International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 1 189 - 192

Secure Hash Based Distributed De-duplication Systems

[1]Poonam N. Patel, [2]Prof. ParulBhanarkar
TulsiramjiGaikwad-Patil College of
Engineering & Technology, Nagpur
poonampatel1308@gmail.com

Abstract: With the unstable development of computerized information, de-duplication procedures are generally utilized to reinforcement information and minimize system and capacity overhead by recognizing and taking out excess among information. Rather than keeping various information duplicates with the same substance, de-duplication takes out repetitive information by keeping stand out physical duplicate and alluding other excess information to that duplicate. De-duplication has gotten much consideration from both the scholarly world and industry in light of the fact that it can significantly enhances stockpiling usage and spare storage room, particularly for the applications with high de-duplication proportion, for example, archival capacity frameworks. Various de-duplication frameworks have been proposed taking into account different de-duplication methodologies, for example, customer side or server-side de-duplications, record level or square level de-duplications. Particularly, with the approach of distributed storage, information de-duplication systems turn out to be more alluring and discriminating for the administration of always expanding volumes of information in distributed storage administrations which inspires endeavors and associations to outsource information stockpiling.

Keywords: *Deduplication, distributed storage system, reliability, secret sharing*

*****

I. 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 forany 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.

 ISSN: 2321-8169
189 - 192

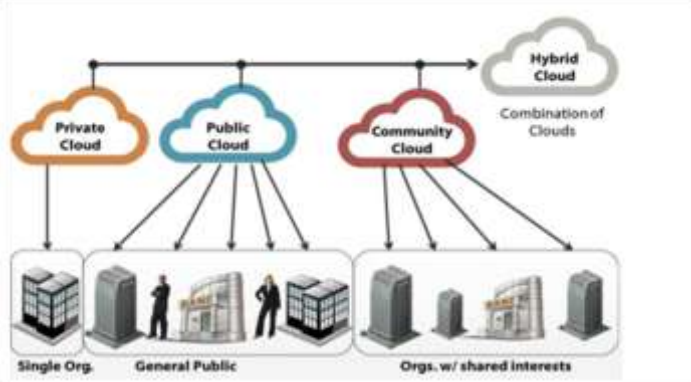

Figure 1.1: Development Models of Cloud

- 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.

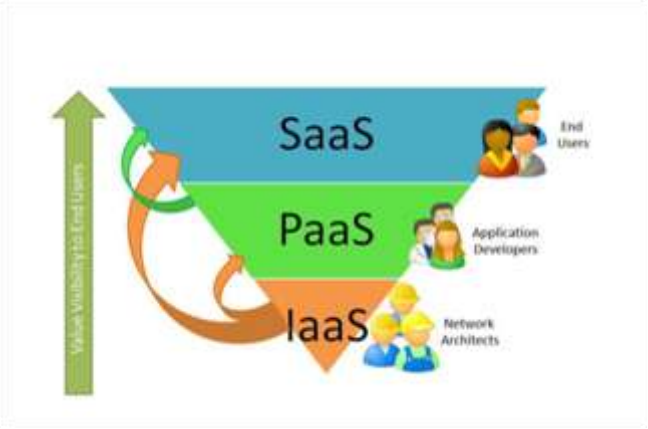

Figure 1.2: Service Models of Cloud

- 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.

II. PROPOSED SYSTEM

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

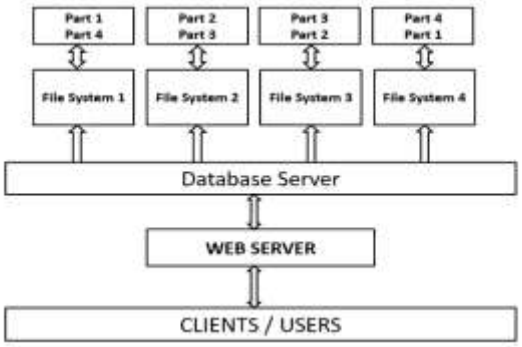

Figure 3.1: Proposed Architecture

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:

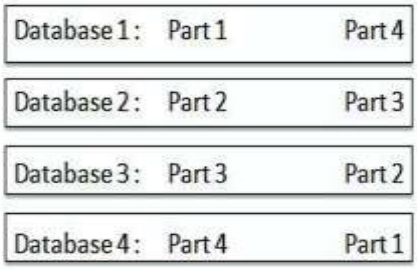

Figure 3.2: 4 Database for Deduplication

III. 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 twisting is done.

IV. CONCLUSION

Cloud computing has come to a development that leads it into a beneficial stage. This implies that the greater part of the fundamental issues with distributed computing have been tended to a degree that mists have gotten to be intriguing for full business misuse. This however does not imply that every one of the issues recorded above have really been comprehended, just that the agreeing dangers can be endured to a sure degree. Cloud computing is in this manner still as much an examination subject, as it is a business sector advertising. For better secrecy and security in distributed computing we have proposed new de-duplication developments supporting approved copy check in cross breed cloud structural planning, in which the copy check tokens of documents are created by the private cloud server with private keys. Proposed framework incorporates verification of information proprietor so it will help to actualize better security issues in distributed computing.

REFERENCES

[1] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," in *IEEE Transactions on Parallel and Distributed Systems*, 2014, pp. vol. 25(6), pp. 1615–1625.

[2] M. Li, C. Qin, P. P. C. Lee, and J. Li, "Convergent dispersal: Toward storage-efficient security in a cloud-of-clouds," in *The 6th USENIX Workshop on Hot Topics in Storage and File Systems*, 2014.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in *USENIX SecuritySymposium*, 2013.

[4] J. Xu, E.-C. Chang, and J. Zhou, "Weak leakage-resilient client-side duplication of encrypted data in cloud storage," in *ASIACCS*, 2013, pp. 195–206.

[5] D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Side channels in cloud services: Deduplication in cloud storage." *IEEE Security & Privacy*, vol. 8, no. 6, pp. 40–47, 2010.

[6] A. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, and J. C. S. Lui, "A secure cloud backup system with assured deletion and version control," in *3rd International Workshop on Security in Cloud Computing*, 2011.

[7] A. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, and J. C. S. Lui, "A secure cloud backup system with assured deletion and version control," in 3rd International Workshop on Security in Cloud Computing, 2011.

[8] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems." in ACM Conference on Computer and Communications Security, &. Cheng. Danezis, and V. Shmatikov, Eds. ACM, 2011, pp. 491–500.

[9] D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Side channels in cloud services: Deduplication in cloud storage." IEEE Security & Privacy, vol. 8, no. 6, pp. 40–47, 2010.

[10] P. Anderson and L. Zhang, "Fast and secure laptop backups with encrypted de-duplication," in Proc. of USENIX LISA, 2010.

[11] H. Shacham and B. Waters, "Compact proofs of retrievability," in ASIACRYPT, 2008, pp. 90–107.

[12] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable data possession at untrusted stores," in Proceedings of the 14th ACM conference on Computer and communications security, ser. CCS '07. New York, NY, USA:

[13] Z. Wilcox-O'Hearn and B. Warner, "Tahoe: the least-authority filesystem," in Proc. of ACM StorageSS, 2008.

[14] J. S. Plank, S. Simmerman, and C. D. Schuman, "Jerasure: A library in C/C++ facilitating erasure coding for storage applications - Version 1.2," University of Tennessee, Tech. Rep. CS-08-627, August 2008

[15] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, "Reclaiming space from duplicate files in a serverless distributed file system." in ICDCS, 2002, pp. 617–624.