A COST EFFICIENT MULTI-CLOUD DATA HOSTING SCHEME USING CHARM

Dr G. Kiran Kumar[1], K. Sai Prasad[2], G. Divya Jyothi[3], T.G.Lakshman Kumar[4]

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 CHARM, 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 one cloud services down still file is available. The second is redundant 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. The system makes sure we get optimum price. On an evaluation of the performance of CHARM, the results show that our system is able to model accurately the cloud performance 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, CHARM

I. INTRODUCTION

Cloud computing uses shared resources, shared computing resources and common services used by multiple users. The services are provided on demand depends on the load and requirements. The computing machines are not owned but they are leased based on the requirements of the organization. We use the shared pool of resources rather owning the specific resources. They are billed as per usage and there is no fixed cost associated with the usage of the resources. They can be configured, upgraded and downgraded as per the need with basic setup/configuration. Cloud services come with pre-installed software's and configurations which make the users to start using the services without any time delay. Cloud computing provides variety of choices and combinations which gives maximum flexibility to the organizations to cater their requirements or environments. Cloud provide ability to grow the computing resources like disk capacity, processing capacity and network without any effort or initial installation. The servers needs no maintenance as they are placed in the third party data centers. Companies do not need to build or operate their own data centers. It surely reduces the upfront cost associated with them. The network administrators can access the systems remotely using browsers and tools.

The usage of cloud services has grown phenomenally in recent years. The cloud data centers are spread across multiple geo locations around the globe. The cloud enables the scaling of application to larger scale as it provides huge computing resources, data storage and network capacity. Companies can start providing high resource services in no time.

---

[1] *Department of CSE, MLR Institute of Technology, Hyderabad, India*
[2] *Department of CSE, MLR Institute of Technology, Hyderabad, India*
[3] *Department of CSE, MLR Institute of Technology, Hyderabad, India*
[4] *Department of CSE, MLR Institute of Technology, Hyderabad, India*

Examples of such applications running as cloud services between sites ranging from office collaborative tools (Google Drive, Microsoft Office 365), file hosting services(Dropbox, OneDrive, Box, amazon cloud drive) to mail services (Gmail, , mail.com), apart from these services they are used in huge data processing, generating analytics, huge processing like simulations, climate predictions.

Enterprises are moving to cloud has increased in recent years. They outsource their services to third party services to reduce their burden of maintaining huge data centers and having to employ team to handle the hardware, network and other infrastructure activities. Majority of the organizations use single cloud provider for all their computing requirements. Due to new entry of new cloud service provider's organizations have huge choice instead of using single service. Organizations can use multiple clouds for their reasons do not need to stick to single cloud provider. It helps in risk mitigation and avoid downtime due to problem with any data center or provider. There are many public clouds like Google compute engine, Microsoft azure, amazon web services, salesforce and soft layer from IBM.

The cloud services infrastructure which uses hardware from the single vendor/provider to get the best service or get optimal performance for reasons like compatibility issues and services agreements. These exhibit great heterogeneities while considering both the working performances and the pricing policies. The various cloud vendors build and design their own respective infrastructures, and keep upgrading with various techniques to make their services competitive. This diversity is responsible for the variety in the observed performance and differential pricing for the cloud services providers. The pricing schemes are distinct in both pricing Levels and charging items depending on the web operations, bandwidth consumption and storage spaces [1].
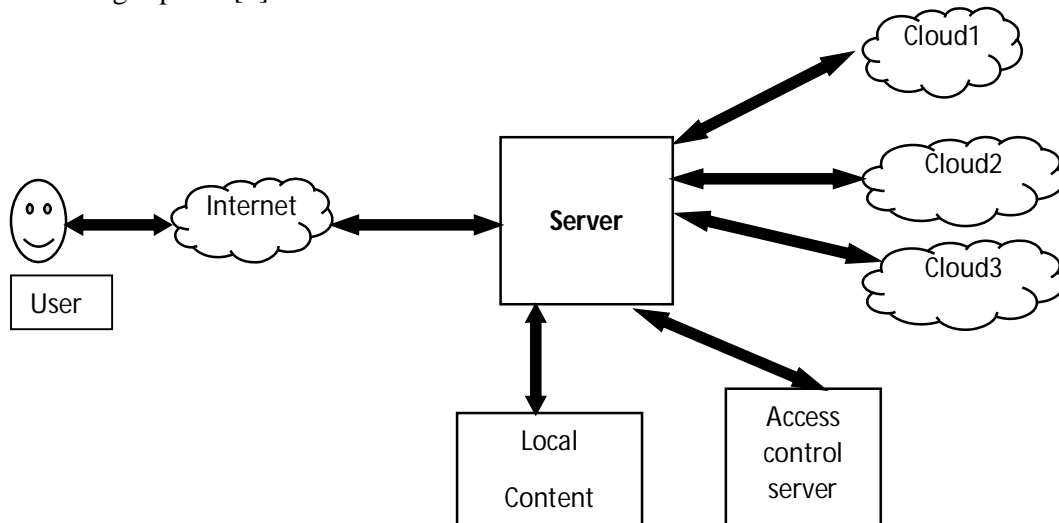

