ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 12, December 2016

Survey on Cloud Data Storage Service Strategies

[1] Pooja Rade [2] Vaishali Barkade
[1][2] Dept. of Computer Engineering
Rajashri Shahu College of Engineering, Tathawade,
Pune, Savitribai Phule Pune University,
Pune, India.

Abstract :-- **The use of cloud computing boosts the concept of storage outscoring which brings the concept of secure remote data auditing which we can see in research literature. Not so far from now some of the research taken the problem of security as well as efficient public data integrity auditing for the shared dynamic data. Be that as it may, these methods are still not secure against the collusion of cloud storage server and revoked group users revocation in practical cloud storage system. This paper survey on different cloud data storage service methods, there pros and cons, and discusses the infrastructure of cloud storage framework and additionally the design of cloud data storage, researches the insights about the configuration of Distributed File System inside cloud data storage.**

Key Words: -- **Cloud Service, cloud computing, data hosting, vendor lock.**

I. INTRODUCTION

Recently, study on cloud computing have examined that the online information hosting service have an incredible fame. Online storage may refer to PC information storage on a medium or a device that is under the control of a processing unit, which is capacity that is not offline storage, online record storage offer by a document hosting services, cloud storage, a model of organized enterprise storage.

A document hosting service, distributed storage service, online record stock storage supplier, or digital locker is an Internet hosting service particularly planned to host client files. It permits clients to transfer records that could then be accessed to over the web from another PC, tablet, smart phone or possibly by other users, after a password or other verification is provided. Characteristically, the services permit HTTP access, and some of the time FTP access. Related services are content-showing hosting services(i.e. video and picture), virtual storage, and remote backup.

Some online record storage services offer space on a for every gigabyte basis, and many times incorporate a data transfer capacity cost segment too. Normally these will be charged month to month or yearly for instance, Carbonite. A few organizations offer the free of cost service, depending on promoting income. Some hosting service don't put any breaking point on how much space the client's record can expend. A few administrations require a software download which makes documents just accessible on PCs which have that software installed; others permit clients to recover files through any web browser. With the expanded inbox space offered by webmail services, numerous clients have begun utilizing their webmail services as an online drive. A few locales offer free boundless document storage yet have a limit on the file size. A few destinations offer extra online storage limit in return for new client referrals.

It is a model of information storage in which the digital information is put away in logical pools, the physical storage traverses different servers (and regularly areas), and the physical environment is ordinarily possessed and managed by a hosting organization. These cloud storage suppliers are responsible for keeping the information accessible and available, and the physical environment secured and running. Individuals and associations purchase or rent storage limit from the suppliers to store users, organization, or application information. Distributed storage services might be accessed to through a co-found cloud PC benefit, a web administration application programming interface (API) or by applications that use the API, for example, cloud desktop storage, a distributed storage portal or Web-based substance services frameworks.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 12, December 2016

Security of stored information and information in transit might be a worry while storing away sensitive information at a distributed storage supplier. Clients with particular records-keeping requirements, for example, open organizations that must hold electronic records as per statute, may experience confusions with utilizing distributed computing and storage. For example, the U.S. Branch of Defense assigned the Defense Information Systems Agency (DISA) to maintain records administration products that meet the greater part of the records maintenance, personally identifiable information (PII), and security (Information Assurance; IA) prerequisites.

