# Selection of Effective Clouds for Data Hosting Services

**D. Yasashwini Raj, I. Sai Hanvitha, C. Sri Shankar, K. Manish**
Under the Guidance of **Dr. Akella Amarendra Babu**
B. Tech, Department of Computer Science and Engineering,
St. Martin's Engineering College, Hyderabad, Telangana, India.

## Abstract

In the current trend companies are moving their computing infrastructure to cloud computing from the traditional servers. The cloud hosting provides more features compared to the traditional hosting. Ability to grow at faster rate, handle huge load at the time of surge traffic. Enterprises do not need to buy/invest in hardware. We lease the infrastructure in cloud. Cloud will take away burden of server maintenance. Cloud give us flexibility to pay as per usage not a fixed cost per machine/server, we do not worry about idle servers. Organizations do not worry about the throughput or capacity of the servers as cloud provides the high availability and ability to grow and shrink as per the demand, this feature makes systems to handle huge traffic at short durations. Cloud provides replication across different servers which are spread across multiple geographical locations in the world. Cloud data centers are located in multiple locations makes us to host the server near to the location of users. Clouds are expensive compared to the traditional services. They are not flexible in terms of costs. In this paper, we explain the concept of Selection of Effective Clouds for Data Hosting Services, which provides cheaper data hosting using multiple cloud services which ensures high availability of data. It provides two key features which makes it better than the single cloud hosting. The first is using more than one cloud service provides high availability, even if one cloud services are down, still the file will available. The second is replication algorithm to store data in multiple clouds and retrieve the data from multiple clouds without any data loss. The distribution load is based on the pricing of clouds. This system makes sure we get optimum price. On an evaluation of the performance of our system, results show that it is able to model accurately the performance of cloud and to leverage this for efficient data dissemination, being able to reduce the monetary costs and transfer time.

**Keywords**—Cloud computing, Multi-Cloud, Data Hosting, Cost Efficient, Cloud Storage.

## 1. Introduction

Cloud computing get its name as a metaphor for today's internet world. Existing clouds such as Amazon S3, Windows Azure, Google cloud storage belongs to great differences in terms of working performances and pricing policies. Selecting suitable clouds and appropriate redundancy strategy to store data with minimum cost and ensured availability plays a major role is storing of data. Many Cloud vendors develop their infrastructure and keep on upgrading them with newly emerging technologies.

The cloud services can be accessed by using different models like Single service provider and multiple services providers. The problem with these models is easy for hacking and not guaranteed availability. Data outsourcing to cloud storage servers is raising trend among many firms and users owing to its economic advantages. This essentially means that the owner (client) of the data moves its data to a third party cloud storage server which is supposed to be charging a fee - faithfully store the data with it and provide it back to the owner whenever required. As data generation is far out spacing data storage it proves costly for small firms to frequently update their hardware whenever additional data is created. Also maintaining the storages can be a difficult task. Storing of user data in the cloud despite its advantages has many interesting security concerns which need to be extensively investigated for making it a reliable solution to the problem of avoiding local storage of data. Many problems like data authentication and integrity outsourcing encrypted data and associated difficult problems dealing with querying over encrypted domain are present. In this paper we deal with the problem of implementing a protocol for obtaining a proof of data possession in the cloud sometimes referred to as Proof of irretrievability (POR).

Any such proofs of data possession schemes do not, by itself, protect the data from corruption by the archive. It just allows detection of tampering or deletion of a remotely located file at an unreliable cloud storage server. To ensure file robustness other kind of techniques like data redundancy across multiple systems can be maintained. In this paper, we focus on efficient and heuristic based data hosting scheme for heterogeneous multi-cloud

environment and flexible transaction scheme. This scheme intelligently puts data into multiple clouds with minimized monetary cost and guaranteed availability.