Figure 1. Basic principle of multi-cloud data hosting

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 [2].

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. That makes difficult the user to migrate to 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 [4][5]

II. PROBLEM DESCRIPTION

A. *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 the existing industrial and enterprise based data hosting systems, data availability are usually guaranteed by replication or erasure coding. For replication, replicas are put into several clouds, and a read access is only served by the cheapest cloud, such that the cloud that charges a minimal amount for out-going bandwidth and GET operation [6].
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 other method is the erasure coding. An erasure code is a forward error correction (FEC) code for the binary erasure channel, in which a message of k symbols is transformed into a longer message (code word) with n symbols, such that the original message can be recovered only from a subset of the n symbols. Here, the fraction $r = k/n$ is called the code rate, the fraction $k\square/k$, is called the reception efficiency, where $k\square$ denotes the number of symbols required for recovery. In cloud computing, data is encoded into „n$\square$ blocks including „m$\square$ data blocks and „n [] m$\square$ coding blocks, and these blocks are put into „n$\square$ different clouds. In this case, the data availability can be guaranteed within a lower storage space, but a read access has to be served by multiple clouds that store the corresponding data blocks [7].
In the multi-cloud scenario also, the replication techniques and the erasure coding mechanisms are used to handle multiple scenarios and strategies, the execution and usage of this are very different.

B. *Disadvantages of the Existing System*

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 no concept of user data vulnerable to potential attacks, a quarantined and secure hybrid architecture ensuring a dependable and reliable service is missing.

III.   PROPOSED MODEL

A. *The proposed CHARM scheme*

In this paper, we propose using a novel, efficient, and heuristic-based data hosting scheme for heterogeneous multi-cloud environments with high availability, named "CHARM". It intelligently accommodates data into multiple clouds with minimized monetary cost and guaranteed availability, considering the differential pricing strategies, availability requirements,

and data access patterns. It selects suitable clouds and an appropriate redundancy strategy to store data, while monitoring the variations of pricing policies and data access patterns, and adaptively triggers the transition process between different data storage modes.

In CHARM, the two widely used redundancy mechanisms, i.e., replication and erasure coding, are combined into a uniform model to meet the required availability in the presence of different data access patterns.

B. System Modules
1) MULTI-CLOUD

To provide for the cloud computing services, several regions are divided around the globe, and every region is usually provided with several data centers that belong to the same or different cloud providers. Technically, a user of the system may access any of the data centers in a certain region, but the user would experience an altogether different performance.

