essential thought of phase top-down specialization technique is to achieve high scalability by means of making a trade-off among scalability as well as data utility. The two phases of our approach are on the basis of two levels of parallelization that are provisioned by MapReduce on cloud. In first phase, actual data sets are divided as a group of minute data sets, and these are anonymized in parallel, constructing intermediate results and in second one, intermediate results are included to one, and later anonymized to attain consistent k-anonymous datasets. To accomplish high scalability, we parallelize numerous jobs on data partitions in first phase, but resulting anonymization levels are not the same [6]. To get hold of finally constant anonymous data sets, second phase is essential to put together intermediate results and additionally anonymize complete data sets.

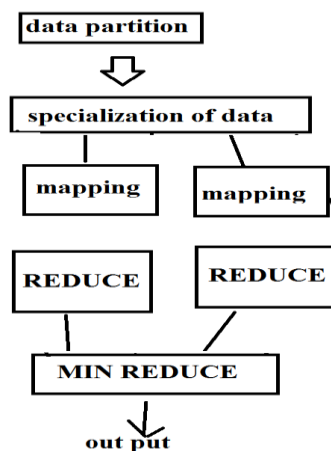


Fig1: An overview of Map Reduce structure

#### 4. CONCLUSION:

One of the vital issues regarding the technology of cloud computing is privacy and concern aggravates in circumstance of cloud computing even though some privacy issues are not innovative. Our work mainly spotlight on scalability issue of top-down specialization method anonymization, and make use of MapReduce to anonymize extensive data sets earlier than data are processed by various other MapReduce jobs, that arrive at privacy preservation. Top-down specialization technique makes use of anonymization level to manage specialization procedure. To make complete use of parallel capability of MapReduce on cloud, specializations necessary in the process of anonymization are split into two phases. In initial phase, actual data sets are

separated as a group of minute data sets, and these are anonymized in parallel, constructing intermediate results and in second one, intermediate results are integrated to one, and later anonymized to attain consistent k-anonymous datasets. In both phases of projected approach, we intend a group of pioneering MapReduce jobs to achieve specialization computation within an extremely scalable means. The essential consideration of phase top-down specialization technique is to achieve high scalability by means of making a trade-off among scalability as well as data utility.

#### REFERENCES

- [1] B. Fung, K. Wang, L. Wang, and P.C.K. Hung, "Privacy- Preserving Data Publishing for Cluster Analysis," *Data and Knowledge Eng.*, vol. 68, no. 6, pp. 552-575, 2009.
- [2] N. Mohammed, B.C. Fung, and M. Debbabi, "Anonymity Meets Game Theory: Secure Data Integration with Malicious Participants," *VLDB J.*, vol. 20, no. 4, pp. 567-588, 2011.
- [3] L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," *Int'l J. Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557-570, 2002.
- [4] K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan, "Sedic: Privacy-Aware Data Intensive

Computing on Hybrid Clouds,” Proc. 18th ACM Conf. Computer and Comm. Security (CCS ’11), pp. 515-526, 2011.

[5] X. Xiao and Y. Tao, “Personalized Privacy Preservation,” Proc. ACM SIGMOD Int’l Conf. Management of Data (SIGMOD ’06), pp. 229-240, 2006.

[6] J. Ekanayake, H. Li, B. Zhang, T. Gunarathne, S.-H. Bae, J. Qiu, and G. Fox, “Twister: A Runtime for Iterative Mapreduce,” Proc. 19th ACM Int’l Symp. High Performance Distributed Computing (HDPC ’10), pp. 810-818, 2010.



**K.Srilata**, Graduated in B.Tech,CSE,Sree Visvesvaraya Inst. of Tech. & Science,Mahabubnagar Dist.



**Ch.Venkatesh** Graduated in B.Tech CSE in 2011 from Arjun College of Technology & Sciences,R.R. Dist . He received Masters Degree in M.Tech [CSE] Arjun College of Technology & Sciences,R.R. Dist. Presently he is working as Assistant Professor in CSE Dept. in Arjun College of Technology & Sciences, Hayathnagar,R.R. Dist Telangana State, India