Implementing e-Library Using Cloud Computing

Prof. Simran Singhani[1], Aniket Goswami[2], Mayank Jain[2], Poonam Chaursia[2]

[1]*Professor* [2]*Student*

[1]simran.s@xavierengg.com

[2]aniketgoswami465@gmail.com

[2]jain19121997@gmail.com

[2]poonamchaursia1998@gmail.com

Abstract— **Cloud Computing is a set of pooled or shared resources that are delivered to users over the internet. The internet becomes a cloud and the data, applications etc are available from any device with which we can connect to the internet. This enables accessing resource remotely. Cloud Computing focuses on deployment of remote servers and software network together. It is a technology through which we can ensure data security, recovery and availability. It works on pay-per-use principle. Users have to pay according to their usage. In this paper, we present Implementing e-Library using Cloud computing, an approach through which we can reduce the storage overhead on the cloud for the purpose of reducing storage cost and preserving privacy. Users usage of cloud would be used to trace patterns in order to do prediction.**

Keywords— **Cloud Computing, File Compressors, Storage Overhead, Efficiency, Security, Prediction.**

I. INTRODUCTION

The term Cloud refers to a Network or Internet. In other words, we can say that Cloud is something, which is present at remote location. Cloud can provide services over network. The idea of cloud computing is based on a very fundamental principal of 'reusability of IT capabilities'. Cloud computing is the delivery of computing services over the Internet. Cloud services allow individuals and businesses to use software and hardware that are managed by third parties at remote locations. When you store your photos online instead of on your home computer, or use webmail or a social networking site, you are using a "cloud computing" service. The cloud computing model allows access to information and computer resources from anywhere that a network connection is available. Cloud computing provides a shared pool of resources, including data storage space, networks, computer processing power, and specialized corporate and user applications. Instead of keeping data on your own hard drive or updating applications for your needs, you use a service over the Internet, at another location, to store your information or use its applications. With the advent of this technology, the cost of computation, application hosting, content storage and delivery is reduced significantly. Cloud computing is a practical approach to experience direct cost benefits and it has the potential to transform a data center from a capital- intensive set up to a variable priced environment. Examples of cloud services include online file storage, social networking sites, webmail, and online business applications.

The following definition of cloud computing has been developed by the U.S. National Institute of Standards and Technology (NIST):

"Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, four deployment models and three service models".

i. Characteristics

On-demand self service: It means that customers can request and manage their own computing resources. Cloud Computing allows the users to use web services and resources on demand.

Broad network access: Since cloud computing is completely web based, it can be accessed from anywhere and at any time.

Resource pooling: Cloud computing allows multiple tenants to share a pool of resources. One can share single physical instance of hardware, database and basic infrastructure.

Rapid elasticity: Scaling of resources means the ability of resources to deal with increasing or decreasing demand. Resources can be scaled up or down automatically as customer needs change.

Measured service: In this service cloud provider controls and monitors all the aspects of cloud service. Resource optimization, billing and capacity planning etc. depend on it.

ii. Deployment Models

Public cloud: The Public Cloud allows systems and services to be easily accessible to the general public. Public cloud may be less secure because of its openness.

Private cloud: The Private Cloud allows systems and services to be accessible within an organization. It offers increased security because of its private nature.

Community cloud: The Community Cloud allows systems and services to be accessible by group of organizations.

Hybrid cloud: The Hybrid Cloud is mixture of public and private cloud. However, the critical activities are performed using private cloud while the non-critical activities are performed using public cloud.

iii. Service Models

IaaS(Infrastructure as a service): The user gets resources, such as processing power, storage, network bandwidth, CPU and power. Once the user acquires the infrastructure, he/she controls the OS, data, applications, services, host-based security. eg-Amazon Web Services, Rackspace

PaaS(Platform as a Service): The user is provided the hardware infrastructure, network and operating system to form a hosting environment. The user can install his/her applications and activate services from the hosting environment. eg- Microsoft Azure, Google App Engine

SaaS(Software as a Service): The user is provided access to an application. He/she does not control the hardware, network, security or operating system. eg- Google, Microsoft, Ramco