The latency of some data centers may be very low while that of others may be intolerably high. CHARM chooses a set of clouds for storing data from all the available clouds which meet the performance requirement, that is, they can offer acceptable throughput and latency when they are not in outage.
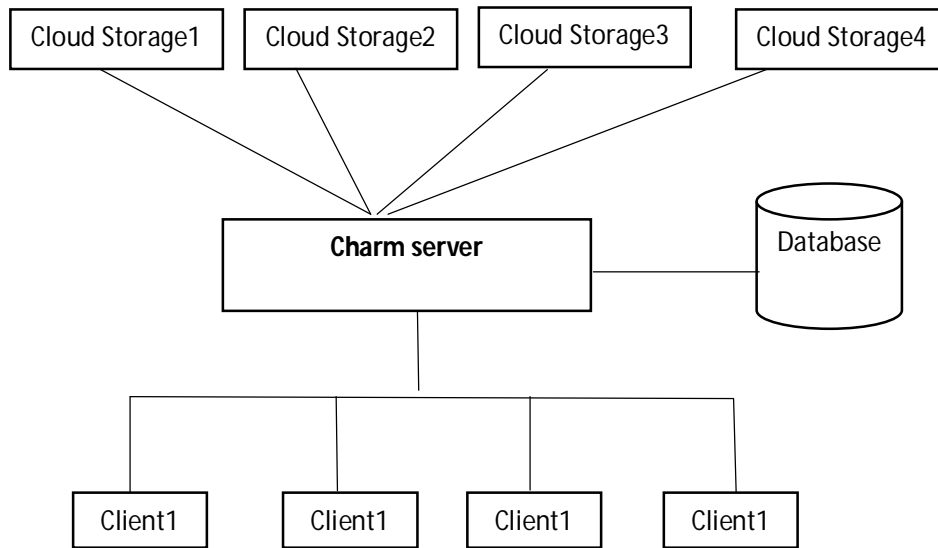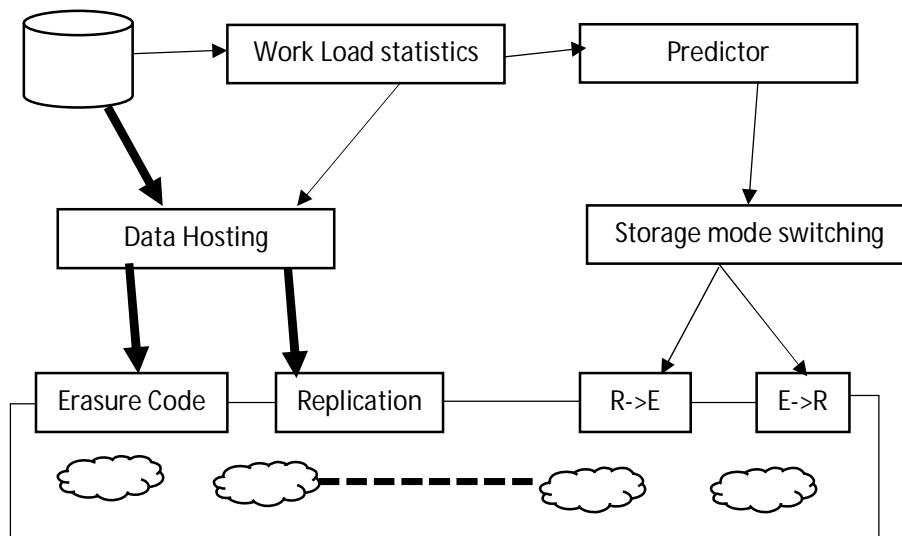
Figure 2 CHARM SYSTEM

Figure 3.THE ARCHITECTURE OF CHARM. "R" REPRESENTS REPLICATION AND "E" REPRESENTS ERASURE CODING

1)   DATA HOSTING

The data hosting model is located in the proxy. There are four main components in CHARM: 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.