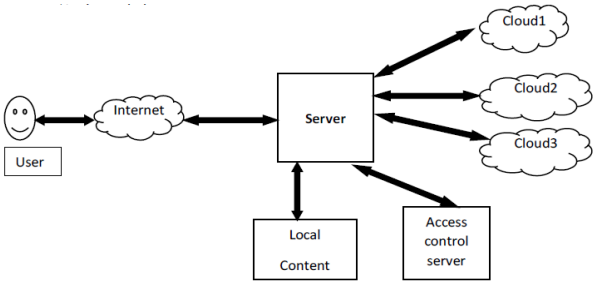


Figure 1. Basic principle of multi-cloud data hosting

Multi cloud vendor idea is to use of two or more cloud providers to store data across various services providers to get redundant service and to reduce the loss due to cloud data center downtime or server crash or network issues for a specific data center for the cloud. Such a failure can occur in hardware, software, or infrastructure. This plan also improves enterprises availability and disaster recovery and contingency plans. We can also overcome exclusive agreements with the cloud providers. With wide variety of vendors and cloud services providers we can cater to different customers and users. Various vendors provides different performances and rules which may not be matched to customer's needs. Usually the customers will put all there files and services in single cloud and depend on exclusive rules of the cloud they may not move to another cloud as it may not support the environment and structure the files or services are stored. The environments, files locations, services and other network makes the tough choice for the customers to move to new cloud providers. Vendor lock-in can be of three types with respect to cloud computing, The first is the platform lock-in where we use virtualization platforms like VMware or Zen it is very difficult for the user to migrate from one platform to another as the technology will not be compatible or support in new platform. The second is the data lock-in where data can't be moved from one cloud to another as they have different file systems or architecture to store files or data. This makes user difficult to move the data across multiple cloud architectures. The third is tools lock-in where the tools developed in one cloud platform are not supported or compatible in another platform. The user should develop tools in new cloud platform which is cost sensitive and also time consuming, these implements tend to solely be capable to run information or to manage data or apps that live in the vendor's particular cloud environment

## 2. Literature Survey

Data Storage and data integration has received a lot of attention at the data management and application level. Mansouri Y, Toosi, A.N., Buyya [1] Authors dealt with the problem of multi-cloud storage with a focus on availability and cost factors. Because of more cost the customers are unable to choose suitable cloud. There is a concern about moving large amount of data into a single cloud is similar to vendor lock-in risk. Depsky[2] author deals with stores data, even critical data into multiple clouds assuming data availability and security. Shaik. AafreenNaaz [3] author reviewed the cloud computing features provides more benefits to the users in terms of low cost and availability of data, providing security to the cloud computing is a main factor. The single cloud service provider for outstanding is not trusted because of failure in service availability and possibility of attacker like malicious virus which corrupts the stored data.

Here a multi-cloud is emerged by inter clouds or cloud of clouds where research related to single cloud problems can be addressed by using multi-cloud. Many new tools like Apache library cloud which provides a unique interface on different clouds for convenient deployment of multi-cloud services information given in [4]. This methodology helps in communication between different clouds. The advantages and disadvantages of erasure coding and replication in peer-to-peer system is given in paper[5] and [6].Here the mechanism in multi cloud environment cannot be compared because it is proved very different from the results in two works. A research is done on data hosting in Grid/peer-to-peer storage systems has stated in [9], [10], [11], [12]. Here the authors deal with the prominent feature of storage system is that storage nodes are unstable. A similar work on storage system in [13], which erasure coding and replication in multiple data centers are discussed here the author deal with cache in the primary data centers. The heterogeneity of multi cloud and the selection of clouds are not considered .the cache helps in storing back of file when accessed by erasure coding frequent data swap inevitably induces additional cost which makes long competitive when compare to other data hosting. M.P.Papazogloneetc[14] has reviewed that cloud computing technology has main drawback vendor lock-in. The cloud service developers will not allow to get service for free and does not allow to mix and match applications and services. Hence they introduced cloud blueprint so that developers to mix and match services for free of cost. By this it is facilitate to mix and match the configuration, application and stacking the resources into cloud. This approach provides simplified method for provisioning and automating cloud services and also applications run dynamically on fully virtualized clouds.

R.Thandeeswaran et al[15] has reviewed that security need to be addressed as major concern for handling critical application and sensitive information. The use of multiple cloud has following advantages

1. Exchange of data from multiple clouds.

2. Selection of clouds based on price and services.

S.Oritz Jr. [16] has reviewed that many of the industries are lack in extension of adopting cloud computing technology. The implementation leads to instability in area such as security and interoperability in turn this leads to vendor lock-in. hence the standardization is introduced which involves virtualization which play very important role in cloud computing. The data hosting schemes mentioned in our paper focus on different aspects like vendor lock-in, selecting suitable data hosting strategy, optimization of performance guaranteeing flexible availability and security.