It is the attending utilization of two or more cloud services to minimize the risk of across data loss or downtime because of a restricted part disappointment in a distributed computing environment. Such a failure can happen in hardware, software, or framework. A multi-cloud technique can likewise enhance general endeavor execution by avoiding from " vendor lock-in " and utilizing distinctive frameworks to address the issues of various accomplices and customers Existing cloud show incredible heterogeneities in term of both working exhibitions and estimating policies. Distinctive cloud merchants build their individual infrastructure and continue redesigning them with recently developing gears. They likewise plan diverse framework architectures and apply different procedures to make their administrations aggressive. Such framework diversity qualities prompt recognizable execution varieties across cloud sellers. In recent years, most of the enterprises and organizations are using cloud to host their data into the cloud. It will result to reduce the IT maintenance cost and enhance the data reliability. In this cloud storage, digital data is stored in logical pools, the physical storage spans multiple servers (and often locations), and the physical environment is typically owned and managed by a hosting company. These cloud storage providers are responsible for keeping the data available and accessible, and the physical environment protected and running. People and organizations buy or lease storage capacity from the providers to store user, organization, or application data. There are numerous cloud vendors as well as their heterogeneous pricing policies are available, in which customers may getting confused with which cloud(s) are suitable for storing their data and what hosting strategy is cheaper. To solve this problem, comprehensive analysis should be done for customers understanding, that will help to decide which storage is suitable for them. They started is use to select several suitable clouds and an appropriate redundancy strategy to store data with minimized monetary cost and guaranteed availability. The second is triggering a transition process to re-distribute data according to the variations of data access pattern and pricing of clouds.

II. LITERATURE REVIEW

This paper [1] proposes a novel data hosting scheme called as CHARM which integrates two key functions desired, based on comprehensive investigation of different state-of-the-art cloud vendors. The first is selecting a few reasonable clouds and an appropriate redundancy procedure to store information with minimized monetary cost and ensured accessibility. The second is triggering a transition procedure to re-circulate information as indicated by the varieties of information access pattern and pricing of clouds.

This paper [2] addresses a basic yet critical query: Is the current information sync traffic of cloud storage services productively utilized? They first characterize a novel metric named TUE to evaluate the Traffic Use Efficiency of information synchronization. In view of both real-world traces and exhaustive analyses, they examine and characterize the TUE of six generally utilized cloud storage services.

This paper [3] proposes the update-batched delayed synchronization (UDS) component to address the traffic overuse issue. Acting as a middleware between the client's file storage framework and a cloud storage application, UDS batches updates from clients to significantly decrease the overhead caused by session maintenance traffic, while preserving the fast document synchronization that clients expect from cloud storage services. Besides, they expand UDS with a backwards compatible Linux kernel modification that further progresses the execution of cloud storage applications by decreasing the CPU utilization.

In this paper [4] authors present DEPSKY, a system that improves the availability, integrity and confidentiality of information stored in the cloud through the encryption, encoding and replication of the data on

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 12, December 2016

diverse clouds that form a cloud-of-clouds. They deployed their system using four commercial clouds and used PlanetLab to run clients accessing the service from different countries. They observed that their protocols improved the perceived availability and, in most cases, the access latency when compared with cloud providers individually.

This paper [5] presents an analysis of file system snapshot and five-month access trace of a campus cloud storage system that has been deployed on Tsinghua campus for three years. The system provides online storage and data sharing services for more than 19,000 students and 500 student groups. Authors report several data characteristics including file size and file type, as well as some access patterns, including read/write ratio, read-write dependency and daily traffic. They find that there are many differences between cloud storage system and traditional file systems: their cloud storage system has larger file sizes, lower read/write ratio, and smaller set of active files than those of a typical traditional file system.

In this paper [6], authors conduct the first systematic study on optimizing content multihoming, by introducing novel algorithms to optimize both performance and cost for content multihoming. In particular, they design a novel, efficient algorithm to compute assignments of content objects to content distribution networks for content publishers, considering both cost and performance. They also design a novel, lightweight client adaptation algorithm executing at individual content viewers to achieve scalable, fine-grained, fast online adaptation to optimize the quality of experience (QoE) for individual viewers.

In this paper [7], authors introduce Scalia, a cloud storage brokerage solution that continuously adapts the placement of data based on its access pattern and subject to optimization objectives, such as storage costs. Scalia efficiently considers repositioning of only selected objects that may significantly lower the storage cost. They prove the cost-effectiveness of Scalia against static placements and its proximity to the ideal data placement in various scenarios of data access patterns, of available cloud storage solutions and of failures.

