ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

A Survey: Efficient Group Based Data Retrieval from Cloud Storage Using Data Mining Technique

Ankita Prajapati, Dipak.C.Patel

M. Tech Student, Department of Computer Engineering, UV Patel College of Engineering, Kherva, Gujarat. India

Assistant Professor, Department of Computer Engineering, UV Patel College of Engineering, Kherva, Gujarat. India

ABSTRACT: Now a day's user of computer system need everything on hand without location dependency with least cost and efficiently. The Cloud computing comes with lots of benefits like the user can store lots of information on Cloud server and able to access from anywhere, anytime. With this important advantage, this technology also has some issues like, Security of user information, quick retrieval of useful information from the various large amount of data and many more.So, to increase the use of Cloud and to deal with the various issue related to cloud environment as mentioned above, a survey is presented on an efficient group based data retrieval mechanism from the cloud storage using particular Data mining techniques. The survey shows the basic knowledge of cloud area and give the brief idea about the data retrieval techniques. This survey paper may helpful to understand the techniques used by different researchers in the field of cloud data retrieval and anyone can get initial startup in this domain of keyword based retrieval.

KEYWORDS: cloud computing, data mining, keyword based retrieval, IaaS, cloud storage.

I. INTRODUCTION

On the Internet, a large amount of data is distributed, heterogeneous, dynamic, and more complex. Every day people have to deal with targeted advertising, by using data mining techniques organization become more efficient by lessen costs. Large amount of data are not handle by Traditional Data Storage systems and also difficult for traditional analytic tools to analyze the large amount of Data. So Cloud Computing is capable of solving the problem of storage, analyzing and handling the data on a distributed network. In cloud computing environment applications and techniques of Data mining are very much needed because Data Privacy, data security and efficient retrieval of the data from the cloud storage is major issues while storing the data in a Cloud environment. So implementation of data mining techniques in Cloud computing will allow the users to quickly retrieve meaningful information from virtually integrated data warehouse that decreases the costs of storage and infrastructure.

The rest of the paper is organized as follows. Section 2 illustrates Survey on various efficient keyword based data retrieval approaches from the cloud storage available with their pros and cons. Section 3 contains overall comparison among all these approaches followed by conclusion in Section 4.Last section contains the list of references used.

II. SURVEY ON VARIOUS EFFICIENT KEYWORD BASED DATA RETRIEVAL APPROACHES FROM THE CLOUD STORAGE

Authors of [6] described that in recent days, due to the fast development of the information technology, the amount of data increasing every day. Large amount of data carries a lots of information. The difficulty of the organizations or individual may increase for finding relevant knowledge from the large amount of the data. At another side Cloud computing platform have a characteristics like high virtualization and high availability which can perform dynamic resource scheduling and allocation, but this requires a need of data mining concepts. In this paper authors mentioned that by combining existing data mining technology with cloud computing technology is a quite feasible way for

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

achieving efficiency. The authors also described the process of data mining, structure of cloud computing in business module and the structure of data mining platform based on cloud computing .By using simple block diagram. Distributed computing have a two feature like parallel computing and distributed storage. Parallel computing and distributed file storage are provided by cloud computing platform, so it is a good solution on both the two levels. Due to development in cloud computing which brings new developing direction in data mining platform. Due to this new generation of data mining platform is possible.

Authors of [5] described that Recently, Cloud computing is widely used technology. It is highly recommended of security of the data and data privacy in cloud computing, because many organization use a cloud platform to store the day, among of them some are personal which may kept secret, so the data to be stored on cloud platform in encrypted form. This may require additional processing and the care should be taken that, the burden on overall system should not be increased. In this research paper authors gives a detail review of various techniques by listing some limitations also, which helps user for storing data in secured manner and efficiently access of the data from cloud server. After that authors of [5] introduce a very efficient and secure system which decrease the burden of the system and also decrease complexity and enhance overall performance of the system. But authors doesn't shows any simulation results.

Authors of [16] proposed a scheme for securely keyword based searching mechanism. The user send a query and Cloud server retrieve matched file which contains the keyword and send back to client. They used privacy preserving mechanism to protect their data from unauthorized access. They compared proposed cryptographic primitive with the existing primitive cryptographic mechanism. The authors proposed data security but it is applicable to single keyword search. The proposed scheme name order-preserving symmetric encryption (OPSE) they guaranteed that this scheme works efficiently as compared to other similar approaches available. They also shows the practical experimental result for displaying efficiency of proposed scheme.

Authors at [1]proposed a model where they focus on file searching based onmultiple keywords. They arguedthat thetraditional keyword based searchsupports Boolean search which shows weatherfile may contain the keyword or not,without any relevance of datafiles.And the ranked based file retrieval using a single keyword having a poorresult. Authors also mentioned that ranking on server side which is based onorder-preserving encryption(OPE) breaks privacy of data.So, the authors provide scalable systerm withminimize information lekage.Their model prevent overload by working at userside for ranking files,where consume less bandwidth. Theyperform analysis whichshows efficiency of their proposed solution.

Authors at [4] described that, now today user remotely store own secret data on cloud .So in cloud computing, the authors focused on encrypted data which are remotely stored. In [4] authors argued that the user can search into the encrypted data using keywords without decrypting it in traditional searchable encryption schemes. These techniques of searchable encrypted data using keyword support Boolean search method, which is not sufficient. So the authors proposed secure ranked keyword search over large amount of data files which are in encrypted form in cloud, in which user retrive the rank-ordered file. So, authors also define OPSE technique and one to more order preserving mapping for retrieve efficient data from cloud.

