Efficient Cloud Storage Environment for Secure Client Side Using Deduplication Scheme

¹Ms. Snehal Jinnewar; ²Mr. Prasanna Padole & ³Prof. S. W. Mohod

ABSTRACT

With the internet getting so popular data sharing and security of personal data has gain much more importance than before. Deduplication is a technique for eliminating duplicate copies of data. Deduplication system improves storage utilization while reducing reliability. In this paper we propose a system that will remove 50% more duplication as composed to existing system. The proposed system will also provide better security due to file splitting and encryption. The cloud used in the proposed system is Google Drive.

Keywords:- Cloud Computing; Encryption; Deduplication; Privacy.

INTRODUCTION

Cloud computing is Internet based development and use of computer technology. It is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources. In concept, it is a model shift whereby details are abstracted from the users who no longer in control over the technology infrastructure "in the cloud" that supports them. The term cloud is used as a symbol for the Internet. It is a style of computing in which instead of keeping data on your own hard drive or updating applications for your needs, you use a service over the internet at other location which is managed by the third party. Typical cloud computing services provide common business applications online that are accessed from a web browser, while the software and data are stored on the servers over the Internet on a pay-for-use basis. All the costs associated with setting up a data center such as procuring a building, hardware, redundant power supply ,cooling systems, upgrading electrical supply, and maintaining a separate Disaster Recovery site can be passed on to a third party

vendor. Since the customer is charged only for computer services used, cloud computing costs are a fraction of traditional technology expenditures.

Cloud provide different types of deployment model such as public cloud, community cloud, private cloud, hybrid cloud. All of them have different properties and the customer can use any of them according to their requirement. Cloud also provides different types of services for customers. These services are broadly divided into three categories: Infrastructure as a Service(IAAS), Platform as a Service(PAAS), and Software as a Service (SAAS).

Engineering development and its selection are two discriminating effective variables for any business/association. Cloud computing is a late innovation ideal model that empowers associations or people to impart different administrations in a consistent and practical way. Cloud computing exhibits an opportunity for pervasive frameworks to power computational and stockpiling assets to

achieve assignments that would not typically be conceivable on such asset obliged gadgets. Distributed computing can empower programming and base planners to construct lighter frameworks that last more and are more convenient and versatile. Regardless of the favorable circumstances distributed computing offers to the originators of pervasive frameworks, there are a few impediments and constraints of distributed computing that must be tended to.

1.1 Deployment Models

Deploying cloud computing can differ depending on requirements. There are four different deployment models, each with specific characteristics that support the needs of the services and users of the clouds in particular ways.
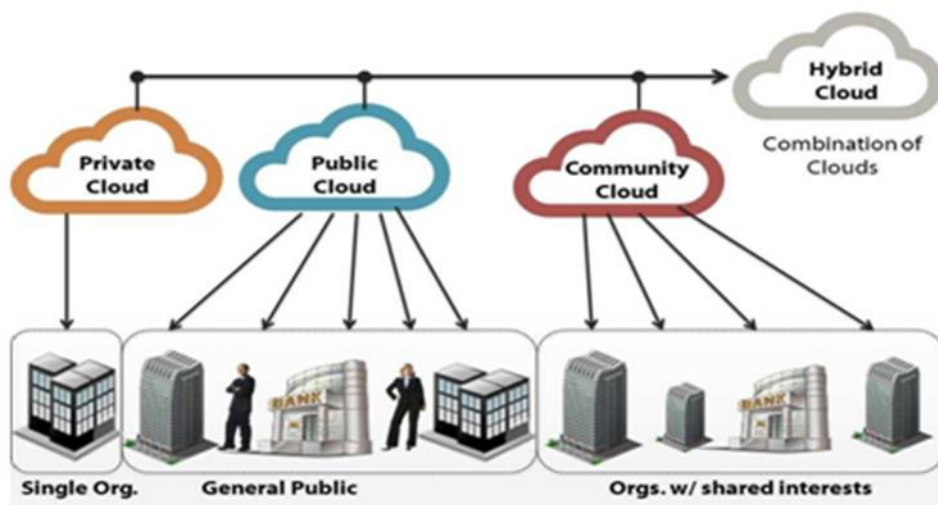

Figure 1.1: Development Models of Cloud

- Private Cloud : The cloud infrastructure has been deployed and is maintained and operated for a specific organization. The operation may be in-house or with a third party on the premises.