A. History

Cloud Computing [5] is said to be the biggest thing the Internet happened. In fact, cloud computing could well be the greatest technical breakthrough. At the beginning era of technology, the Client-Server architecture was popular along with the mainframe and terminal application. With the revolution in the mass storage capacity, the file servers gained the popularity for storing huge amount of information. In 1990, the giant connecting concept - Internet, finally got enough computers attached to it and the connection of those machines together create a massive, interconnected shared pool of storage that won't be possible by a single organization or institution to afford. There comes the concept of "grid". 'Grid Computing' requires the usage of application programs to divide one large system processing to several thousands of machines. But there lies the disadvantage; that if a single part of a software node fails the processing or working, other pieces of that software nodes may also fail to process. So, this 'grid'-based working concept didn't become so fruitful. On the other hand, cloud computing involves the concept of 'grid', except that it provides on-demand resource provisioning. On the first milestone of cloud technology, Salesforce.com engraved its name in 1999. It pioneered the technique of delivering enterprise application via a simple website. The next development was in 2002 by Amazon's Web Service (AWS). Then in 2006, Amazon launches their EC2 (Elastic Compute Cloud) - a commercial web service. EC2/S3 became the 1st accessible cloud technology infrastructure service. In 2009, another big milestone engraved the name of Google with Web 2.0. Google and others started to offer browser-base application via Google apps and other apps. Then came Microsoft's Azure - both Microsoft and Google deliver services in a way that is reliable and easy to consume.

B. Benefits

The benefits of Cloud Computing [6] are low up-front costs, rapid return on investment, rapid deployment, customization, flexible use and solutions that can make use of new innovations. Some other benefits to users include scalability, reliability and efficiency. Scalability means that cloud computing offers unlimited processing and storage capacity. The cloud is reliable in that it enables access to applications and documents anywhere in the world via the Internet. Cloud computing is considered efficient because it allows organizations to free up resources to focus on innovation and product development. It does not require to install a specific piece of software to access or manipulate cloud application. With the massive Infrastructure that is offered by Cloud providers today, storage & maintenance of large volumes of data is a reality. Sudden workload spikes are also managed effectively & efficiently since the cloud can scale dynamically. Also, Information in the cloud is not as easily lost when compared to the paper documents or hard drives.

C. Why cloud are popular

Cloud services are popular because they can reduce the cost and complexity of owning and operating computers and networks. Since cloud users do not have to invest in information technology infrastructure, purchase hardware or buy software licences.

"According [5] to a study by the International Data Group, 69 percent of businesses are already using cloud technology in one capacity or another, and 18 percent say they plan to implement cloud-computing solutions at some point. At the same time, Dell reports that companies that invest in big data, cloud, mobility, and security enjoy up to 53 percent faster revenue growth than their competitors". As this data clearly shows, an increasing number of tech-savvy businesses and industry leaders are recognizing the many benefits of the cloud computing trend. But more than that, they are using this technology to more efficiently run their organizations, better serve their customers, and dramatically increase their overall profit margins.
Gmail, Google Drive, TurboTax and even Facebook and Instagram are all cloud-based applications. For all of these services, users are sending their personal data to a cloud- hosted server that stores the information for later access. And as useful as these applications are for personal use, they're even more valuable for businesses that need to be able to access large amounts of data over a secure, online network connection. For example, employees can access customer information via cloud-based CRM software like Salesforce from their smartphone or tablet at home or while traveling, and can quickly share that information with other authorized parties anywhere in the world.

D. Concerns

One major hang up that many organizations have when it comes to adopting a cloud computing solution is the issue of security. For one thing, a cloud host's full-time job is to carefully monitor security, which is significantly more efficient than a conventional in-house system. And while most businesses don't like to openly consider the possibility of internal data theft, the truth is that a staggeringly high percentage of data thefts occur internally and are perpetrated by employees. When this is the case, it can actually be much safer to keep sensitive information off-site.

"RapidScale claims that 94 percent of businesses saw an improvement in security after switching to the cloud, and 91 percent said the cloud makes it easier to meet government compliance requirements."