# 3. Overview of the System

## 1) Architecture

There are two main factors and all the data are stored in the clouds, first the replication indicates no repetition of files in one cloud it avoids duplicate files so that more storage space is available in the cloud. And there will never be a work load for the admin and no complexity in the database. First, the application interacts with the cloud and sends all the data to the server, if the user is interacting with the cloud1 or requesting the cloud for example the cloud1 is busy it is forwarded to the cloud2 no time is wasted while shifting of files takes place. Then, the predictor is used to gather the file frequency, the given time is only for one month and in the next month its changes and many algorithms supports for the collecting the right data in the clouds. Then, the storage modes shift towards the storing capacity, that decides which data is suitable for cloud because storing process must not be a complexity in the cloud.

## 2) Module Description

**1. Multi-cloud:**

Around the world many different data centers have been situated. There may be a same or different data centers contains a similar or distinguished cloud suppliers. Data can be viewed by a user's in some region but he may be satisfied with the data or may not be depending on own views. He may get different outcomes some of them may be high and some of them may be low. But in our project it chooses a particular cloud for a particular data which is suitable according to the file and information which gives a good output and overcomes the limitation from the present system without effecting the other information. When server has a less load of a work we may slow down the increasing of operations and can also implement other data.

**2. Cloud Storage:**

As the cloud has become popular and getting benefits day by day, its importance is it provides security and protects the data complexity, the information which is saved is full of safe. Some may try to take the information present in the cloud due to encryption, if sometimes hacked there may be no use, In our project keys are generated and forwarded to their respective accounts, if the user mail is sometimes hacked there will be no use because the data will be in the form of cipher text, clients information is protected and maintained safety.

**3. Owner module:**

Owner modules is about to accepting the clients and their pricing models, when accepted they take their permission of admin and using secret key he uploads the files for a particular file particular key is generated and each owner gets a unique code.

**4. User module:**

This module helps the clients to download the files, first he register and logins with a given id and select the clouds he gets the generation of a secret key takes place to his mail and request to the clouds for a particular file with the id when admin gives the permission the session key is generated so that by using that key he can view the file and make in the form of decrypted file and also can be download.

**5. predictor and proxy modules:**

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. Predictor helps us to encrypt the data according to the size, if the data is small it uses replication or else heuristic algorithm and encrypts the data. Proxy sends the message to the owner in case of the cloud failure and backup of information.

## 3) Overview

**Existing System:**

Z. Li, C. Jin, T. Xu, C. Wilson, Y. Liu, L.Cheng, Y. Liu, Y. Dai,and Z.-L. Zhang [3], in this paper the author introduced the various advantages of Cloud storage services like Google drive, Drop box and many other. In existing industrial data hosting systems, data availability are usually guaranteed by replication or erasure coding. In the multi-cloud scenario, we also use them to meet different availability requirements, but the implementation is different. For replication, replicas are put into several clouds, and a read access is only served (unless this cloud is unavailable then) by the "cheapest" cloud that charges minimal for out-going bandwidth. For erasure coding, data is encoded into n blocks including m data blocks and n*m coding blocks, and these blocks are put into n different clouds. In this case, though data availability can be guaranteed with lower storage space, a read access has to be served by multiple clouds that store the corresponding data blocks. Consequently, erasure coding cannot make full use of the cheapest cloud as what replication does.

**Proposed System:**

In this paper, we propose a novel cost-efficient data hosting scheme with high availability in heterogenous multi-cloud. It intelligently puts data into multiple clouds with minimized monetary cost and guaranteed availability. Specifically, we combine the two widely used redundancy mechanisms, i.e., replication and erasure coding, into a uniform model to meet the required availability in the presence of different data access patterns. Next, we design an efficient heuristic-based algorithm to choose proper data storage modes. Moreover, we implement the necessary procedure for storage mode transition by monitoring the variations of data access patterns and pricing policies. We evaluate the performance by using both trace driven simulations and prototype experiments. The traces are collected from two online storage systems, both of which possess hundreds of thousands of users. In the prototype experiments, we replay samples from the two traces for a whole month on top of four mainstream commercial clouds: Amazon S3, Windows Azure, Google Cloud Storage, and Aliyun OSS.

# 4. Methodology

## Working Techniques:

The data hosting model is located in the proxy. There are four main components in Data Hosting, Storage Mode Switching (SMS), Workload Statistic, and Predictor. Workload Statistic keeps collecting and tackling access logs to guide the placement of data. It sends the statistic information to Predictor which guides the action of Storage Mode Switching. Data Hosting stores data using, the two widely used redundancy mechanisms, replication and erasure coding, depending upon the size and access frequency of the data. SMS decides whether the storage mode of the 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, without disturbing the Web services. Predictor is used to predict the future access frequency of files.

