ISSN 2348 – 9928
IJAICT Volume 5, Issue 7, July 2018 Doi:01.0401/ijaict.2015.11.20 Published on 05 (7) 2018

# DATA ANONYMITY USING BOTTOM UP APPROACH AND MAPREDUCE ON CLOUD

***Ms. S. Vijayalakshmi***
***Department of Information Technology***
***Christ College of Engineering and Technology***
***Puducherry, India***

***Ms. S. Jayasri, Ms. J. Surekha, Ms. P. Thilagavathi***
***Department of Information Technology***
***Christ College of Engineering and Technology***
***Puducherry, India***

***Abstract— At present, the size of information in numerous cloud applications increments massively as per the Big Data pattern, in this way making it a test for usually utilized programming instruments to catch, oversee and process such expansive scale information inside a mediocre slipped by time. In huge information applications, information protection is a standout amongst the most concerned issues on the grounds that handling vast scale security delicate information sets frequently obliges processing force gave by open cloud administrations. Accordingly is challenge for existing anonymity ways to attain to protection conservation on security delicate huge scale information sets because of their inadequacy of versatility. In this paper we propose an adaptable Propelled Bottom up speculation approach for inform anonymity in light of Map Reduce on cloud.***

***Keywords—Data Anonymity, Bottom-up Generalization, Map-Reduce, cloud***

## I. INTRODUCTION

Distributed computing gives gigantic reckoning force and capacity limit through using countless PCs together, empowering clients to send applications cost-viably without overwhelming framework speculation. Cloud clients can lessen tremendous forthright speculation of IT foundation, and focus all alone center business. Be that as it may, various potential clients are still reluctant to exploit cloud because of security and security concerns. A very versatile two-stage TDS Approach for information anonymity in light of Map Reduce on cloud. To make full utilization of the parallel capacity of Map Reduce on cloud, specializations needed in an anonymity methodology are part into two stages. In the first, unique information sets are parceled into a gathering of littler information sets, and these information sets are anonymity in parallel, creating transitional results. In the second one, the moderate results are incorporated into one, and further anonymity to attain to predictable k-unknown information
sets. The proposed paper we are going to embrace is the adaptable protection safeguarding mind full examination and planning on expansive scale information sets. We are right now utilizing the methodology or the system named base up procedure. Upgraded adjusted planning techniques are required to be produced towards general versatile protection conservation mind full information set booking. The significant contributions of this paper are as per the following.

The protection safeguarding for information investigation, impart and mining is a testing examination issue. Centralized methodologies presumably experience the ill effects of low effectiveness and versatility when taking care of substantial scale information sets, so utilizing virtual machines have the most elevated calculation and capacity. Cloud computing is internet-based computing in which large groups of remote servers are networked to allow sharing of data-processing tasks, centralized data storage, and online access to computer services or resources. It is the computational procedure of finding examples in vast information sets including strategies at the convergence of manufactured intelligence, machine learning, statistics and database framework. Information mining is the procedure of dissecting information from alternate points of view and outlining it into helpful data. It permits clients to investigate information from a wide range of measurements or edges, arrange it, and outline the connections recognized.

## II. EASE OF USE

### *2.1 A Scalable Two-Phase Top-Down Specialization Approach for Data Anonymization Using Map Reduce on Cloud*

In the Top down Specialization (TDS) [1], is an iterative process starting from the topmost domain values in taxonomy trees of attributes. Each round of iteration consists of three main steps, namely, finding the best specialization, performing specialization and updating values of the search metric for next round. Such process is repeated until k-anonymity is violated, to expose the maximum data utility. The goodness of specialization is measured by a search matrix.

### *2.2 Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data.*

Related works on search able encryption focus on single keyword search or Boolean keyword search, and rarely sort the search results [2]. In this paper, for the first time, we define and solve the challenging problem of privacy preserving multi-keyword ranked search over encrypted cloud data (MRSE).

***2.3 Data Anonymization Approaches for Data Sets Using Map Reduce on Cloud.***

In this information age, huge amounts of data are collected and mined every day. The process of data publication is becoming larger and complex day by day. Cloud computing is the most popular model for supporting large and complex data, most organizations are moving towards to reduce their cost and elasticity features [3]. However cloud computing has potential risk and vulnerabilities. One of major problem in moving to cloud computing is its security and privacy concerns. Cloud computing provides powerful and economical infrastructural resources for cloud users to handle ever increasing data sets in big data applications.