The key to this amped-up security is the encryption of data being transmitted over networks and stored in databases. Enterprises are reluctant to buy an assurance of business data security from vendors. They fear losing data to competition and the data confidentiality of consumers. In many instances, the actual storage location is not disclosed, adding onto the security concerns of enterprises. In the existing models, firewalls across data centers (owned by enterprises) protect this sensitive information. In the cloud model, Service providers are responsible for maintaining data security and enterprises would have to rely on them.

II. PROBLEM STATEMENT

Since long people are being using hardcopies of books/documents which they want to read. These hardcopies of books/documents require space to store them and also with increase in their number it gets difficult to manage them, as we need actual physical space like a store room or a library. And also with time they get damaged. Due to digitalization, people have started using books/documents in computer readable format like ebooks, pdfs, documents. For this current situation readers will require space to store these files. Thus, "Implementing an elibrary using cloud computing" will not only overcome the problems associated with the traditional method of storing hardcopies but also provide storage space on cloud. It will provide an approach through which we can reduce the storage overhead on cloud for the purpose of saving data storage cost and privacy preservation. Users usage of cloud would be used to trace patterns inorder to do prediction.

III. RELATED WORK

In [1] the authors have proposed an approach through which we can reduce the storage overhead of the data centers in clouds. This approach can be applicable to the intermediate data sets also for the purpose of saving data storage cost and privacy preservation. An approach is developed which leads to storage of data with lesser data storage overhead. It provides options of storing data with lesser storage space to the data owners. The new approach gives significant reduction in the size of data which data owner is going to store over the clouds. In future we can improve compression techniques. We can apply different type of compression techniques according to the type of the file, So that we can improve the performance of this new compression based storage system.

To have efficient cloud storage confidentiality, [2] uses encryption and obfuscation as two different techniques to protect the data in the cloud storage. The paper uses the confidentiality parameter to address the data security problems. All the data must be encrypted or obfuscated before it is sent to the cloud database. This paper proposeda new cryptographic technique which is applied to address this problem. Algorithms are proposed for encryption and obfuscation technique.The proposed technique is secure to store the cloud users data in the cloud storage.

In [3], the authors have 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. The paper defines two methods of duplication first one is file deduplication and the second one is blocked deduplication. After surveying relevant literature in this area, the 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. Cloud data storage redefines the security issues targeted on customer's outsourced data.

In [4], authors proposed a secured cost-effective multi-cloud storage (SCMCS) model in cloud computing which holds an economical distribution of data among the available service providers in the market, to provide customers with data availability as well as secure storage. The results show that, the proposed model provides a better decision for customers according to their available budgets. In this work, to mitigate the threats facing cloud storage, the proposed model extends the cloud data storage to include multiple service providers, where each cloud storage represents a different service provider.

IV.   PROPOSED SYSTEM

Cloud computing allows us online access of the resource remotely. It benefits us while we use it for distributed data intensive applications with its capability of large data storage and high computation power.

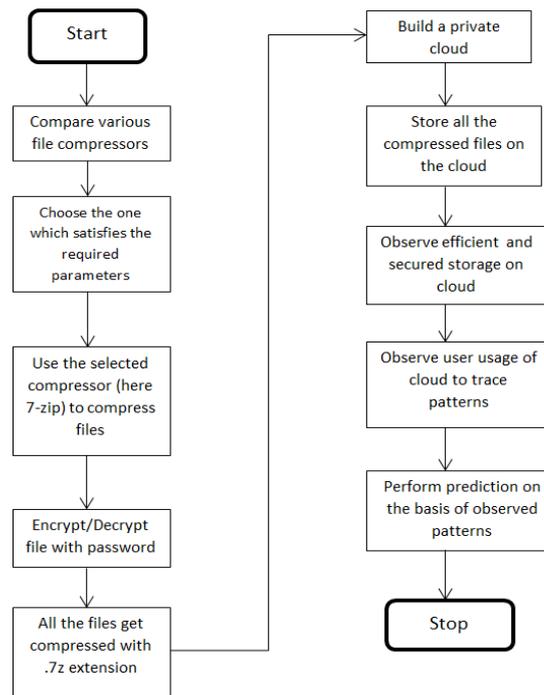Our proposed system consists of three phases-

1)   Comparison analysis of compressors.