- Community Cloud : The cloud infrastructure is shared among a number of organizations with similar interests and requirements. This may help limit the capital expenditure costs for its establishment as the costs are shared among the organizations. The operation may be in-house or with a third party on the premises.

- Public Cloud : The cloud infrastructure is available to the public on a commercial basis by a cloud service provider. This enables a consumer to develop and deploy a service in the cloud with very little financial outlay compared to the capital expenditure requirements normally associated with other deployment options.

- Hybrid Cloud : The cloud infrastructure consists of a number of clouds of any type, but

the clouds have the ability through their interfaces to allow data and applications to be moved from one cloud to another. This can be a combination of private and public clouds that support the requirement to retain some data in an organization, and also the need to offer services in the cloud.

1.2 Service Models

Once a cloud is established, use of cloud computing services in terms of business models can differ depending on requirements. The primary service models being deployed are of three types. Each of service provides different properties and are used according to user requirements.
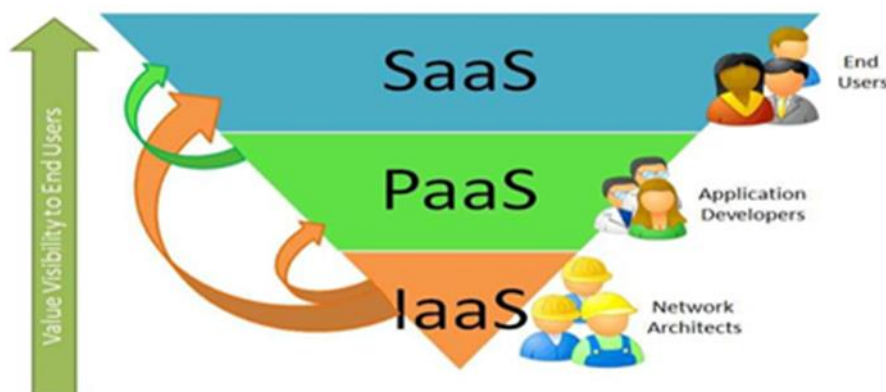

Figure 1.2: Service Models of Cloud

- Software as a Service (SAAS) : In this model, consumers has the ability to access and use an application or service that is hosted in the cloud. Cloud providers install and operate application software in the cloud and cloud users access the software from cloud client. This eliminate the need to install and run the applications on the users own computer which simplifies maintenance and support of the software. Microsoft is expanding its involvement in this area, and as part of the cloud computing option for Microsoft Office 2010, its Office Web Apps are available to Office volume licensing customers and Office Web App subscriptions through its cloud-based Online Services.

- Platform as a Service (PAAS) : In this model, consumers has access to the platforms, allowing them to install their own software and applications in the cloud. The operating systems and network access are not managed by the consumer. The cloud provider delivers a computing platform i.e. OS, database, web server etc. Application developers can develop and run their software solution on a cloud platform without the cost and complexity of buying and

managing the underlying hardware and software layers.

- Infrastructure as a Service (IAAS) : It is a form of cloud computing that provides virtualized computing resources over the Internet. It offers highly scalable resources that can be adjusted on-demand. A third part provider hosts hardware, software, servers, storage and other infrastructure components on behalf of its users. IaaS customers pay on a per-use basis, typically by the hour, week, or month.

RELATED WORK

There are numerous issues with current cloud and their architectures. Some of them are clients are regularly tied with one cloud supplier, figuring parts are firmly coupled, absence of SLA backings, absence of Multi-tenure backings, Lack of Flexibility for User Interface.

A standout amongst the most critical issues identified with cloud security dangers is information respectability. The information put away in the cloud may experience the ill effects of harm amid move operations from or to the distributed storage supplier. Cachinet al. give samples of the danger of assaults from both inside and outside the cloud supplier, for example, the as of late assaulted Red Hat Linux's appropriation servers. Another illustration of broke information happened in 2009 in Google Docs, which set off the Electronic Privacy Information Center for the Federal Trade Commission to open an examination concerning Google's Cloud Computing Services. Another case of a danger to information respectability as of late happened in Amazon S3 where clients

experienced information debasement.

