Improving Data Availability by using VPC Strategy in Cloud Environment

Priti Bhute[1], Kumud Kamble[2], Tejashwini Kore[3], Prof. Shravani Jashthi[4]

[1,2,3,4]*Department of Computer Science and Engineering, Wainganga College of Engineering and Management Dongargaon Nagpur, Maharashtra*

---***---

Abstract: *Due to the increasing popularity of cloud computing, emerging concerns users are having to contend with include undesirable internet marauders such as worms, viruses, hackers and financially motivated cyber-terrorists. This paper analyses the basic problem of cloud computing data storage, data security. The cloud computing is playing a very important role to provide effective cost availability, scalability also minimize the response time of applications and make balancing load for cloud. Here we work on Virtual Private Cloud (VPC) and data replication strategy (DRS) to improve data storage in cloud environment. Data replication is a well known technique that consists in storing a multiple copies of data, called replicas, at multiple nodes. It aims to increase data availability, reduce bandwidth consumption and achieve fault tolerance. For example- Amazon web services (AWS), Amazon Microsoft Azure etc.*

Keywords: cloud computing; cloud storage; load balance; VPC strategy, Data replication strategy

1. INTRODUCTION

Cloud computing is the use of a network of remote servers hosted on the internet to store, manage and process data rather than a local server. In simple terms, it means storing or accessing your data over the internet. In distributed system data replication strategy is observed as the common way to enhance availability and reliability of data storage service. Also determines the nodes where replicas are placed the total number of replicas across the clusters is referred to as the replication factor.

Cloud computing is a term, which involves virtualization, distributed computing, networking, software and web services. A cloud consists of several elements such as clients, datacentre and distributed servers. It includes fault tolerance, high availability, scalability, flexibility, reduced overhead for users, reduced cost of ownership, on-demand services etc. The properties of cloud computing in data storage with acceptable efficiency as well as low cost enhanced the accessible capability of data over the internet.

1.1 cloud computing:

Cloud computing refers to the practice of having a shared pool of resources remotely, which can be "rented" out by organizations for their use, rather than using their own resources for the same.

It has many advantages such as:

_You pay only for the resources that you use. Hence, it is highly economical for small-scale purposes.

_It is on-demand and can be easily expanded as per the requirements of the user, thus helping the applications running on it easily scale.

_It can be access from anywhere over the internet irrespective of what device is being used since the servers are present remotely.

_Maintenance of the resources is in the hands of the cloud service provider and the user doesn't need to worry about it.

_Reliability is offered by cloud service providers since they have data centres on multiple sites and can easily transfer the load to another data centres in case of any problem.

Cloud can be broadly categorized into public, private and hybrid.

_In this project we work on private cloud. A private cloud hosts one enterprise as a user. Various departments may be present in the cloud, but all are in the same enterprise. A private cloud also includes provisioning and metering facilities that enable fast deployment and removal where applicable. Private clouds are completely managed and maintained by your organization.

There are many benefits to going with a private cloud model. Most of these benefits centre around your ability to monitor and control what goes on in the cloud environment.

VPC allows the user of the cloud environment to define their private own works may of them like manipulated them, control them, recreate them.

A virtual private cloud is public cloud capability that provide you the ability to define and control isolated virtual network and then deploy cloud resources into those networks. For example: AWS, VM ware, Amazon Microsoft Azure.

Here are some service model which are used in cloud computing which are as follows:
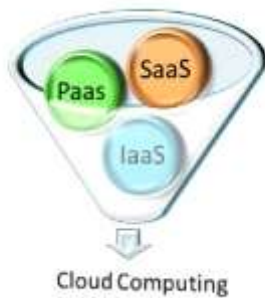

Fig-1.1 Service Models of Cloud Computing

IaaS: cloud base services, pay-as-you-go for services such as storage, networking and virtualization.

PaaS: hardware and software tools available over the internet.

SaaS: software that's available via a third–party over the internet.

2. RELATED WORK

Data replication is one of the most common methodologies in cloud environment to decrease the waiting time as well as bandwidth consumption by creating of replicas in most appropriate site and consequently enhance the data availability. Since, data replication and replica managements is a hot issue for researches in distributed system.

1] Wei et al. [1] presented a cost effective dynamic replication management scheme referred as CDRM. Firstly, they model the availability and replica number relationship and set the number of replication in lowest possible values to satisfy the user availability requirement. If the replica number is lower than the proposed threshold by model, more replicas must be generated in data nodes of cluster. CDRM used from this characteristics as well as blocking probability of data nodes as criteria for replica placement. The authors introduced CDRM with HDFS and concluded that CDRM is able to enhance the performance and load balancing of HDFS as default.

2] Hussein et al. [2] illustrated an Adaptive replication strategy (ARS) in the cloud environment. They studied the availability and file accessibility as affecting parameters to enhance the data file reliability on the base of estimation of the user access to block of each file. Moreover, ARS extends the replication of large-scale various file on different sites by minimization of cost according to the heuristic method. ARS is able to determine the most popular file on the base of evaluation in the recent history of file data access using HLES time series. When, replication factor (i.e., estimation on the base of popularity of file) is lower than specific value, the replication is done. Therefore, employments of heuristic method enable ARS to determine the best threshold during replication process. Simulation results confirm the adaptive strategy improves availability in cloud environment. Consideration of one parameter, i.e., the popularity degree, replica placement is the most restriction of proposed method.

