Guaranteed Availability of Cloud Data with Efficient Cost

Vignesh M[1], Rajesh T M[2]

[1] M. Tech, Dept. of Computer Science and Information Technology, Dayananda Sagar University, Bengaluru, India
[2]Assistant Professor, Dept. of Computer Science and Engineering, Dayananda Sagar University, Bengaluru, India
---***---

Abstract - *Cloud computing is an innovative proficiency in the field of information and expertise. It provides so many things in terms of "As-A-Service" basis. Cloud Computing is the long-fantasized visualization of computing as a utility, where users can unconvincingly store up their records into the cloud so as the benefit of cloud on-demand high availability applications and services from a shared pool of configurable computing resources is achieved. In recent days, with increasing the population of world, the cloud based storage space popularity is also increasing rapidly. The main advantage of using cloud storage is, customer can reduce their expenditure for purchasing and maintain the storage area, but they have to use the cloud in optimized manner. In this work, we propose a cost proficient multi-cloud for the data with high accessibility. The proposed approach comprises of two techniques, the first is choosing a few reasonable cloud storage providers and a suitable repetition procedure to store information with minimalized monetary cost and guaranteed data availability. The second is setting up a redundant storage method to re-distribute the data among different available data access patterns and cloud pricing models. The proposed technique not just saves around 20 percent of monetary cost additionally shows comprehensive flexibility to information and value changes*

Key Words: Multi-cloud, data hosting, cloud storage

1.INTRODUCTION

   Cloud computing will be the core of information infrastructure in future. It provides all kinds of services for the users. The considerable service happening by the cloud is nothing but storage competence. Storage outsourcing is fetching more and more attractiveness to both industry and academic outstanding to the advantages of low cost, ease of access, and easy distribution. As one of the storage outsourcing form, cloud storage gains broad interest in recent years. Many companies, such as Amazon, Google, and Microsoft provide their own cloud storage space services, where users can upload their files to the servers, access them from a variety of devices, and share them with the others even though cloud storage services are extensively adopted in current days, there still stay behind many security issues and potential threats.

   Cloud computing provides different types of services to the users over the network. It enables enterprises to consume resources as a utility just like consumption of

electricity. Data hosting services provide users with a efficient and reliable way to store data and this stored data can be accessed from anywhere, on any device, and at any time. The term Cloud computing is defined as an internet based computing which offers on demand access to shared pool of resources on pay per use basis. Cloud computing provides distributed environment which is essential to develop large scale applications rapidly. In recent years data hosting services became more popular so that there are many cloud service providers offering data hosting services. In most of the cases companies moving towards hosting their data into a single cloud. However, in market there are several options are available from various cloud vendors and hosting service

1.1 Heterogeneous clouds

There are various cloud vendors exhibiting variations in working performances and pricing policies. They design with different system architectures and apply various techniques to provide better services. So that customers are unable to understand which clouds are suitable to host their data, this is called vendor lock in risk. It is inefficient for an organization to host all the data in a single cloud. It does not provide guaranteed availability

1.2 Multi cloud data hosting

Multi cloud data hosting is to distribute data across numerous clouds to increase more availability of the data and to minimize the loss of data, system failure risk due to a centralized component failure in a cloud environment. A failure can occur in hardware, software, or infrastructure, such a strategy also improves the overall enterprise performance by preventing possible risks such as vendor lock-in.

2. LITERATURE REVIEW

   There are many benefits of cloud computing, it also brings new challenges in maintaining data integrity and high reliable data availability. Cost is also a major concern when switching to multi-cloud system. The author Nitesh Shrivastava et al [1] made a research related to single and multi-cloud cost, security and availability based scenario. The work purposes to sustenance the usage of multi-cloud environment over single cloud in order to reduce the risk.

Prof.V.N.Dhawas et al [2] proposed a secured cost-effective multi-cloud storage prototype in cloud computing which holds an efficient distribution of data amongst the available cloud Service Providers, to deliver customers with high data availability as well as secure storage. Providing improved privacy and guaranteeing data availability can be achieved by fragmenting the customers data blocks into data pieces and putting them among the various available Service Providers. High data availability in peer-to-peer data hosting scheme demands data redundancy. In this comparison, the author HakimWeatherspoon et al [3] compares two common redundancy schemes: replication and erasure coding. In contrast to previous comparisons, the authors have taken the features of the nodes that encompass the overlap into consideration, and investigate that in some cases the advantages from coding are restricted, and possibly will not be worth its disadvantages.

Zhenhua Li et al [4], Cloud storage services such as Dropbox, Google Drive, and Microsoft OneDrive provide users with a convenient and reliable method to store and share data from anyplace, on any device, and at any time. The keystone of these services is the data synchronization operation which automatically maps the changes in users' local filesystems to the cloud through a series of network communications in a appropriate timely manner. If not structured properly, the tremendous amount of data synchronization congestion can potentially cause troubles to both cloud service providers and cloud consumers.