2)   Private cloud.

3)   Prediction on cloud.

Here we plan to compare various programs which are able to create, open or extract compressed files. A file compressor is a program that can be used to compress a single file and reduce the file size on disk or to combine multiple files into a single one (the archive) reducing the size in the same time. Thus, inorder to ensure efficient and secured storage on cloud, choosing the correct file compressor is atmost important. There are various tools like WinZip, PeaZip, BandiZip, 7-Zip etc which help in compression.

We will compare these tools on the basis of certain parameters so that we can choose the right one which can help in efficient and secured storage on cloud. Efficient storage on cloud can be established by compressing the files first and then storing them on cloud. In our case, .7z has the maximum file size reduction, so all the files will compressed and saved with .7z extension. This would reduce the overall storage overhead on the cloud. Also, .7z provides features like encrypting with password, choosing encryption method etc which ensures confidentiality and security of files.

After the construction of cloud, the users will start using the cloud as per their requirement. When the cloud will be used by users for around a period of time we will be able to get trace patterns out of it on the basis of users past usage of cloud. Here we will require a data set, on the basis of this data set we can perform prediction.

V. DESIGN

```
Start ──────────────────────────────────► Build a private
  │                                            cloud
  ▼                                             │
Compare various                                 ▼
file compressors                         Store all the
  │                                      compressed files on
  ▼                                          the cloud
Choose the one                                  │
which satisfies the                             ▼
required                                 Observe efficient and
parameters                               secured storage on
  │                                          cloud
  ▼                                             │
Use the selected                                ▼
compressor (here                         Observe user usage of
7-zip) to compress                       cloud to trace
files                                        patterns
  │                                             │
  ▼                                             ▼
Encrypt/Decrypt                          Perform prediction on
file with password                       the basis of observed
  │                                          patterns
  ▼                                             │
All the files get                               ▼
compressed with ───────────────────────►     Stop
.7z extension
```

After comparing various compressors the one which satisfies the required parameters will be selected. The file will be encrypted/decrypted with the help of password only. The compressed files will be stored on the cloud here will be able to observe efficient and secured file storage. Also, on the basis of user usage of cloud will be able to trace patterns through which we can do prediction.

VI. EVALUATION AND ANALYSIS

WinZip and PeaZip work very poor when it comes at compressing pdf, ppt and word file though they work moderately well with excel files but not that well compared to 7-Zip and BandiZip. Also winzip is a paid software, user gets free service for some days and after that he needs to purchase it. The compression speed of PeaZip is more than 7-Zip but as it is not good at compressing the required file formats it won't be suitable for our purpose.

BandiZip and 7-Zip both work the best when it comes at compressing pdf, ppt, word and excel file. But BandiZip has less file size reduction than 7-Zip. 7-Zip is distinguished by simplicity and is free for home and commercial use. It is working slower than the majority of its competitors, but that doesn't means that it takes a long duration to compress or extract a file, its user friendly. Also it has support for all popular compression formats.

