Comparative Study of Efficient Storage Techniques in Cloud Computing

Mamta[1], Fatima[2], Mr. Mayank Deep Khare[3], Dr. Chandra Shekhar Yadav[4]

Student, CSE, NIET, Greater Noida, India [1,2]

Assistant Professor, CSE, NIET, Greater Noida, India [3]

Professor, CSE, NIET, Greater Noida, India [4]

Abstract: From the past few years, there has been a fast progress in the cloud, with the increasing number of companies using the resources from the cloud, it is very important for protecting the data from different users, which are using centralized resources. Cloud storage services avoid the cost expensive on software, personnel maintenance and provide better performance less storage cost and scalability. But maintenance of these stored data in a secure manner is a typical task in the cloud environment. In cloud computing data is stored on remote servers and accessed from the internet. Now a day in cloud computing efficiently storage of data is a big concern to solve this problem many storage techniques has been introduced. In this paper, we surveyed several existing storage techniques with their benefits and proposed approaches. This review paper helps to identify the future research areas and methods for improving the existing approaches drawbacks.

Keywords: Cloud Computing, Service model of cloud computing, Data storage, data security, service model of cloud computing, storage techniques.

I. INTRODUCTION

Cloud Computing is the storing of data and applications on remote servers and accessing them via the internet rather than saving or installing them on our personal or office computer.

The term "cloud" is used because data and applications are stored on a cloud or connected web servers. The cloud can be accessed via the cloud computing system interface software that can be simple as using a web base service which helps all the applications and files that you will need for your job and personal life.

The cloud is being used not only to store data but also inexpensive, efficient, flexible alternative to the computer to maintaining in-house computing equipment and software.

The cloud computing architecture is comprised of two parts. The front end and back end.

The front end represents a computer that you as a client sees, this side requires you to access the cloud computing system.

The back end represents our cloud computing system is comprises computers, servers, and data storage system which store all your files and information. There is a central server that ministers the system, managing traffic in the client to ensure everything runs smoothly, in addition, the central server follows the several rules called protocols. The central server also used software called middleware that allows the networked computers communicate with each other.

Cloud computing is a viable option for data storage due to various reasons as follow.

First: cloud computing applications are limitless, it allow access the applications in data from anywhere so as long as we can link with the cloud to the internet none of our data will become find to a single hard drive or location.
Second: with the movement of our file to the web. We no longer have to pay for expensive high memory computers we simply need a device that is powerful to run the middleware needed to convert to the cloud system.
Third: servers and digital storage take up physical space, so no physical space is needed at the front end [2].

It reduces all IT problems and cost. But when storing user's data into cloud data storage, security plays an important role. Sometimes user stores their sensitive data on cloud storage environment. This causes some serious security issues. So providing security to such type of critical data is one of the major issues in cloud computing [11]. Storage and computation services are almost equivalent because in cloud computing charges are according to their usage [9]. Efficient storage is also a great issue in cloud computing because our ultimate aim is to reduce the storage space. This paper discussed various techniques to store data efficiently and also analyzed some advantages and proposed approaches of those techniques which give direction to the future research.

II. CLOUD COMPUTING

Cloud computing is "An emerging It development, deployment, and delivery model, enabling real-time delivery of products, services and solutions over the internet (i.e., enabling cloud services)"[8].

A. Deployment models of Cloud Computing
*   Private cloud: a cloud infrastructure operated by a single organization and it can be managed internally or by a third party.
*   Community cloud: it is used by distinct groups (or communities) of organizations. The communities' benefits from public cloud capabilities and also know their neighbors.
*   Public cloud: the capability resources are shared with the provider's other customers. No awareness of their neighbors.
*   Hybrid cloud: it is a composition of two or more clouds. It uses the public cloud for general computing and stores customer's data in private clouds.
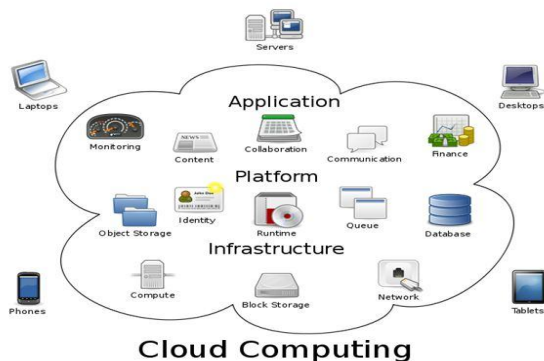

Fig.1   Cloud computing structure

B. Service models of Cloud Computing
*   Iaas (Infrastructure-as-a-service): Iaas cloud offers virtual machines, file storage, firewalls, load balancers and IP addresses.
*   Paas (platform-as-a-service): Paas cloud provides an operating system, execution runtime, database, web server and dev tools.
*   Saas (software–as-a-service): Saas cloud provides Email, CRM, virtual desktop, communication, and games.

III. LITERATURE SURVEY

Mostly storage techniques are proposed for reducing storage space; I have reviewed some of them which give better result in terms of storage space.