***2.4 Security and Privacy Issues in Cloud Computing***

Cloud computing transforms the way information technology (IT) [4] is consumed and managed, promising improved cost efficiencies, accelerated innovation, faster time-to-market, and the ability to scale applications on demand. According to Gartner, while the hype grew exponentially during 2008 and continued since, it is clear that there is a major shift towards the cloud computing model and that the benefits may be substantial. However, as the shape of the cloud computing is emerging and developing rapidly both conceptually and in reality, the legal/contractual, economic, service quality, interoperability, security and privacy issues still pose significant challenges.

***2.5 Security and Privacy for Map Reduce***

Airavat enables the execution of trusted and untrusted Map Reduce computations on sensitive data, while assuring comprehensive enforcement of data providers privacy policies .We present Airavat, a Map Reduce-based system which provides strong security and privacy guarantees for distributed computations on sensitive data. Airavat is a novel integration of mandatory access control and differential privacy [5]. Data providers control the security policy for their sensitive data, including a mathematical bound on potential privacy violations.

## III. PREPARE YOUR PAPER BEFORE STYLING

The current framework comprises of a versatile two-stage top-down specialization (TDS) way to anonymity huge scale information sets utilizing the Map Reduce system on cloud [6].In the first, unique information sets are divided into a gathering of littler information sets, and these information sets are anonymity in parallel, creating moderate results.

In the second one, the moderate results are incorporated into one, and further anonymity to accomplish reliable k-mysterious information sets.

In the Top down Specialization (TDS), is an iterative procedure beginning from the highest area values in scientific categorization trees of characteristics. Each round of cycle comprises of three primary steps, to be specific, discovering the best specialization, performing specialization and upgrading estimations of the quest metric for next round. Such process is rehashed until k-secrecy is violated, to uncover the greatest information utility.

### *3.1 Two-phase top-down specialization*

Three parts of the TPTDS approach, in particular, information allotment, anonymity level fusing, and information specialization. We propose a TPTDS way to lead the reckoning needed in TDS in an exceptionally adaptable and productive design. The two periods of our methodology are in light of the two levels of parallelization provisioned by Map Reduce on cloud.

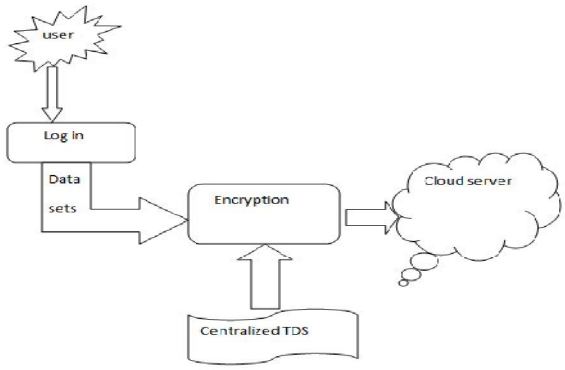


***Fig 1: TDS- working process***

### *3.2 Map Reduce version of centralized TDS*

MRTDS plays a core role in the two-phase TDS approach, as it is invoked in both phases to concretely conduct computation. Basically, a practical Map Reduce program consists of Map and Reduce functions, and a Driver that coordinates the macro execution of jobs.

### *3.3 MRTDS Driver*

Usually, a single Map Reduce job is inadequate to accomplish a complex task in many applications. Thus, a group of Map Reduce jobs are orchestrated in a driver program to achieve such an objective. MRTDS consists of MRTDS Driver and two types of jobs, i.e., IGPL Initialization and IGPL Update.

## IV. USING THE TEMPLATE

The proposed paper we are going to adopt is the scalable privacy preservation aware analysis and scheduling on large-scale data sets. Optimized balanced scheduling strategies are expected to be

developed towards overall scalable privacy preservation aware data set scheduling. We are currently using the approach or the technique named bottom up process.

An adaptable Advanced Bottom up speculation approach for information anonymity in light of Map Reduce on cloud. Unique datasets are part up into a gathering of littler datasets, and these datasets are anonymity in parallel, creating moderate results. At that point, the middle of the road results is consolidated into one, and further anonymity to accomplish steady k-unknown information sets.

Anonymity is not invulnerable counter measures that compromise current anonymity techniques can expose protected information in released datasets.