3] Agnesswaran S et al. in [6] proposed a replica management policy that the system maintains data availability level by establishing a model. The method first calculates the number of blocks that need to replicated for each file according to the availability thresholds and then choose the best replica placements node according to the bandwidth and storage cost.

4] Mistarihi H et al. in [8] proposed the replica policy that the system select the best block needing to be copied due to the requested numbers of block life cycle and then put the block copy to the smallest blocking rate node. The above work all concerned on the distributed storage system.

5] Bonvin Nicolas et al. in [10] proposed a cost efficient replication scheme across the data centres that dynamically adopts the number of replica S employed per partition to the query load, while maintaining availability guarantees in case of failures. However, this work does not consider load balancing and real time characteristics of availability.

3. PROPOSED SYSTEM

Usually clouds with its unique characteristics provides the elasticity as well as scalability for user by employment of large and power-consuming data centres and usable strategies in distributed storage system, e.g., GFG, HDFS.
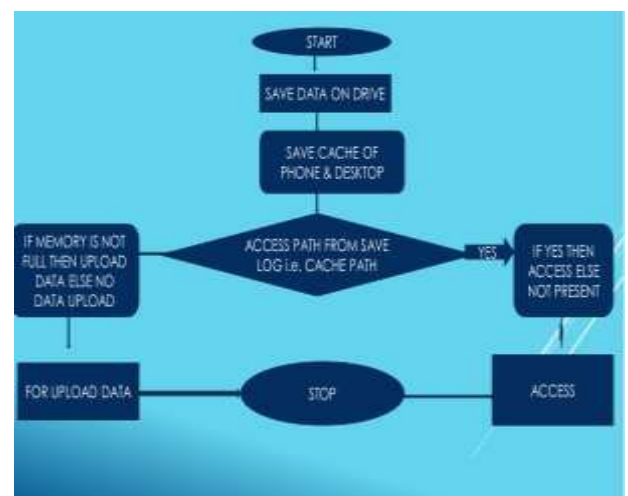

Fig 3.1: Flow diagram for proposed system.

The proposed replication strategy has three different steps which are the identification of data file to replicate, placing new replicas, and replacing replicas. In the first step, it finds the most requested files for replication. In the second step, it selects the best site by consideration of the frequency of requests for replica, the last time the replica was requested, failure probability, centrality factor and storage usage for storing new replica to reduce access time. In the third step, the replacement decision made in order to provide better resource usage. The proposed strategy as certain the importance of valuable replicas based on the number of accesses in future, the availability of the file, the last time the replica was requested, and size of replica. Our proposed algorithm evaluate by cloud simulator and results confirmed the better performance of hybrid replication strategy in terms of mean response time, effective networks usages, replications frequency, degree of imbalance and number of communications.

4. CONCLUSION

Cloud data management is increasingly attracting researcher's attention. This study is a good technique which decreases the waiting time of the user and increases the data availability and provides data security, scalability. Combine replication strategy (CRS) distributes the work load and improve the response time of the user and access the files.

5. FUTURE SCOPE

Cloud computing is having an increasing impact on world of search engine optimization (SEO). Given the importance of SEO to virtually every business with a website, this trend can only grow in 2020 and beyond. Cloud computing will, of course, have a major role to play in the development of Internet of Things, as it relies enormously on machine-to-machine communication.

Of course, there are many advantages to using cloud technology such as the spinning of extra resources and the flexible pay-for-use consumption model.

The cloud is set to become indispensable to businesses and the need of cloud computing solutions grow increasingly complex. In 2020, these trends will be impacting businesses across the world, and it is important for organization to keep up-to-date with their requirements so that they can adjust their cloud architecture as necessary.

REFERANCES

1] Q. Wei, B. Veeravalli, B. Gong, L. Zeng, and D. Feng, "CDRM: A cost effective dynamic replication management scheme for cloud storage cluster," In IEEE International Conference on Cluster Computing, pp. 188-196, 2010.

2] M. Hussein and M.H. Mousa, "A light-weight data replication for cloud data centre environment," International Journal of Engineering and Innovative Technology (IJEIT). Vol. 1.pp. 169-175, 2012.

3] H. Zhang, B. Lin, Z. Liu and Wenzhong Guo, "Data replication placement strategy based on bidding mode for cloud storage cluster." In IEEE web information system and application conference," pp.207-212, 2014.

4] L. Chen and Doan B. Hoang, "Adaptive data replicas management based on active data-centric framework in cloud environment." In IEEE International Conference on High Performance Computing and communication. 2013.

5] A. Rajalakshmi, D. Vijayakumar, and K.G. Srinivasagan, "An Improved dynamic data replica selection and placement in cloud," In International Conference on recent trends in information technology, pp. 1-6, 2014.

6] Agnesswaran S, Janakiram D. Node capability aware replica management for peer-to-peer grids [J]. system man and Cy-berntics part A: system and humans, 2009, 39(4): 807-818.

7] D. W. Sun, G. R. Chang, Sh. Gao, L. Jin, and X. W. wang, "Modeling a dynamic data replication strategy to increase system availability in cloud computing environment," Journal of computer science and technology, vol. 27, No. 2, pp.256-272, 2012.

8] Mistarihi H, Yong C, replica management in distributed file system[J]. International journal of computer science and network security, 2008, 22(8): 22-32.

9] Grossman R L. the case for cloud computing [J]. IT professional, 2009, 11(2): 23-27.

10] Bonvin N, papaioannou T G, Aberer K. Dynamic cost-efficient replication in data clouds[C]// proceedings of the first workshop on automated control for datacenters and clouds. ACM, 2009: 49-56.