A. Optimal Cloud Storage System
In 2013 the authors proposed this storage technique. In this cloud data storage which requires no effort is acquiring more popularity for individual, enterprise and institutions data backup and synchronization [6]. In this system, a high-level architecture is described for a cryptographic storage service.
This architecture consists of three components
*   A data processor (DP): it processes data before it is sent to the cloud.
*   Data verifier (DV): it checks whether the data in the cloud has been tamper.
*   Token generator (TG): it generates tokens which enable the cloud storage providers to retrieve segments of consumer data.
A taxonomic approach is used for achieving cloud storage service optimality [7].
The proposed scheme contributes storage system definition, storage optimality for storage services and controller architecture for storage which is aware of optimality.

B. Process of Access And Store Small Files With Storage
In the same year, authors proposed another storage technique. Here to support internet services broadly, Hadoop distribution files system (HDFS) is used [8]. It is a highly scalable file system which adds commodity servers and disks to extend storage and IO bandwidth there are several reasons examined for small file trouble of native Hadoop distribution file system. The burden on Name node of Hadoop distribution file system is enforced by a lot of small files, for data placement correction are not considered prefetching mechanism is not also presented. For overcoming these small size problems, they proposed an approach that improves the efficiency of small file system [5]. Large clusters used where thousands of servers both host directly attached storage and execute user application task.

C. Cost Efficient Approach For Cloud Storage
In 2015 authors proposed an efficient approach for cloud storage. It reduces the storage overhead of the data centers in clouds. In this paper, they used intermediate data sets for saving data storage cost and privacy preservation. Here cloud users basically focus on storing data in the limits of data storage space while backup their data and generation of intermediate data sets.
In that case, when the user has not enough expenditure to invest in occupying larger cloud storage space, the user become bounded to store fewer data over the clouds which act that it is an unwanted limitation of cloud computing [3].
Intermediate data sets obtain very large storage space to be saved. This problem is overcome by the data compression techniques. But sometimes it is very difficult to compress and decompress the data sets when they are needed. So if we compressing all data of intermediate data sets become less efficient. So if we define a threshold value after applying heuristic algorithm [10] of selecting data sets that

need to encrypted. The threshold can be set in such a manner that the data sets which having less frequency of accessing should be compressed. So by using this approach, we can save significant storage space which reduces data storage overhead.

D. Faster Data Availability And Restoration In Disaster Cloud Data

In the same year 2015, a new technique is introduced for cloud storage. This technique is used to reduce the impact of cloud data disaster which can occur to the data center in clouds. In this model, they define user backup server and storage server. The storage server provides storage services and the backup server provides backup, restoration, and responses at direct services.

In this system user first upload their data, after when data has been uploaded successfully, it is then encrypted and dispatches to the storage server. After that message is stored as a backup by the backup server where it is again decrypted and stored it in another secure directory for dealing with disaster issue [4].

This approach is applicable to the data which is stored in the cloud and it also saved users time.

E. High-level Authorized Deduplication Method For Cloud Storage System

This technique is introduced in 2016 for cloud storage system. We know that the data storage is the biggest problem in cloud computing environment. For reducing storage space, save bandwidth and share data with others in cloud storage using public key cryptosystem that produces constant size ciphertext such that the active delegation of decoding rights for any set of ciphertexts but its leads to many privacy problems[12].

To avoid this problem hybrid cloud has been used. In this scheme, the confidential data can be stored in private cloud and other data are stored in public cloud but the exponential growth of increasing volume of data in the cloud has raised many storage challenges.

To overcome these challenges Nandhini [1] proposed a deduplication technique. It is a recent data compression technique which eliminates redundant data as well as improves the cloud storage and bandwidth utilization.

The data deduplication method removes redundant data by keeping only one physical copy and referring other redundant data to that copy.

In this paper, they define two methods of deduplication first one is file deduplication and the second one is blocked deduplication. The file deduplication method eliminates duplicate copies of the same file. And the block deduplication eliminates a duplicate block of data that occur in non-identical files.

The authorized data deduplication protects the data confidentiality by differential privileges of users in the duplicate check.

TABLE I COMPARATIVE ANALYSIS OF PROPOSED APPROACH, ADVANTAGES AND FUTURE SCOPE

Year	Storage scheme	Proposed approach	Advantages	Future scope
2013 [IJST]	Optimal cloud storage system	In this paper, they proposed a new NubiSave prototype and taxonomic approach for optimal cloud storage service.	It proposed generic architecture served as a blueprint for the optimal storage controller. NubiSave is available freely.	We can integrate NubiSave with frontends for future research.
2013 [IJETAE]	The process of Access and Store Small Files with Storage.	In this paper, they proposed a prefetching technique for storage. Which makes better access efficiency.	It improves the access ability of small files. The cut-off point is used to improve I/O performance.	In this cut-off point, the formula is not available. It will be identified in future.
2015 [IEEE]	Cost Efficient Approach for Cloud Storage	In this paper, they proposed an approach to cloud data storage, privacy, and storage of intermediate data sets in the cloud.	This approach gives a significant reduction in the size of data and reduces the storage overhead of data centers in the cloud.	Improve compression based storage system by applying a different type of compression techniques according to the type of file as a future work.
2015 [IEEE]	Faster Data Availability and Restoration in Disaster Cloud Data.	In this paper, they proposed an approach through which they can reduce the impact of cloud data disaster which occurs to the data centers in the cloud.	This approach is applicable to the data stored in the cloud and it saves user's time.	We can improve this by using the splitting method as a future work.
2016	High level Authorized	In this paper, they	It saves storage space by	It provides