Understanding the integral system characteristics is crucial in designing and optimization of cloud storage system, and few studies have systematically explored the data features and access patterns. Authors S.Liu, et al [5] [6] have found that there are much diversity amongst cloud storage system and traditional file systems: the cloud storage system has high file sizes, smaller read and write ratio, and low set of active files than that of a typical traditional file system. With a trace driven mockup, they found that the cache effectiveness can be enhanced by 5 times using the inputs from authors observations. To tackle the challenges in migrating the enterprise application services into multi-cloud environment -based deployments, organizations processes are partially put on-premise and partly in the multi-cloud. The use multi-cloud architectures enable organizations to benefit from cloud-based architectures, though accepting application performance rations, and privacy limitations on which all services could be migrated to the multi-cloud.

With the increasing popularity of cloud storage is leading many enterprises to consider moving data out of their own data centers and into the cloud. However, achievement for cloud storage providers can be a significant risk to clients, as it becomes very expensive to shift storage providers. In the work presented by author Hussam Abu-Libdeh at al [7], achieves a instance for using RAID-like techniques using disks and file systems in the cloud storage, writer claims that by method of striping the user's data into multiple cloud service providers can consent customers to prevent from vendor lock-in risk, reduce the cost of switching cloud providers, and improved fault-tolerate provider failures. with introduce Redundant Array of Cloud Storage, a proxy technique that transparently blowouts the storage load across providers. The author evaluates a model of the system and evaluate the costs incurred and benefits reaped.

The study on storage system Manorama et al [8] the author implements erasure coding and replication across multiple data centers are conferred that with the cache is primary for data centers. The heterogeneity of multi cloud and the identifying of clouds are not worth the advantage, the cache benefits in storing back of data when accessed by erasure coding frequent data swap certainly induces additional cost which is longer competitive when compare to other data hosting schemes.

An extensive research is done on data hosting in peer-to-peer storage systems the authors has stated in [9],[10],[11],[12] deals with the major feature of storage system is that storage nodes are not stable. There are many merits and demerits of peer-to-peer system mechanism in multi cloud environment cannot be compared since it is proved very different from the results of other techniques. In the existing model, the enterprise based data hosting systems, data high availability is usually defined by replication or erasure coding. In the replication process, usually replicas are placed into numerous clouds, and a read access is only permitted by the low-priced cloud, so that the cloud that charges a minimal price for an out-going bandwidth and GET operation. Data replication is used for systems that are hosted in hybrid cloud where data is stored in multiple locations that are accessed my multiple applications. If the data is distributed across the multiple geographical locations there will be huge latency to access the data and process it compared to the data stored in single datacenter or location. It might also come under multiple government laws and regulations. The existing systems works on single server which cannot ensure a guaranteed backup for the optimal security and availability of the system. In the existing model concept of user data vulnerable to potential attacks are not implemented
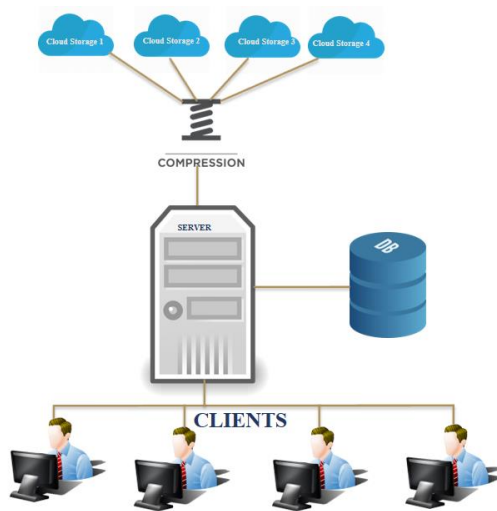
3. PROPOSED METHODOLOGY

Fig -1: Architecture of the proposed methodology

In system model think about 3 kinds of entity: the cloud service providers, the server and cloud users. Each time when, the user uploads the file to the cloud storage system, the application server decides based on the size of the file, the file requirement need of the user, the bandwidth, application server decides which cloud providers to use and upload the files with compression so as the reduce the monetary cost. These are achieved through the following phases

3.1 Multi Cloud

In the designed Application, we are proving the four-cloud storage sever for storing data from all the available clouds storage which meet the performance requirement, that is, they can offer acceptable throughput and latency when they are not in outage. The storage mode transition does not affect the performance of the service. Since it is not a latency-sensitive process, the priority of transition operations can be decreased.

3.2 Data Hosting

Data Hosting scheme stores data by means of replication or erasure coding, according to the size and frequency of the data access. Storage Mode Switching (SMS) will decide the replication process for storing the client data. The implementation of changing storage mode runs in the background, in order not to impact Client Application. Data Hosting and SMS are two significant modules in proposed work. The data hosting decides storage mode and the clouds that the data should be stored in.

