IJRRECS/June 2015/Volume-3/Issue-6/4166-4170 **ISSN 2321-5461**

INTERNATIONAL JOURNAL OF REVIEWS ON RECENT ELECTRONICS AND COMPUTER SCIENCE

MANAGING OF PRIVACY PRESERVATION FOR EXTENSIVE DATA SETS

P.Nipuna Reddy[1], A.Radha Rani[2]

[1]M.Tech Student, Dept of CSE, Malla Reddy Engineering College for Women, Hyderabad, T.S, India

[2]Associate Professor, Dept of CSE, Malla Reddy Engineering College for Women, Hyderabad, T.S, India

ABSTRACT:

In cloud computing, an issue of privacy was considered as serious issue. Quite a lot of distributed algorithms are projected to defend confidentiality of numerous data sets which are retained by numerous parties. Data sets have turn out to be huge such that anonymizing of these data sets has turned out to be a substantial challenge for established anonymization algorithms. Our research mostly spotlight on scalability concern of top-down specialization anonymization, and is, consequently, orthogonal and harmonizing to them. Our research makes the most of MapRedue to anonymize extensive data sets earlier than data are additionally processed by other MapReduce jobs, which arrives at preservation of privacy. In our work, we put forward an extremely scalable two-phase top-down specialization method for data anonymization on basis of MapReduce on cloud. In our research, we control MapReduce, which is an extensively adopted framework of parallel data processing, to tackle scalability difficulty of top-down specialization (TDS) strategy for extensive data anonymization. Since MapReduce computation concept is comparatively effortless, it is still a challenge to plan appropriate MapReduce jobs in support of top-down specialization. The top-down specialization, offers a good trade-off among data utility as well as data reliability, is extensively functional for data anonymization. Combined by means of cloud, MapReduce turn out to be more influential as well as flexible since cloud can recommend infrastructure resources on demand.

Keywords: Cloud computing, Data anonymization, MapReduce, Top-down specialization.

1. INTRODUCTION:

Considerable computation power as well as storage capacity was provided by cloud computing using huge number of commodity computers mutually and permitting users towards organizing of applications inexpensively devoid of assets of heavy infrastructure [1]. Anonymization of data has been expansively adopted for preserving of data privacy in non-interactive data publishing as well as situations of sharing. Data anonymization refers towards concealing of sensitive data for data record owners. Several algorithms concerning anonymization with numerous operations of anonymization were projected in existing works. The extent of data sets that require anonymizing in several cloud applications augments extremely in conformity with cloud computing. Extensive data processing frameworks such as MapReduce was incorporated with cloud to make available controlling computation ability for applications thus, it is promising to accept such frameworks to tackle scalability difficulty of anonymizing extensive data for preserving of privacy. Our research makes the most of MapRedue to anonymize extensive data sets earlier than data are additionally processed by other MapReduce jobs, which arrives at preservation of privacy [2][3]. In our work, we put forward an extremely scalable two-phase top-down specialization method for data anonymization on basis of MapReduce on cloud. To make complete use of equivalent ability of MapReduce on cloud, specializations necessary in an anonymization procedure are divided into two phases such as: initially original data sets are divided into a group of minor data sets, which are anonymized in equivalent, producing intermediary results. In second one, intermediary results are included into one, and additionally anonymized to attain consistent k-anonymous datasets. The two phases of our introduced approach are on basis of two levels of parallelization which are provisioned by MapReduce on cloud.

2. METHODOLOGY:

Cloud users can decrease vast upfront savings of IT infrastructure, and focus on own core business. In cloud computing, an issue of privacy was considered as serious issue. Data privacy is revealed with lesser effort by malevolent cloud users because of failures of several established measures of privacy protection on cloud. Issues of data privacy should be addressed without delay

earlier than data sets are shared on cloud. Quite a lot of distributed algorithms are projected to defend confidentiality of numerous data sets which are retained by numerous parties. Our research mostly spotlight on scalability concern of top-down specialization anonymization, and is, consequently, orthogonal and harmonizing to them. Data sets have turn out to be huge such that anonymizing of these data sets has turned out to be a substantial challenge for established anonymization algorithms. The researchers have begun towards examining of scalability difficulty of extensive data anonymization. In our research, we control MapReduce, which is an extensively adopted framework of parallel data processing, to tackle scalability difficulty of top-down specialization (TDS) strategy for extensive data anonymization. Even though several distributed algorithms were projected they mostly spotlight on protected anonymization of data sets from numerous parties, to a certain extent than the scalability features. Since MapReduce computation concept is comparatively effortless, it is still a challenge to plan appropriate MapReduce jobs in support of top-down specialization. The top-down specialization, offers a good trade-off among data utility as well as data reliability, is extensively functional for data anonymization. For the most part of top-down specialization algorithms are centralized, outcomes in their failure in handling extensive data sets [4]. Our research makes the most of MapRedue to anonymize extensive data sets earlier than data are additionally processed by other MapReduce jobs, which arrives at preservation of privacy.