Table 1. Survey Table

1.	CHARM: A Cost-efficient Multi-cloud Data Hosting Scheme with High Availability	Proposes a novel data hosting scheme (named CHARM) which integrates two key functions desired.	CHARM Model	CHARM saves monetary cost and also exhibits sound adaptability to data and price adjustments.	Need to improve performance.
2.	Towards Network-level Efficiency for Cloud Storage Services	Quantifies and analyzes the data sync traffic usage efficiency (TUE) of six widely used cloud storage services, using a real-world trace and comprehensive experiments.	TUE of cloud storage services	Data sync traffic is unnecessary and can be avoided or mitigated by careful design of data sync mechanisms.	Need to enhance the cloud storage designers to enhance their system and software, and meanwhile guide the users to pick appropriate services
3.	Efficient Batched Synchronization in Dropbox-like Cloud Storage Services	Identify a pathological issue that causes cloud storage applications to upload large amount of traffic to the cloud: many times more data than the actual content of the user's files.	Dropbox data sync mechanism	UDS significantly reduces the traffic overhead of cloud storage applications.	Need to enhance the system to reduce the CPU usage of cloud storage applications.
4	DepSky: Dependable and Secure Storage in a Cloud-of-Clouds	present DEPSKY, a system that improves the availability, integrity and confidentiality of information stored in the cloud through the encryption, encoding and replication of the data on diverse clouds that form a cloud-of-clouds	DepSky Model	It provides confidentiality and improved availability at a cost roughly double of using a single cloud for a practical scenario.	It does not consider different redundancy mechanisms.
5	Optimizing Cost and Performance for Content Multi-homing	conduct the first systematic study on optimizing content multihoming, by introducing novel algorithms to optimize both performance and cost for content multihoming	Control Framework	Content multihoming algorithms reduce publishing cost	It uses pure replication
5.	DONAR: Decentralized Server Selection for Cloud Services	Presents DONAR, a distributed system that can offload the burden of replica selection, while providing these services with a sufficiently expressive interface for specifying mapping policies.	DONAR Model	This distributed algorithm is accurate and efficient, requiring little coordination among the mapping nodes to adapt to changing client demands	Pure replication is applied in this scheme
6.	A Profit Maximization Scheme with Guaranteed Quality of Service in Cloud Computing	Double resource renting scheme is designed firstly in which short-term renting and long-term renting are combined aiming at the existing issues.	A Multi-server Model	This scheme outperforms the SQU scheme in terms of both of service quality and profit.	Vendor Lock-In Risk
7.	RACS: A Case for Cloud Storage Diversity	Make a case for applying RAID-like techniques used by disks and file systems, but at the cloud storage level.	RACS proxy Model	It enables cloud storage customers to explore trade-offs between overhead and mobility.	Pure erasure code is applied in this scheme.

III. PROPOSE SYSTEM

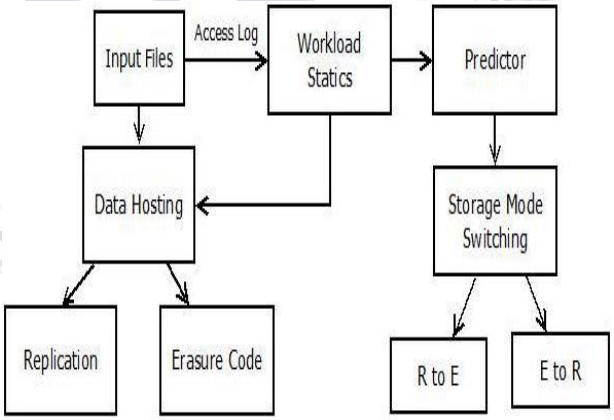

Fig 1. Propose System

This system proposes a novel cost-effective information facilitating plan with high accessibility in heterogeneous multi-cloud, named "CHARM". It insightfully places information into different clouds with minimized financial expense and ensured accessibility. In particular, system join the two broadly utilized redundancy systems, i.e., replication and deletion coding, into a uniform model to meet the required accessibility in the presence of diverse information access designs. Next, system designs an effective heuristic-based calculation to choose appropriate

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 12, December 2016

information storage modes. Additionally, system actualizes the essential method for storage mode transition by checking the varieties of information access designs and pricing policies. System evaluates the performance of CHARM using both trace driven simulations and prototype experiments. The traces are collected from two online storage systems: AmazingStore [7] and Corsair [8], both of which possess hundreds of thousands of users.