3.3 File Uploading with Hashing

In File uploading module, User has to select the file to upload to cloud by selecting the number of copies of replication required to store. While uploading our application will reading file size in kb .Then it will select the best cloud storage server based on the storage availability, Pricing Cost, Predictor, Size etc.. For integrity Verification process it will generate the Hash Key (HK1) using MD5 algorithm and it will keep it in the user DB. Finally based on the Replication details the File will be store in the Cloud Storage Server.

3.4 Storage mode Switching

In Storage Mode Switching proxy, it will get the Cloud Storage Server credential from the DB like Size of the file in KB, Cost for the Uploading File. Availability Cloud Storage based on this credential it will get the Cloud Storage Server Details. Then Using FTP Protocol, it will connect to cloud to store the user data.

3.5 Cost Process

To Show the Multi-Cost Data Hosting, we have to collect the Cloud Storage Server Details like Cost, availability, Price etc., The Cost Details will be calculating by in KB.

3.6 File Download

When User wants to download the File from the Multi Cloud Storage Server, User has to select the file from the Charm Application then the Charm Server will select the corresponding cloud sever details from the DB and also it will check for the cloud availability for downloading if cloud is not available then it will download from another cloud server. While downloading it will generate the HASH Key (HK2), then it will check for the HK1 &HK2 for the Integrity Check. Finally, it will be download to the user local System

4. CONCLUSION AND FUTURE SCOPE

This paper expands Guaranteed Availability of Cloud Data with Efficient Cost that ensures High Availability, a novel storage scheme which guides customers to distribute data among clouds efficiently in large, geographically distributed and highly dynamic environments. As the demand for decentralization and delocalization of data is growing, Cloud services are experiencing rapid development and the services based on multi-cloud also become prevailing. A major challenge to moving services to the clouds is the capital involvement. The server provides a superior solution to the complexities involving as to which storage mode to use and which clouds to place data in. It appeals to the particularities of multi-cloud environment, and hosts data into multiple clouds cost-effectively, while guaranteeing flexible availability and avoiding vender lock-in. The

advantages of proposed methodology, multi-cloud approach is where an enterprise uses two or more cloud services, hence minimizing the risk of widespread data loss or outage due to a component failure, as in a single cloud computing environment. Data Hosting with low cost cloud storage Server will make effective use for client benefit. In this present work, if any one of the cloud service is inactive, the admin gets a notification in this screen, along with the available clouds and costs and push the data to the available clouds manually. In the future, we can automate the process by automatically migrate the data to the available clouds.

REFERENCES

[1] NiteshShrivastava, Ganesh Kumar A survey on cost effective multi-cloud storage in cloud computing" International Journal of Advanced Research in Computer Engineering & Technology, Volume 2, Issue 4, April 2013

[2] Prof.V.N.Dhawas,PranaliJuikar "A Secured Cost Effective Multi-Cloud Storage in Cloud Computing" International Journal of Scientific & Engineering Research, Volume 4, Issue 5, May-2013 1017 ISSN 2229-5518.

[3] HakimWeatherspoon and John D. Kubiatowicz "Erasure Coding Vs. Replication: A Quantitative Comparison" 2012

[4] Zhenhua Li 1;2, Christo Wilson 3, Zhefu Jiang 4, Yao Liu 5, Ben Y. Zhao 6, Cheng Jin 7, Zhi-Li Zhang 7, and Yafei Dai. Efficient Batched Synchronization in Dropbox-like Cloud Storage Services

[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] Mohammad Hajjat, Xin Sun, Yu-Wei Eric Sung, David Maltz, Sanjay Rao Kunwadee Sripanidkulchai and Mohit Tawarmalani,Cloudward Bound: Planning for Beneficial Migration ofEnterprise Applications to the Cloud,2010

[7] Hussam Abu-Libdeh, Lonnie Princehouse, Hakim Weatherspoon RACS: A Case for Cloud Storage Diversity

[8] Y. Ma, T. Nandagopal, K. P. Puttaswamy, and S. Banerjee, "An Ensemble of Replication and Erasure Codes for Cloud File Systems," in INFOCOM. IEEE, 2013

[9] M. Pitkanen, R. Moussa, M. Swany, and T. Niemi, "Erasure Codes for Increasing the Availability of Grid Data Storage," in AICT-ICIW. IEEE, 2006

[10] H. B. Ribeiro and E. Anceaume, "Datacube: A P2P Persistent Data Storage Architecture Based on Hybrid Redundancy Schema," in PDP. IEEE, 2010

[11] A. Duminuco and E. W. Biersack, "Hierarchical Codes: A Flexible Trade-off for Erasure Codes in Peer-to Peer Storage Systems," Peerto- Peer Networking and Applications, vol. 3, no. 1, pp. 52–66, 2010

[12] W. Lin, D. Chiu, and Y. Lee, "Erasure Code Replication Revisited," in P2P. IEEE, 2004.

BIOGRAPHIES

Vignesh M
M.Tech, Computer science and Information Technology
Dayananda Sagar University, Bengaluru, India

Dr. Rajesh T M
Assistant Professor, Department of Computer science and Engineering, Dayananda Sagar University, Bengaluru, India