Volume 6, Issue 9, September 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Novel Approach to an Efficient and Reliable Data Hosting in Multi Cloud Environment

S. Deepika
MTech (CSE) Priyadarshini College of Engineering,
Nellore Andhra Pradesh, India

A. Ravi
Asst.Professor Priyadarshini College of Engineering,
Nellore Andhra Pradesh, India

Abstract: ***As cloud computing became popular, more and more organizations are interested to host their data into the cloud to reduce maintenance cost and enhance the data reliability. There are several cloud service providers exhibiting great variations in their pricing policies and working performances. Due to the enormous growth of heterogeneous cloud data centers the users are now able to put their data in any of the cloud data centers. Usually customers put their data in a single cloud and simply trust to luck. However the problem is choosing a reliable cloud data center which performs well and also to understand which hosting strategy is better. To ensure reliability cloud service providers uses replicas. It is essential to load balance among the cloud data centers for better performance. Based on the Existing schemes, we proposed a novel data hosting scheme by selecting several suitable clouds and an appropriate redundancy strategy to store data with minimum cost and high availability.***

Keywords: Data Hosting load balancing, Replica, Eruser coding.

I. INTRODUCTION

In recent years there is a rapid movement of people towards online data hosting services. So that many cloud service providers are offering such services. Data hosting is to store data on a server or other computer so that it can be accessed over the internet. Sometimes companies required particular resources for limited period of time then they need not to purchase those resources. Companies can use resources over a network on pay per use basis.

Cloud computing provides different types of services to the users over the network. It enables companies to consume resources as a utility just like 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. Cloud computing is internet based computing which provides on demand access to shared pool of resources and data 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 became available from various cloud vendors

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

Multi cloud data hosting:

Multi cloud data hosting is to distribute across multiple clouds to gain more availability of the data and to minimize the risk of data loss or system failure due to a centralized component failure in a cloud computing environment. Such a failure can occur in hardware, software, or infrastructure. Such a strategy also improves the overall enterprise performance by avoiding potential risks such as "vendor lock-in".

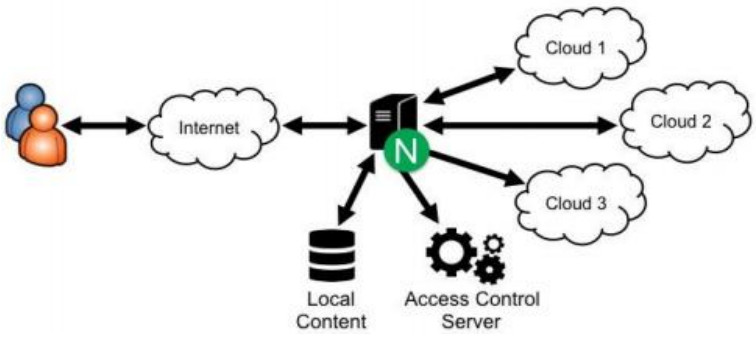

Fig 1.1 Multi Cloud Architecture.

II. BACKGROUND

In existing cloud data hosting systems, availability of data are usually guaranteed by replication or erasure coding. In the multi-cloud environment we also use the above two mechanisms to achieve different availability requirements, but the implementation is different for both of them. Replication is achieved by using redundancy, replicas are placed in several clouds[1], to read data it accesses the "cheapest" cloud that charges minimal out-going bandwidth and GET operation unless it is unavailable. Data replication is suitable for systems with distributed applications. For erasure coding, there are m data blocks and data is encoded into n blocks. M data blocks and n-m coding blocks are placed into n different clouds. In this case, compared with replication data availability is guaranteed with lower storage space, to read data multiple clouds need to be accessed which are storing the corresponding data blocks. However erasure coding cannot make full use of the cheapest cloud as replication. In the multi-cloud scenario bandwidth is generally (much) more expensive than storage space. In the multi-cloud scenario the replication techniques and the erasure coding mechanisms are used to meet different availability requirements, but the implementation of these are very different.

The two problems related to multi cloud are

- How to choose appropriate clouds in the presence of heterogeneous pricing policies which provides minimum monetary cost.
- How to meet different cloud availability requirements of different hosting services.

III. RELATED WORK

Micheal vrable proposed blue sky[2] , a network file system backed by cloud storage. The blue sky stores data persistently in a cloud storage provider such as windows Azure or Amazon S3 allowing users to take advantage of the large storage capacity and reliability of cloud providers to avoid the need for dedicated server hardware.

Harsha v.madhyastha [3] proposed a storage configuration compiler (SCC)for cluster applications. SCC automates cluster configuration decisions based on formal specifications of application behavior and hardware properties.

M. Pitkanen [4] described the design of a highly available grid data storage system. Redundant data is computed using Reed Solomon codes. The level of availability can be chosen for each individual file.

IV. PROPOSED WORK

We propose a new allocation strategy, a combination of replication and erasure [8] coding for data storage in cloud data centers using load balancing based on redundancy mechanisms.