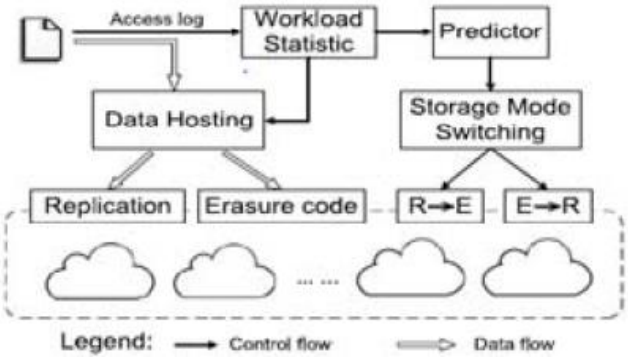


**Figure 2:** System architecture

## Modules

**1) Cloud Storage Security**

It uses the availabilities declared in the SLAs of cloud services. Most of the proposed schemes assume cloud storage service providers or the third parties handling key management, are trusted and cannot be hacked. Some entities may however, intercept communications between the users and cloud storage providers and then convince the storage providers to release user secrets by using illegal practices. In this case, encrypted data are assumed to be known and storage providers are requested to release user secrets. To avoid such situations, special kind of encryption scheme called deniable encryption scheme is employed that aims at building an encryption scheme that could help cloud storage providers avoid this predicament. In this approach, the storage providers create fake user account, from which the outside coercers can only obtain forged data from a user's stored cipher text.

**2. File Download**

When the user wants to download the file from the Multi Cloud Storage Server, user has to select the file from the Application, where the Server will select the corresponding cloud sever details from the associated database and also, will check for the cloud availability for downloading. If the cloud is not available, then it will download from another cloud server.

**3. Cost Process**

We generate the storage mode table based on the data available about the various clouds guaranteeing higher availability. We use data for different file sizes varying from 1KB to 1GB and varying read counts to calculate their corresponding storage modes. This storage mode table depends on prices of the available clouds and

required availability. If the prices of the clouds change, the table will change accordingly. The total cost includes storage/bandwidth/operation costs and transition cost. When the price adjustment occurs, it re- calculates the storage mode table, and uses it to store data and implement transition. This way it demonstrates great adaptability, when compared with the various other schemes that cannot change the clouds as dynamically depending upon the price fluctuation.

**2) Techniques Used:**

Combining Replication and Erasure Coding

In existing industrial data hosting systems, data availability are usually guaranteed by replication or erasure coding. In the multi-cloud scenario, we also use them to meet different availability requirements, but the implementation is different. For replication, replicas are put into several clouds, and a read access is only served by the "cheapest" cloud that charges minimal for out-going bandwidth. For erasure coding, data is encoded into n blocks including m data blocks and n□m coding blocks, and these blocks are put into n different clouds. In this case, though data availability can be guaranteed with lower storage space (compared with replication), a read access has to be served by multiple clouds that store the corresponding data blocks. Consequently, erasure coding cannot make full use of the cheapest cloud as what replication does. Still worse, this shortcoming will be amplified in the multi-cloud scenario where bandwidth is generally more expensive than storage space.

## 6. Results and Discussions

Home page which includes Data owner, Proxy server and User. The different symmetric key is generated for a different file when uploaded by the data owner. The process done by users when downloading, they first decrypt the file and then downloaded. Session key is generated to the user accounts to download the file, if user enters the invalid session key their account is blocked.

**1. Concern of Erasure Coding**

The computational complexity of erasure coding is one of the most significant concerns, because it needs to implement lots of multiplication operations. The CPU resource may become the bottleneck of the applications and services which apply erasure coding. Recently, however, this is not the case for erasure coding scenarios any more, coding complexity can be addressed easily using commodity CPUs, making erasure coding more popular in storage systems. Moreover, we give an upper limit to guarantee high performance, which reduces the computational overhead further. How to set the upper limit is the problem that the real system developers have to deal with through real world measurements.

**2. Other System Concerns**

As a holistic storage system, there are several other factors to be considered, such as cache strategies, geographical data consistency, etc. However, we only focus on the data hosting strategy to minimize monetary cost while meeting flexible availability requirements. Though we have considered the complexity and feasibility when designing this strategy, the system design is out of the scope of this paper, and we put the detailed system design of multi-cloud data hosting into future work.

## 7. Conclusions and Future Scope

Cloud services are experiencing rapid development and the services based on multi-cloud also become prevailing. One of the most concerns, when moving services into clouds, is capital expenditure. So, in this paper, we design a novel storage schem which guides customers to distribute data among clouds cost-effectively. It makes fine-grained decisions about which storage mode to use and which clouds to place data in. The evaluation proves its efficiency.