One of the outcomes that they propose is to use a Byzantine blemish tolerant replication tradition inside the cloud. Hendricks et al. express that this outcome can sidestep data pollution made by a couple parts in the cloud. On the other hand, Cachinet al. declare that using the Byzantine blemish tolerant replication tradition inside the cloud is inadmissible in light of the way that the servers having a spot with cloud suppliers use the same structure foundations and are physically set in the same spot [1]. According to Garfinkel, an other security danger that may happen with a cloud supplier, for instance, the Amazon cloud organization, is a hacked mystery key or data intrusion. If some person becomes acquainted with an Amazon account mystery key, they will have the ability to get to most of the account's events and resources.

In spite of the way that cloud suppliers are aware of the noxious insider risk, they expect that they have essential responses for alleviate the issue [1]. Rocha and Correia [1] center possible aggressors for Iaas cloud suppliers. For outline, Grosse et al. [1] propose one outcome is to keep any physical access to the servers. In any case, Rocha and Correia [1] battle that the aggressors depicted in their work have remote get to and needn't trouble with any physical access to the servers. Grosse et al. [1] propose a substitute result is to screen OK to get access to the servers in a cloud where the customer's data is secured. In any case, Rocha and Correia [1] declare that this segment is profitable for watching laborer's behavior to the extent whether they are after the assurance course of action of the

association or not, in any case it is not fruitful in light of the way that it distinguishes the issue after it has happened.

A substitute technique to secure dispersed registering is for the data holder to store mixed data in the cloud, and issue deciphering keys to endorsed customers. By then, when a customer is denied, the data supervisor will issue re-encryption requests to the cloud to re-scramble the data, to keep the repudiated customer from disentangling the data, and to deliver new unscrambling keys to generous customers, so they can continue getting to the data. Of course, since a conveyed registering environment is included various cloud servers, such summons may not be gotten and executed by most of the cloud servers in view of hazardous framework correspondences [3].

A substitute way to deal with secure the data using various pressing and encryption computations and to disguise its region from the customers that stores and recuperates it. The primary complexity is that the system presented by Olfa Nasraoui [2] is an application based structure like which will keep running on the clients own system. This application will allow customers to exchange record of different associations with security quirks including Encryption and Compression. The exchanged records may be gotten to from wherever using the application which is given.

The security of the Olfa Nasraoui [2] model has been examination on the reason of their encryption estimation and the key organization. It has been watched that the encryption count have their own specific qualities; one computation gives security to the detriment of fittings, other is strong however uses more number of keys, one takes

also taking care of time. This region exhibits the diverse parameters which accept a vital part while selecting the cryptographic computation. The Algorithm found most ensuring is AES Algorithm with 256 bit key size (256k) [2].

A rule trick of cloud is data advertising. Cheng-Kang Chu, Sherman S. M. Chow, Wen-Guey Tzeng, Jianying Zhou, and Robert H. Deng [5] exhibit to securely, adequately, and adaptably grant data to others in circulated stockpiling. We depict new open key cryptosystems which convey consistent size figure messages such that capable task of unscrambling rights for any arrangement of figure works are possible. The interest is that one can add up to any arrangement of riddle keys and make them as minimized as a lone key, yet wrapping the power of each and every one of keys being collected. Toward the day's end, the puzzle key holder can release a reliable size aggregate key for versatile choices of figure substance set in appropriated stockpiling, however the other encoded archives outside the set stay mystery [5].

There are distinctive examination challenges in like manner there for grasping circulated registering, for instance, for the most part managed organization level attestation (SLA), security, interoperability and constancy. This examination paper graphs what disseminated processing is, the diverse cloud models and the standard security risks and issues that are at present inside the dispersed figuring industry. This investigation paper furthermore explores the key exploration and challenges that shows in conveyed figuring and offers best practices to organization suppliers furthermore attempts wanting to power cloud organization to upgrade their final result in

this genuine money related air [7].