There are four main components in CHARM:

- Data Hosting.
- Storage Mode Switching (SMS).
- Workload Statistic.
- Predictor.

Workload Statistic keeps collecting and tackling access logs to guide the placement of data. It also sends statistic information to Predictor which guides the action of SMS.

Data Hosting stores data using replication or erasure coding, according to the size and access frequency of the data. SMS decides whether the storage mode of certain data should be changed from replication to erasure coding or in reverse, according to the output of Predictor. The implementation of changing storage mode runs in the background, in order not to impact online service.

Predictor is used to predict the future access frequency of files. The time interval for prediction is one month, that is, we use the former months to predict access frequency of files in the next month. However, we do not put emphasis on the design of predictor, because there have been lots of good algorithms for prediction. Moreover, a very simple predictor, which uses the weighted moving average approach, works well in our data hosting model.

Data Hosting and SMS are two important modules in CHARM. Data Hosting decides storage mode and the clouds that the data should be stored in.

IV. CONCLUSION

In this survey, we have discussed on the different existing cloud data storage service methods. Cloud services are encountering fast improvement and the services based on multi-cloud also become prevailing. One of the most concerns, when moving services into clouds, is capital expenditure. This paper presents various cloud data storage service methods and discuss there advantages and drawbacks. From this survey we have identified some major issues of redundancy, storage problem and replication issues etc. It is necessary to develop a new storage system, which guide users to distribute data among clouds cost-effectively.

REFERENCES

1. Quanlu Zhang, Shenglong Li, Zhenhua Li, Yuanjian Xing, Zhi Yang, and Yafei Dai, “CHARM: A Cost-efficient Multi-cloud Data Hosting Scheme with High Availability”, IEEE Transactions on Cloud Computing, 2015.

2. Z. Li, C. Jin, T. Xu, C. Wilson, Y. Liu, L. Cheng, Y. Liu, Y. Dai, and Z.-L. Zhang, “Towards Network-level Efficiency for Cloud Storage Services,” in IMC. ACM, 2014.

3. Z. Li, C. Wilson, Z. Jiang, Y. Liu, B. Y. Zhao, C. Jin, Z.-L. Zhang, and Y. Dai, “Efficient Batched Synchronization in Dropbox-like Cloud Storage Services,” in Middleware. ACM/IFIP/USENIX, 2013.

4. A. Bessani, M. Correia, B. Quaresma, F. Andr´e, and P. Sousa, “DepSky: Dependable and Secure Storage in a Cloud-of-Clouds,” in EuroSys. ACM, 2013.

5. S. Liu, X. Huang, H. Fu, and G. Yang, “Understanding Data Characteristics and Access Patterns in a Cloud Storage System,” in CCGrid. IEEE, 2013.

6. H. H. Liu, Y. Wang, Y. R. Yang, H. Wang, and C. Tian, “Optimizing Cost and Performance for Content Multihoming,” 2012.

7. T. G. Papaioannou, N. Bonvin, and K. Aberer, “Scalia: An Adaptive Scheme for Efficient Multi-cloud Storage,” in SC. IEEE, 2012.

8. J. S. Plank, "Erasure Codes for Storage Systems: A Brief Primer," The Usenix Magazine, vol. 38, no. 6, pp. 44–50, 2013.

9. P. Wendell, J. W. Jiang, M. J. Freedman, and J. Rexford, "Donar: Decentralized Server Selection for Cloud Services," 2010.

10. H. H. Liu, Y. Wang, Y. R. Yang, H. Wang, and C. Tian, "Optimizing Cost and Performance for Content Multihoming," 2012.

11. H. Abu-Libdeh, L. Princehouse, and H. Weatherspoon, "RACS: A Case for Cloud Storage Diversity," in SoCC. ACM, 2010.

12. T. G. Papaioannou, N. Bonvin, and K. Aberer, "Scalia: An Adaptive Scheme for Efficient Multi cloud Storage," in SC. IEEE, 2012.