[ICACCS]	Deduplication Method for Cloud Storage System.	proposed a deduplication technique which eliminates redundant data as well as improves the cloud storage and bandwidth utilization.	deduplicating the data. The authorized deduplication was proposed to protect the data confidentiality by differential privileges of users in the duplicate check.	authorization to the private firms and protects the confidentiality of important data.

IV. CONCLUSION

Cloud computing is a model for delivering information technology services, whose resources are retrieved from the internet through web-based applications and tools. Even though cloud provides many benefits to users, but the security and privacy of stored data in the cloud are still major issues in cloud storage.

We know that cloud storage is more beneficial than the other traditional storage systems especially in cost reduction, portability, scalability and functionality requirements.

After surveying relevant literature in this area, this paper presents various efficient storage techniques in cloud computing and also presented a comparative analysis of their proposed approach, advantages and future scope which helps for future research on storage methods to provide much better results.

REFERENCES

[1] Nandhini, "A review of high-level authorized deduplication method for cloud storage system", 2016 IEEE (ICACCS 2016).
[2] V.K Saxena,"Cloud computing challenges and implementations", ICEEOT, IEEE 2016 DOI: 10.1109/ICEEOT.2016.7755159.
[3] Mayank Deep Khare,"Storage cost efficient approach for cloud storage and intermediate data sets in clouds", IEEE 2015 DOI: 10.1109/ICGCIOT.2015.7380700.
[4] Aditi Tripathi, Mayank Deep Khare,"Efficient and secure approach for faster data availability and restoration in disaster cloud data management", 2015, ICGIOT, DOI: 1109/ICGIOT.2015.7380628.
[5] A.Rajathi,"A survey on secure storage in cloud computing", International journal of science and technology, ISSN: 0974-6846 April 2013.
[6] T.Sivashakthi,"A survey on storage techniques in cloud computing ", 2013, IJETAE, ISSN 2250-2459, ISO. 9001-2008 certified journal.
[7] Spillner J, Muller J et al (2012).Creating Optimal Cloud Storage System, future Generation Computer Systems,vol29(4),1062-1072.
[8] Dong B, Zheng Q et al. (2012). An optimized Approach for storing and Accessing small files on cloud storage, Journal of Network and Computer Applications,35(6),1847-1862
[9] D. Zissis and D. Lekkas, "Addressing Cloud Computing Security Issues":2011
[10] X. Zhang, H. Du, J. Chen, Y. Lin and L. Zeng,(2011) "Ensure Data Security in Cloud Storage," Proc. International Conference on Network Computing and Information Security (NCIS 11), IEEE Press, pp. 284-287.
[11] H. Takagi, J.B.D. Joshi, and G. Ahn, "Security and Privacy Challenges in Cloud Computing Environments": 2010.
[12] Chuanyi LIU, Dapeng JU, Yu GU, Youhui ZHANG, Dongsheng WANG, David H.c. Du2.Semantic Data De-duplication for Archival Storage System in IEEE Transactions, August 2008.

BIOGRAPHIES

Mamta is a student of M.Tech Computer Science and Engineering, Greater Noida. She received her bachelor Degree in computer science & engineering from Indira Gandhi Institute of Technology (IP University), Delhi. She acquired the first position in all her studies. Her research interest includes Cloud Computing and Networking.

Fatima is a student of M.Tech Computer Science and Engineering & Technology, Greater Noida. She received Master of Computer Applications degree. Her research interest is in Cloud Computing and she has taken awards for participated in cultural activities at the university level. And contributed three years in national services scheme (NSS)

Mayank Deep Khare completed M.Tech (Information Technology) from Madan Mohan Malaviya University of Technology, Gorakhpur in 2015. He is working at NIET, Greater Noida for last one year as Assistant Professor in the department of Computer Science & Engineering. He is equipped with an extraordinary caliber and appreciable academic potency. He has published 9 papers in various international and national conferences.

Dr. Chandra Shekhar Yadav is an Associate Professor and Head, Computer Science & Engineering Department at Noida Institute of Engineering & Technology (NIET), Greater Noida. He has about 17 years of experience in teaching. He received Master of Computer Applications degree from Institute of Engineering & Technology (IET), Lucknow in the year 1998, M. Tech (Computer Science & Engineering) from JSSATE, Noida in the year 2007. He received his Ph.D. degree in Computer Science & Engineering from A.P.J. Abdul Kalam Technical University, Lucknow in the year 2016. He has supervised 07 M.Tech theses. He has published 19 research papers in various national, international journals and conferences.