2)   FILE UPLOADING WITH HASHING

In the File Uploading module, the user selects the file to be uploaded to the cloud by selecting the number of copies of replication required to be stored .While uploading, the charm server reads the file size, while, selecting the best cloud storage server, based on the storage availability, pricing cost, predictor and size. For the Integrity Verification process, it generates the Hash Key (HK1) using MD5 algorithm and keeps it in the user Database. Finally based on the Replication details the File will be stored on the Cloud Storage Server.
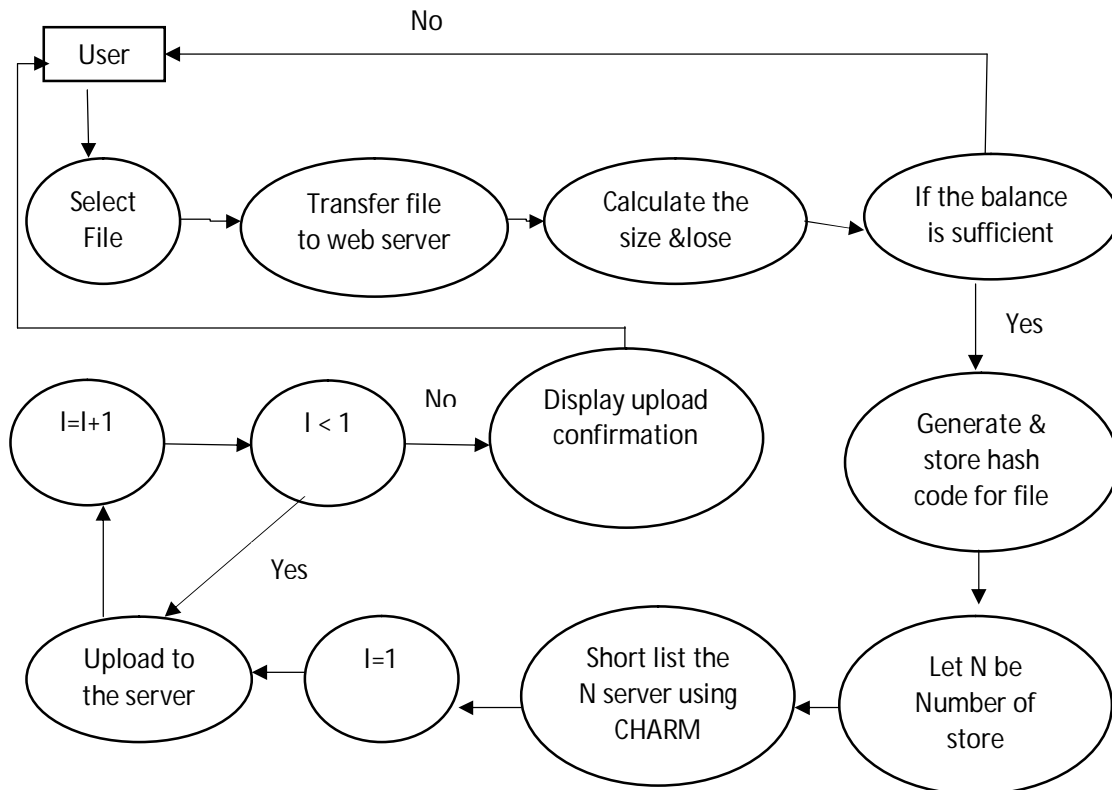

Figure 4. FILE UPLOADING

3)   CLOUD STORAGE SECURITY

CHARM 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 ciphertext.

4)   FILE DOWNLOAD

When the user wants to download the file from the Multi Cloud Storage Server, user has to select the file from the Charm Application, where the Charm 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. While downloading it, the server will generate the HASH Key (HK2), which will check the HK1 &HK2 for the Integrity Check. Finally the file would be downloaded to the user local System.

5)   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 of CHARM includes storage/bandwidth/operation costs and transition cost. When the price adjustment occurs, CHARM 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.

B.   Advantages

☐   A 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.

IV.   CONCLUSION

This paper expands CHARM, a Cost-efficient, Multi-cloud Data Hosting Scheme 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. CHARM 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.

REFERENCES

[1]   Steve Crago, Kyle Dunn, Patrick Eads, Lorin Hochstein, Dong-In Kang, Mikyung Kang, et al -- "Heterogeneous cloud computing" --- University of Southern California / Information Sciences Institute 3811 N. Fairfax, Suite 200. Arlington, VA, 22203 {crago∗, lorin† }@isi. edu,

[2]   Elpula Vijaya Kumar, V. Rama Krishna, P. Raja Sekhar Reddy -- "A Novel Approach for Multi Cloud Storage" --- CSE Dept., AGI, India,

Dr G. Kiran Kumar, K. Sai Prasad, G. Divya Jyothi, T.G.Lakshman Kumar   101

[3]   Z. Li, C. Jin, T. Xu, C.Wilson, Y. Liu, L. Cheng, Y. Liu, Dai, and Z.-L. Zhang, -- "Towards network-level efficiency for cloud storage services," -- in Proceedings of the 2014 Conference on Internet Measurement Conference, pp. 115– 128, ACM, 2014

[4]   Cloud Computing Issues -- https://en.wikipedia.org/wiki/Cloud_computing_issues

[5]   Vendor Lock-In Issues – http://emereo.net/success/cloud-infrastructure-vendor-lock-in/

[6]   Dejene Boru, Dzmitry Kliazovich, Fabrizio Granelli, Pascal Bouvry, Albert Y. Zomaya – "Energy-efficient data replication in cloud computing datacenters"

[7]   Jun Li and Baochun Li – "Erausre Coding For Cloud Storage Systems"