The Algorithm:

Setup (n datacenters)
Allocate m blocks to each dc
Schedule:
Choose a datacenter based on load
For k=1 to n
Check the availability of k^{th} dc suitable for μ
If μ = sflag
Allocate to K
Else
Ealloc (n, μ)
End
// Algorithm for partitioning and choosing a suitable cloud with least cost.
Ealloc (n,μ)
//The output is minimum cost C, The set of the selected clouds H.
1.C←inf;
2.H={}//initially empty.
3.Sort the clouds by S+ μ // Accessibility
4. for m= 1 to n do
A←calculate the availability of G
If A<=Amax then
Mcost←minimamal cost.
If Mcost<C then
H←G.
End

V. EXPERIMENTAL SETUP

The proposed system was developed in java and is tested using Cloudsimulator. The experiment is carried for 10 cloud data centers in cloudsim. Each time when the user wants to place his data in a cloud, the algorithm determines the load and time required to place the data and also evaluates the cloud data center to which the file should be allocated.


```
Starting Cloud...
Total 10 cloud Data Centers are Ready for hosting..
yes
Choose a File to Upload in Cloud Data Center..Arrays & Matrices.txt
Initialising...
Starting CloudSim version 3.0
Datacenter_0 is starting...
Datacenter_1 is starting...
Broker_0 is starting...
Entities started.
0.0: Broker_0: Cloud Resource List received with 2 resource(s)
0.0: Broker_0: Trying to Create VM #0 in Datacenter_0
0.0: Broker_0: Trying to Create VM #1 in Datacenter_0
0.0: Broker_0: Trying to Create VM #2 in Datacenter_0
0.0: Broker_0: Trying to Create VM #3 in Datacenter_0
0.0: Broker_0: Trying to Create VM #4 in Datacenter_0
0.1: Broker_0: VM #0 has been created in Datacenter #2, Host #0
0.1: Broker_0: VM #1 has been created in Datacenter #2, Host #0
0.1: Broker_0: VM #2 has been created in Datacenter #2, Host #0
0.1: Broker_0: VM #3 has been created in Datacenter #2, Host #1
0.1: Broker_0: VM #4 has been created in Datacenter #2, Host #0
0.1: Broker_0: Sending cloudlet 0 to VM #0
0.1: Broker_0: Sending cloudlet 1 to VM #1
```

Fig 1.2 uploading of file by the user

```
General Output

File Acclocated to ..0
Adding: Broker_1
Broker_1 is starting...
200.0: Broker_1: Cloud Resource List received with 2 resource(s)
200.0: Broker_1: Trying to Create VM #100 in Datacenter_0
200.0: Broker_1: Trying to Create VM #101 in Datacenter_0
200.0: Broker_1: Trying to Create VM #102 in Datacenter_0
200.0: Broker_1: Trying to Create VM #103 in Datacenter_0
200.0: Broker_1: Trying to Create VM #104 in Datacenter_0
200.1: Broker_1: VM #100 has been created in Datacenter #2, Host #1
200.1: Broker_1: VM #101 has been created in Datacenter #2, Host #0
200.1: Broker_1: VM #102 has been created in Datacenter #2, Host #1
200.1: Broker_1: VM #103 has been created in Datacenter #2, Host #0
200.1: Broker_1: VM #104 has been created in Datacenter #2, Host #1
200.1: Broker_1: Sending cloudlet 100 to VM #100
200.1: Broker_1: Sending cloudlet 101 to VM #101
200.1: Broker_1: Sending cloudlet 102 to VM #102
200.1: Broker_1: Sending cloudlet 103 to VM #103
200.1: Broker_1: Sending cloudlet 104 to VM #104
200.1: Broker_1: Sending cloudlet 105 to VM #100
200.1: Broker_1: Sending cloudlet 106 to VM #101
200.1: Broker_1: Sending cloudlet 107 to VM #102
```

Fig 1.3 Allocation of a suitable cloud

VI. CONCLUSION

Cloud services are becoming so popular now a day. One of the concern in moving services to cloud is "cost" and secured storage. In this paper we presented a method which uses both replica and "Eruser" techniques in choosing a cloud according to the client needs with a minimum cost.

REFERENCES

[1] Quanlu Zhang; Shenglong Li; Zhenhua Li; Yuanjian Xing; Zhi Yang; Yafei Dai CHARM: A Cost –Efficient Multi-Cloud Data Hosting Scheme with High Availability ,IEEE Transactions on cloud computing ,2015 volume 3 issue 3.

[2] M. Vrable, S. Savage, and G. M. Voelker, "BlueSky: A Cloud-Backed File System for the Enterprise," in FAST. ACM, 2012.

[3] V. Madhyastha, J. C. McCullough, G. Porter, R. Kapoor, S. Savage, A. C. Snoeren, and A. Vahdat, "scc: Cluster Storage Provisioning Informed by Application Characteristics and SLAs," in FAST. ACM,2012.

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

[5] W. Itani, A. Kayssi, A. Chehab, "Privacy as a Service: Privacy- Aware Data Storage and Processing in Cloud Computing Architectures," Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, Dec 2009.

[6] M. Dijk, A. Juels, ”On the Impossibility of Cryptography Alone for Privacy-Preserving Cloud Computing”, HotSec 2010.

[7] P. F. Oliveira, L. Lima, T. T. V. Vinhoza, J. Barros, M. Medard ”Trusted storage over untrusted networks”, IEEE GLOBECOM 2010.

[8] Jun Li and Baochun Li – “Erausre Coding For Cloud Storage Systems”.

[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