The scheme provided in this paper facilitates the client in getting a proof of integrity of the data, a cloud hosting and storage security that collectively deals with security and performance. It is done by two functions, the bit generator function g and the function h which is used for encrypting the data. Hence the storage at the client is very much minimal compared to all other schemes that were developed.

It 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.

## 8. References

[1] Mansouri, Y,Toosi, A.N., Buyya, R.: Brokering algorithms for optimizing the availability and cost of cloud storage services. In: Proceedings of the 2013 IEEE International Conference on Cloud Computing Technology and Science, Washington, DC, USA, vol. 01, pp. 581–589 (2013).

[2]H.Shacham and B. Waters, "Compact Proofs of Retrievability," In Proceedings. of Asiacrypt "08,Dec. 2008.

[3]Shaik. AafreenNaaz, Pothireddygari. Ramya, P.VishunuVardhanReddy.,**"**Cloud Computing: Use of Multi-Clouds" pp. 295-304s.

[4] "ApacheLibcloud"http://libcloud:apache:org/.

[5] H. Weatherspoon and J. D. Kubiatowicz, "Erasure Coding vs. Replication. A Quantitative Comparison," in IPTPS. Springer, 2002.

[6] R. Rodrigues and B. Liskov, "High Availability in DHTs: Erasure Coding vs. Replication," in IPTPS. Springer, 2005.

[7] 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.
[8] H. H. Liu, Y. Wang, Y. R. Yang, H. Wang, and C. Tian, "Optimizing Cost and Performance for ContentMultihoming," 2012.
[9] M. Pitkanen, R. Moussa, M. Swany, and T. Niemi, "Erasure Codes for Increasing the Availability of Grid Data Storage," in AICTICIW. 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.
[13] 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.
[14] M.P. Papazoglou and W. van den Heuvel, "Blueprinting the Cloud," IEEE Internet Computing, Nov. /Dec 2011, pp. 74-79
[15] R. Thandeeswaran, S. Subhashini, N. Jeyanthi1, M. A. SaleemDurai, "Secured Multi- Cloud Virtual Infrastructure with Improved Performance", cybernetics and information technologies XII, (2), pp. 11-22, 2012 Society IEEE, 2013
[16]S. Ortiz Jr., "The Problem with Cloud Computing Standardization," Computer, July 2011, pp. 13-16.
[17]"Aliyun OSS (Open Storage Service)," http://www.aliyun.com/product/ oss.
[18] "Gartner: Top 10 cloud storage providers," http:// www.networkworld.com/news/2013/010313-gartner-cloud-storage265459.html?page=1.
[19] 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.
[20] 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.
[21] C. M. M. Erin Allen, "Library of Congress and DuraCloud Launch Pilot Program Using Cloud Technologies to Test Perpetual Access to Digital Content," The Library of Congress, News Releases, http://www.loc.gov/ today/pr/2009/09-140.html.
[22] A. Li, X. Yang, S. Kandula, and M. Zhang, "CloudCmp: Comparing Public Cloud Providers," in IMC. ACM, 2010.
[22] "Windows Azure pricing updates," http://azure.microsoft.com/en-us/ updates/azure-pricing-updates/.
[23] "Google Cloud Platform pricing updates," http:// googlecloudplatform.blogspot.com/2014/03/google-cloud-platformlive-blending-iaas-and-paas-moores-law-for-the-cloud.html.
[24] "Its Official, The Nirvanix Cloud Storage Service Is Shutting Down,"http://techcrunch.com/2013/09/27/its-official-the-nirvanixcloud-storage-service-is-shutting-down/.
[25] "Shutting down Ubuntu One file services," http://blog.canonical.com/ 2014/04/02/shutting-down-ubuntu-one-file-services/.

[26] "Nirvanix Provides Cautionary Tale For Cloud Storage,"http://www.forbes.com/sites/tomcoughlin/2013/09/30/nirvanixprovides-cautionary-tail-for-cloud-storage/.
[27] "Google Outages Damage Cloud Credibility," https: //www.networkworld.com/news/2009/092409-google-outages-damagecloud.html.
[28]"Rackspace to issue as much as $3.5M in customer credits after outage," http://www.networkworld.com/news/2009/070609-rackspaceoutage.html.
[29] "Summary of the Amazon EC2 and Amazon RDS Service Disruption in the US East Region," http://aws.amazon.com/cn/message/65648/.