Cloud based information stockpiling frameworks have numerous complexities with respect to discriminating/private/touchy information of customer. The trust needed on Cloud stockpiling is so far had been restricted by clients. The part of the paper is to develop trust in Users towards Cloud based information stockpiling. The paper handles key inquiries of the User about how information is transferred on Cloud, kept up on cloud with the goal that there is no information misfortune; information is accessible to just approved User(s) according to Client/User necessity and propelled ideas like information recuperation on catastrophe is connected [8].

Distributed computing is a versatile, fiscally smart, and showed movement stage for giving business or customer IT advantages over the Internet. On the other hand, appropriated processing demonstrates an included level of peril in light of the fact that key organizations are as often as possible outsourced to a pariah, which makes it harder to keep up data security and insurance, help data and organization availability, and show pleasantness. Circulated processing powers various advances (SOA, virtualization, Web 2.0); it moreover acquires their security issues, which we discuss here, perceiving the major vulnerabilities in this kind of systems and the most foremost threats found in the written work related to Cloud Computing and its surroundings furthermore to recognize and relate vulnerabilities and risks with possible arrangements[10].

SYSTEM ARCHITECTURE

In proposed system single file is never stored at a single place nor is duplicated at various database locations.
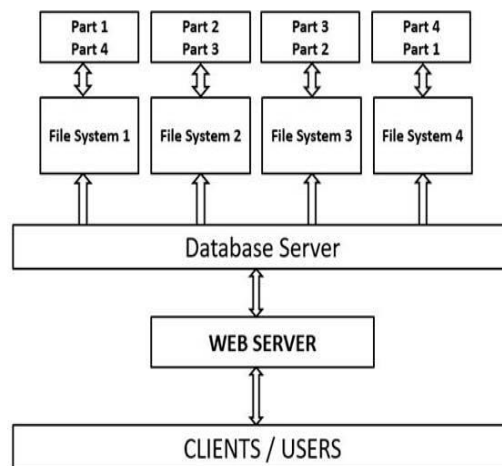

Figure 3.1: Proposed Architecture

Each file is split twice and properly inserted into different file systems. We have considered 4 databases to be used.

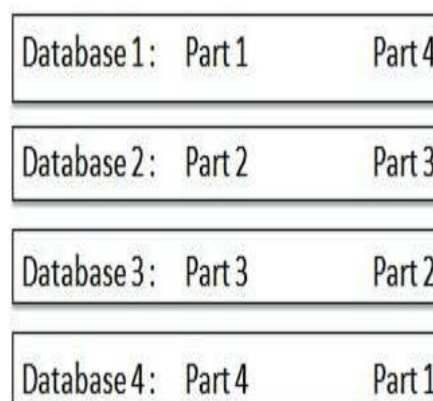The file split sequence is shown as follows:

Figure 3.2: 4 Database for Deduplication

MODULES
4.1 GUI Designing

A graphical user interface or GUI is a type of interface that allows users to interact with electronic devices through graphical icons and visual indicators such as secondary notation, as opposed to text-based interfaces, typed command labels or text navigation. GUIs were

introduced in reaction to the perceived steep learning curve of command-line interfaces (CLIs)which require commands to be typed on the keyboard.

The actions in a GUI are usually performed through direct manipulation of the graphical elements.In addition to computers, GUIs can be found in hand-held devices such as MP3 players, portable media players, gaming devices, smartphones and smaller household, office and industrial equipment. The term GUI tends not to be applied to other low-resolution types of interfaces with display resolutions, such as video games (where HUD is preferred), or not restricted to flat screens, like volumetric displays because the term is restricted to the scope of two-dimensional display screens able to describe generic information, in the tradition of the computer science research at the PARC (Palo Alto Research Center).

4.2 Database Designing

Database design is the process of producing a detailed data model of a database. This data model contains all the needed logical and physical design choices and physical storage parameters needed to generate a design in a data definition language, which can then be used to create a database. A fully attributed data model contains detailed attributes for each entity.

The term database design can be used to describe many different parts of the design of an overall database system. Principally, and most correctly, it can be thought of as the logical design of the base data structures used to store the data. In the relational model these are the tables and views. In an object database the entities and relationships map directly to object classes and named relationships.

However, the term database design could also be used to apply to the overall process of designing, not just the base data structures, but also the forms and queries used as part of the overall database application within the database management system (DBMS).

The process of doing database design generally consists of a number of steps which will be carried out by the database designer.

