Abstract

Most cloud services require users to share personal data like electronic health records for analysis of data or mining, bringing privacy concerns. In many cloud applications at present the scale of data increases in accordance with Big Data, thereby making it a complicated to commonly used software tools to handle and process a large-scale data within a tolerable elapsed time. It is challenging for previous anonymization approaches to achieve privacy preservation on large scale data sets due to insufficiency. The proposed a scalable two-phase top-down specialization (TDS) approach uses MapReduce architecture on cloud to anonymized large scale datasets finally deliberately design a group of innovative MapReduce
jobs to particularly accomplish specialization computation in a highly scalable way. So the
ability of TDS and efficiency of TDS can be significantly improved over existing approaches.

References

- Xuyun Zhang, Laurence T. Yang, Chang Liu, Jinjun Chen,"A Scalable Two-Phase
- P. Mohan, A. Thakurta, E. Shi, D. Song, and D. Culler,"Gupt: Privacy
Preserving Data Analysis Made Easy," Proc. ACM SIGMOD Int&apos;s Conf.
- L. Hsiao-Ying and W. G. Tzeng,"A Secure Erasure Code-Based Cloud Storage
- N. Mohammed, B. C. Fung, and M. Debbabi,"Anonymity Meets Game Theory:
- 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.
- B. Fung, K. Wang, L. Wang, and P. C. K. Hung,"Privacy-Preserving Data
- B. C. M. Fung, K. Wang, and P. S. Yu,"Anonymizing Classification Data for

**Index Terms**

Computer Science
Distributed System

**Keywords**

Data Anonymization
Top-down Specialization
Mapreduce
Cloud
Privacy
Preservation