Authors in [2] described that User can store infinite data on cloud by using limited setup and minimum usage cost. Due to the availability of resources at fast internet speed and low initial investment, the companies are motivated to store their data on the cloud. Authors propose a multi-key word search scheme which is a cluster based over encrypted cloud data. In this research paper the formation of cluster is done by the client. Authors shows the simulation results based on average searching time and comparisons required to retrieve the desired documents from the cloud server. The proposed search scheme preserves the security requirements as proposed by the existing approaches in literature but provides efficient result.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

III. COMPARISON OF VARIOUS RESEARCH SCHEMES

The table below shows a short comparison about the various schemes proposed by a researcher by taking different parameters. The table gives the description about the basic technique used with the benefits that researcher gets as well as the limitations found in schemes.

Ref No.	Keyword Based Searching	Encry... & Decry	Clustering	Ranked Based Result	Algorithm	Simu-lation Results	Remark
[1]	Yes	Yes	No	Yes	RSA & Homo - morphic	Yes	User upload the index and the encrypted File on the cloud server
[2]	Yes	Yes	Yes	No	Symmetric key Encryption	Yes	User upload cluster index, document index & encrypted document.
[4]	Yes	Yes	No	Yes	AES	Yes	User outsourced the index and the encrypted File on cloud server
[5]	Yes	Yes	No	Yes	ECC	No	User upload file with index after encryption process.
[6]	No	No	No	No	No	No	-----
[16]	Yes	Yes	No	Yes	OPSE	Yes	User upload the index and the encrypted Files on the cloud server.

Table-1: Comparative Study

IV. CONCLUSION

Analysis of different research papers show that the thousands of user store large number of data on cloud platform and when it's about the cloud security of data becomes main concern for particular user or organization, there are various algorithms or scheme suggested by many authors. Another one is efficient retrieval of data is major concern while data are stored on cloud platforms. Existing approaches check whether the keyword is exist or not in file by means of Boolean search, single keyword or multiple keywords based retrieval. Multiple keywords based retrieval can be cluster based and formation of clusters at client side consumes more time and more number of data owner involvement require so sometimes it's become difficult to handle. In approach the documents are not retrieved ranked based so after analyzing all these there may be a system in which the formation of cluster are done at the server side it lessen the burden of user or data owner and the documents are retrieved faster and the retrieval of the documents are ranked based by using vector space model.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

REFERENCES

1. D.Pratiba, Dr.G Shobha, Vijay Lakshmi P,"Efficient Data Retrieval from Cloud Storage Using Data Mining Technique", International Journal on Cybernetics and Informatics, Volume 4, No 2, pp.271- 279, 2015.
2. Rohit Handa, Rama Krishna Challa,"A Cluster Based Multi-Keyword Search on Outsourced Encrypted Cloud Data", IEEE 2nd International Conference on Computing for Sustainable Global Development, pp.115-120, 2015.
3. R.Kabilan, Dr.N.Jayaveeran,"Survey of Data Mining Techniques in Cloud Computing", International Journal of Scientific Engineering and Applied Science, Volume 1, Issue-8, pp.123-127, 2015.
4. Ms. M.R.Girme, Prof.G.M.Bhandari, "Efficient Ranked Keyword Search for Achieving Effective Utilization of Remotely Stored Encrypted Data in Cloud", International Journal of Application or Innovation in Engineering and Management Volume 3, Issue 6, pp.105-113, 2014.
5. Akshay D Kapse, Piyush K Ingole, "Secure and Efficient Search Technique in Cloud Computing", IEEE Fourth International Conference on Communication System and Network Technologies, Issue 10.1109, pp.743-747, 2014.
6. Zhu Jia, A, Zhang Ping,"Design and Implementation of Data Mining Platform Based On the Cloud Computing", IEEE Workshop on Advanced Research and Technology in Industry Application, pp.163-165, 2014.
7. Mazhen zhong, "Research of Information Retrieval in the Cloud Computing Environment", IEEE 7th International Conference on Intelligent Computation Technology and Automation, Issue 10.1109, pp.476-479, 2014.
8. Harneet Khurana,Kailash Bahl," An Approach to Mine Frequent Itemsets in Cloud Using Apriori and FP-tree Approach", International Journal of Computing and Technology, Volume 1, Issue 7, pp.387-389, 2014.
9. Ms Aishwarya S. Patil, Ms Ankita S. Patil," A Review on Data Mining Based Cloud Computing", International Journal of Research in Science & Engineering E -Volume 1, Special Issue: 1, pp.1-4, 2014.
10. N. Janardhan, T. Sree Pravallika, Sowjanya Gorantla,"An Efficient Approach for Integrating Data Mining Into Cloud Computing", International Journal of Computer Trends and Technology - Volume 4, Issue 5, pp.1291-1294, 2013.
11. Naskar Ankita, Mrs Mishra Monika R," Using Cloud Computing To Provide Data Mining Services", International Journal of Engineering and Computer Science - Volume 2, Issue 3, pp.545-550, 2013.
12. Ms Aishwarya S. Patil, Ms Ankita S. Patil," A Review on Data Mining Based Cloud Computing", International Journal of Research in Science & Engineering E -Volume 1 Special Issue 1, pp.52-55, 2012.
13. Ruxandra-Ştefania PETRE,"Data Mining In Cloud Computing ", Database System Journal – Volume 3, pp.67-71, 2012.
14. Vishal Jain, Mahesh Kumar Madam," Information Retrieval through Multi-Agent System with Data Mining in Cloud Computing ", IJCTA, volume 3, issue 1, pp.62-66, 2012.
15. Bhagyashree Ambulkar, Vaishali Borkar," Data Mining In Cloud Computing", MPGI National Multi Conference, pp.23-26, 2012.
16. Ning Cao et al., "Secure ranked keyword search over encrypted cloud Data", IEEE International Conference of Distributed Computing Systems (ICDCS), Genoa, Italy, pp.253–262, 2010.