Proceedings of the 10th INDIACom; INDIACom-2016
3rd 2016 International Conference on "Computing for Sustainable Global Development", 16th – 18th March, 2016
Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA)

Performance Enhancement and Analysis of Privacy Preservation Using Slicing Approach Over HADOOP

Mr. Abhang Vikram Kishor
Assistant Professor, Computer Engineering
Amrutvahini COE, Sangamner ,Maharshatra,India
Email Id: abhangv@gmail.com

Mr. Sable Balasaheb Shrimantrao
Assistant Professor, Computer Engineering
Amrutvahini COE, Sangamner ,Maharshatra,India
Email Id: balasaheb.sable@gmail.com

Abstract **– Nowdays every one is working with web resources, web faces many challenges about preserving data of individual user. Previously anonymization techniques are used to preserve data like Generalization and Bucketization. New technique introduced which partitioned data horizontal as well vertical called as Slicing. Slicing is used to handle huge amount of data but processing time is more.Our Proposed work based on hadoop distributed file system. our contribution in this paper is we implemented slicing techniques over hadoop. Our result shows efficiency of slicing improved over hadoop distributed file system. Slicing approach is useful to preserve Medical and Government data.**

Keywords – ***Anonymization,HDFS,Mapper ,Reducer,Slicing***

NOMENCLATURE

HDFS- Hadoop distributed file system.

I. INTRODUCTION

Extracting large amount of data from databases is called as Data Mining. Large amount of data is concern about privacy of own data and organization data. Privacy preservation means hiding important data from users. Li.N, Li.T (2012) divided attributes into three types:1) Unique attributes or Identifiers for e.g. social security number or Privacy number. 2) Quasi Identifiers which are known to individuals or other for e.g. date of birth, gender, city code. 3) sensitive attribute which are not known to everyone for E.g. Salary of person, diseases of person etc. We consider these attributes in our work to preserve data. Below table shows the classification of attributes.[1]

Data Anonymization techniques like bucketization generalization used for privacy preservation but these having some drawbacks. We implemented all these techniques and compare with our proposed system results.

Table 1. Classification of Attributes

Name	DOB	Gender	Zipcode	Disease
Vicky	2/21/86	Male	53415	Brochitis
Pradnya	6/13/81	Female	53415	Broken Arm
Rishi	3/22/76	Male	53403	Heart Disease
Dagadu	1/24/66	Male	53403	Hepatitis
Bala	4/13/89	Female	53403	Flu
Rani	2/28/86	Female	53406	Hang Nail

II. LITERATURE SURVEY

Samarati and Sweeney (1998,2001) works on K-anonymity seeks to prevent the identity disclosure of micro data which is release and based on the quasi-identifier attributes. This work fail to prevent attribute disclosure problem.[3][8] Domingo-Ferrer and Torra (2005) works on micro aggregation of the quasi-identifier attributes. Again this work fails to prevent attribute disclosure problem.[7]
Machanavajjhala (2007) And Li et al (2007) works on the technique l-Diversity and another technique t-closeness. The property of the technique l-diversity is nothing but extension of k-anonymity which one tries tosolve the problem of attribute disclosure.[7][9]

Another technique t-closeness clearly solves the attribute disclosure weakness and gives better results than k-anonymity. Ninghui Li, Ian Molloy, Jian Zhang, Tiancheng Li (2009) works on Various Anonymization techniques like Bucketization and Generalization. They work on Anonymization techniques Generalization and Bucketization

Generalization Loose some amount of data. Bucketization fails to prevent membership disclosure.[10]
Li.N, Li.T March 2012 (IEEE Transaction) works on "Slicing: The new Approach for Privacy Preserving Data publishing". Introduce new technique Slicing which is used to partitions data vertically and horizontally. Slicing used to handle high dimensional data. [1]

We studied all these papers. We analyzed slicing is gives better performance as compare to bucketization

Scanned with CamScanner