.7z is the best format to create an archive and to reduce the size at maximum. There are many programs able to create this type of archive and almost all are able to extract the files. We can use programs like Bandizip, PeaZip or 7- Zip. The only file type where .7z extension isn't very good at reducing file size is .exe (executable program). But we are concerned only with file formats which are used at a larger extent by students which are ppt, xsl, doc and pdf, so to compress them .7z would be the best choice. Keeping in mind that not only we will work with these files, but also the people who will receive them and need to open and extract the content, in order to use it. 7Z (.7z) is the best format to reduce file size and is supported by all major programs (create and extract).
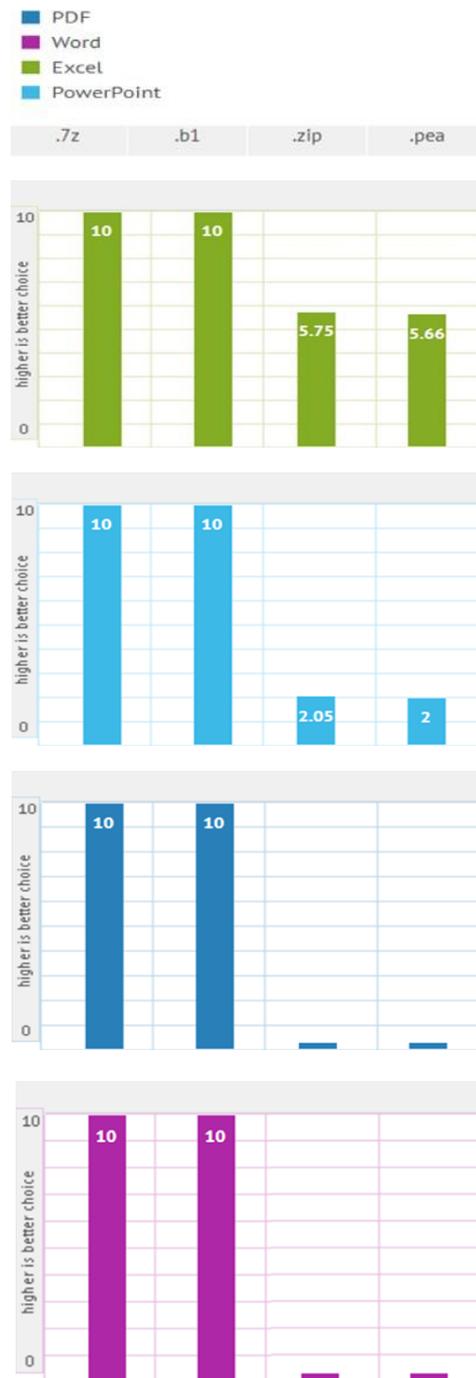
VII.   CONCLUSION

In this project, we have proposed an approach for efficient cloud data storage and privacy. Our approach gives significant reduction in the size of data which data owner is going to store over the cloud. We have compared various file compressors which help in compressing file and provide security with the help of different encryption methods and user password. Efficient storage on cloud can be established by compressing the files first and then storing them on cloud. In our case, .7z has the maximum file size reduction, so all the files will compressed and saved with .7z extension. Lastly, after the construction of cloud, the users start using it, we can trace patterns out of it on the basis of users past usage of cloud. Thus, this can help while taking important cloud related decisions. This project will provide a platform for students to maintain their personalized elibrary on the cloud. This project will overcome all the concerns associated with traditional storage methods, provide efficient storage, better security and accessibilty and would also promote prediction on the basis of users usage of cloud.

REFERENCES

[1] Mayank Deep Khare, Aditi Tripathi, Dr. Pradeep kumar Singh,(2015) "Storage Cost efficient approach for Cloud Storage and Intermediate Data Sets in Cloud," IEEE.

[2] Y. Singh, F. Kandah and W. Zhang,(2011) "A Secured cost-effective multi-cloud storage in cloud computing," IEEE INFOCOM Workshop on Cloud Computing , pp. 619-624.

[3] Mamta, Fatima, Mr. Mayank Deep Khare, Dr. Chandra Shekhar Yadav,(2017) "Comparative Study of Efficient Storage Techniques in Cloud Computing," IJARCCE.

[4] Dr.L.Arockiam, S. Monikandan,(2014) "A Secured Cost-Efficient Cloud Storage Confidentiality to Ensure Data Security," IEEE.

[5] https://www.w3schools.in/cloud-computing/history-of-cloud- computing/

[6] https://www.salesforce.com/hub/technology/benefits-of-cloud/

[7] https://en.wikipedia.org/wiki/Cloud_computing

[8] https://www.topattack.com/list/best-file-compression-software- review/35

[9] https://www.topattack.com/list/best-free-file-compression-software- review/38

[10] https://www.topattack.com/list/file-archive-formats-test-and- comparison/23