4.3 Connecting Website to Cloud Database

This section provides a sample script that creates a very simple webpage. You can use this webpage to test that your MySQL database is working. You can also use it as a very simple calculator. You copy the script and paste it into a text editor. Then you modify the script with your own hostname, user name, password, and database instance name information and save the changes. Finally, you copy the script to your cloud server and execute the script to display the simple webpage and test your connection to your database instance.

Your web server must be in the same region as your database instance.

4.4 File Encryption and Splitting

If the file contains sensitive information, you can encrypt the file while compressing it. Option -e encrypts the file with the given password, and the receiver should know this password for decrypting it. If the file size exceeds the specified limit after compressing also, then split the files

4.5 Removing Duplications and Testing

Removing duplication means repeated data should be deleted so that this space will be

made available for another purpose. so the less space will require and another task can be perform with that space and after that twsting is done.

CONCLUSION

IaaS is the establishment layer of the Cloud Computing conveyance demonstrate that comprises of numerous segments and innovations. Every segment in Cloud framework has its helplessness which may affect the entire Cloud's Computing security. Cloud computing business develops quickly notwithstanding security concerns, so coordinated efforts between Cloud gatherings would aid in overcoming security difficulties and push secure Cloud Computing administrations.

In this paper we said a percentage of the security worries about cloud computing furthermore proposed a framework that can help enhance the security of cloud IaaS administrations. Our methodology is intended to be executed in a multi nature.

REFERENCES

[1] Mohammed A. Alzain , Eric Pardede , Ben Soh , James A. Thom "Cloud Computing Security: From Single To Multi-Clouds", 45th Hawaii International Conference On System Sciences 2012.

[2] Olfa Nasraoui, Member, IEEE, Maha Soliman, Member, IEEE, Esin Saka, Member, IEEE, Antonio Badia, Member, IEEE, And Richard Germain "Ensuring Data Integrity And Security In Cloud Storage", IEEE TRANSACTIONS ON CLOUD AND DATA ENGINEERING, VOL. 20, No. 2, February 2013.

[3] Qin Liu ,Chiu C.Tan ,Jiewu, And Guojun Wang "Reliable Re-Encryption In Unreliable Clouds", IEEE Communications Society Subject Matter Experts For Publication In The IEEE Globecom 2011 Proceedings.

[4] Wei-Tek Tsai, Xin Sun, Janaka Balasooriya "Service-Oriented Cloud Computing Architecture", 2010 Seventh International Conference On Information Technology

[5] Cheng-Kang Chu, Sherman S. M. Chow, Wen-Guey Tzeng, Jianying Zhou, And Robert H. Deng, Senior Member, IEEE "Key-Aggregate Cryptosystem For Scalable Data Sharing In Cloud Storage", IEEE Transactions On Parallel And Distributed Systems. Volume: 25, Issue: 2. Year: 2014

[6] Mell-Peter, Grance-Timothy "The NIST Definition Of Cloud Computing", September 2011.

[7] C. Cachin, I. Keidar And A. Shraer "Trusting The Cloud", ACM SIGACT News, 40, 2009, Pp. 81-86. Clavister, "Security in The Cloud", Clavister White Paper, 2008.

[8] H.Mei, J. Dawei, L. Guoliang And Z. Yuan "Supporting Database Applications As A Service", ICDE'09:Proc. 25thintl.Conf. On Data Engineering, 2009, Pp. 832-843.C. Wang, Q. Wang, K. Ren and W. Lou, "Ensuring Data Storage Security In Cloud Computing", ARTCOM'10: Proc. Intl. Conf. On Advances In Recent Technologies In Communication And Computing, 2010, Pp. 1-9.

[9] Keiko Hashizume, David G Rosado, Eduardo Fernandez-Medina And Eduardo B Fernandez "An Analysis Of Security Issues For Cloud Computing", Hashizume Et Al. Journal Of Internet Services And Applications 2013.

[10] Gehana Booth, Andrew Soknacki, and Anil Somayaji "Cloud Security: Attacks and Current Defenses", 8th ANNUAL SYMPOSIUM ON INFORMATION ASSURANCE (ASIA'13), JUNE 4-5, 2013, ALBANY, NY.