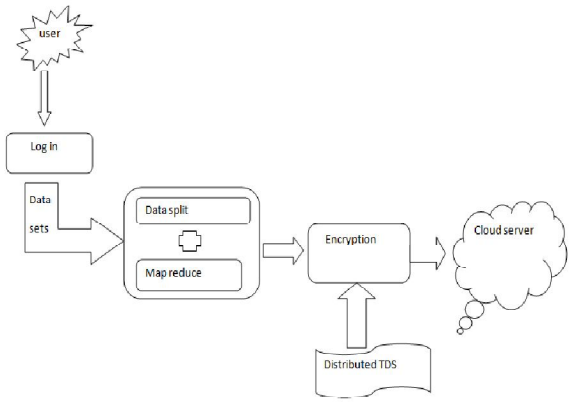


***Fig 2: BUG- working process***

## V. SIGNIFICANCE OF PROPOSED SYSTEM

i. The privacy preservation for data analysis, share and mining is a challenging research issue [7].
ii. Centralized approaches probably suffer from low efficiency and scalability when handling large-scale data sets.
iii. Virtual machines have the highest computation and storage capability.

## VI. MODULES

### *6.1 User Interface*

To connect with server user must give their username and password then only they can able to connect the server. If the user already exits directly can login into the server else user must register their details such as username, password and Email id, into the server. Server will create the account for the entire user to maintain upload and download rate.

### *6.2 File Uploading*

In this module, the data is given by customer requests go to server where administrator maintains all files and responsible for storing that files into cloud. File was uploaded using http request method, for each file unique key generated.

### *6.3 Key Generation*

In this model we take user uploaded file, that file is split into multiple parts using mapping technique. Each and every part was assigned unique id .that key used for privacy preserving of files. This kind of data encryption is called data anonymity [8].

### *6.4 Splitting Of File*

Each file was split into multiple parts. Those parts are mapped using mapping technique. Splitting of file part size is equal and depends upon file size.

### *6.5 File Storing using Map-reduce*

Map Reduce program is composed of a Map procedure that performs filtering and sorting. Reduce procedure that performs storing of data into database. The key contributions of the Map Reduce framework are not the actual map and reduce functions [9], but the scalability and fault-tolerance achieved for a variety of applications by optimizing the execution engine once.

## VII. CONCLUSION

In cloud environment, the privacy preservation for data analysis, share and mining is a challenging research issue due to increasingly larger volumes of data sets, thereby requiring intensive investigation. We investigated the adoption of our approach to the bottom-up generalization algorithms for data anonymity. Demonstrate that our approach can significantly improve the privacy preservation, scalability and efficiency in large data set on cloud.

## References

[1] Xuyun Zhang, Laurence T. Yang, Senior Member, IEEE, Chang Liu, and Jinjun Chen,Member, IEEE A Scalable Two-Phase Top-Down Specialization Approach for Data Anonymization Using MapReduce on Cloud 2014.

[2] Ning Cao†, Cong Wang‡, Ming Li†, Kui Ren‡, and Wenjing Lou†Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data 2011.

[3] N. Mohammed, B. Fung, P.C.K. Hung, and C.K. Lee, Centralized and Distributed Anonymization for High-Dimensional Healthcare Data," ACM Trans. Knowledge Discovery from Data, vol. 4, no. 4, Article 18, 2010.

[4] B.C.M. Fung, K. Wang, and P.S. Yu, "Anonymizing Classification Data for Privacy Preservation," IEEE Trans. Knowledge and Data Eng., vol. 19, no. 5, pp. 711-725, May 2010.

[5] Jaydip Sen Innovation Labs, Tata Consultancy Services Ltd., Kolkata, INDIA Security and Privacy Issues in Cloud Computing2012.

[6] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A View of Cloud Computing," Comm. ACM, vol. 53, no. 4, pp. 50-58, 2010.

[7] H. Takabi, J.B.D. Joshi, and G. Ahn, "Security and Privacy Challenges in Cloud Computing Environments," IEEE Security and Privacy, vol. 8, no. 6, pp. 24-31, Nov. 2010

[8] S. Chaudhuri, "What Next?: A Half-Dozen Data Management Research Goals for Big Data and the Cloud," Proc. 31st Symp. Principles of Database Systems (PODS '12), pp. 1-4, 2012.

[9] Indrajit Roy Srinath T.V. Setty Ann Kilzer Vitaly Shmatikov Emmett Witchel Airavat: Security and Privacy for MapReduce.

[10] B.C.M. Fung, K. Wang, R. Chen, and P.S. Yu, "Privacy-Preserving Data Publishing: A Survey of Recent Devel- opments," ACM Computing Surveys, vol. 42, no. 4, pp. 1-53, 2010.