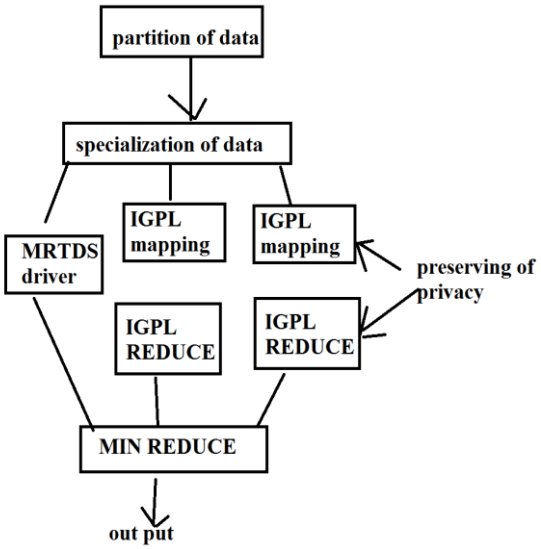

Fig1: Dataflow in Map Reduce structure

3: AN OVERVIEW OF PROPOSED APPROACH:

In recent times, data privacy preservation of data privacy was broadly investigated. We control MapReduce, which is an extensively adopted framework of parallel data processing, to tackle scalability difficulty of

top-down specialization (TDS) strategy for extensive data anonymization. An overview of data flow in Map Reduce structure was shown in fig1. In our work, we put forward an extremely scalable two-phase top-down specialization method for data anonymization on basis of MapReduce on cloud and consist of three components of such as data separation, anonymization level integration, as well as data specialization. Specializations necessary in an anonymization procedure are divided into two phases such as: initially original data sets are divided into a group of minor data sets, which are anonymized in equivalent, producing intermediary results. In second one, intermediary results are included into one, and additionally anonymized to attain consistent k-anonymous datasets. The two phases of our introduced approach are on basis of two levels of parallelization which are provisioned by MapReduce on cloud. Combined by means of cloud, MapReduce turn out to be more influential as well as flexible since cloud can recommend infrastructure resources on demand [5]. The fundamental idea of two phases of our introduced approach is to put on high scalability by building a trade-off among scalability as well as utility of data and ensure the measure of data privacy preservation, since two phases of our introduced approach produces k-anonymous data sets ultimately. MapReduce on cloud contains two levels of parallelization, specifically job level as well as task level. To attain high scalability, numerous jobs were parallelized on data partitions in initial phase; however resulting anonymization levels are not matching. To get hold of ultimately constant anonymous data sets, second phase is essential to put together intermediary results and additionally anonymize complete data sets. Subroutine was run over each of partitioned data sets in analogous to build complete use of the job level parallelization concerning MapReduce [6].

4. CONCLUSION:

Anonymization of data has been expansively adopted for preserving of data privacy in non-interactive data publishing as well as situations of sharing. Several algorithms concerning anonymization with numerous operations of anonymization were projected in existing works. Issues of data privacy should be addressed without delay earlier than data sets are shared on cloud. Numerous distributed algorithms are

projected to defend confidentiality of numerous data sets which are retained by numerous parties. Data processing frameworks such as MapReduce was incorporated with cloud to make available controlling computation ability for applications thus, it is promising to accept such frameworks to tackle scalability difficulty of anonymizing extensive data for preserving of privacy. Our research mostly spotlight on scalability concern of top-down specialization anonymization, and is, consequently, orthogonal and harmonizing to them. Even though several distributed algorithms were projected they mostly spotlight on protected anonymization of data sets from numerous parties, to a certain extent than the scalability features. In our work, we put forward an extremely scalable two-phase top-down specialization method for data anonymization on basis of MapReduce on cloud and consist of three components of such as data separation, anonymization level integration, as well as data specialization. For the most part of top-down specialization algorithms are centralized, outcomes in their failure in handling extensive data sets.

REFERENCES

[1] L. Hsiao-Ying and W.G. Tzeng, "A Secure Erasure Code-Based Cloud Storage System with Secure Data Forwarding," IEEE Trans. Parallel and Distributed Systems, vol. 23, no. 6, pp. 995-1003, 2012.

[2] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM, pp. 829-837, 2011.

[3] P. Mohan, A. Thakurta, E. Shi, D. Song, and D. Culler, "Gupt: Privacy Preserving Data Analysis Made Easy," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '12), pp. 349- 360, 2012.

[4] 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.

[5] 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.

[6] 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.