Volume 5, Issue 4, 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Replication for Improving Availability & Balancing Load in Cloud Data Centres

Manisha Kalkal*, Sona Malhotra
Department of Computer Science & Engineering
UIET, KUK, Haryana, India

Abstract- This paper focuses on the problem of availability of data on the data nodes and replication process. Cloud computing service providers require a system which can handle a large number of requests at a time. For processing the huge cloud of requests for data access, services need to be highly available. System keeps multiple copies of the blocks of data on different nodes by replication. Load on system may increase when there is high demand of a data is present and one or more sites providing that data fails. Data can easily be recovered from another node in case of one nodes unavailability. Workload division is a method for distributing workloads across all the available nodes by using replication. Brief introduction about the system framework for accessing data is also provided in this paper on basis of study.

Keywords— Cloud Computing, Cloud Data stores, Availability, Data Replication, Workload Balancing

I. INTRODUCTION

Cloud computing is an emerging technology which provides a lot of opportunities for online distribution of resources or services. According to National Institute of Standards and Technology (NIST) cloud computing is a model for enabling universal, well-located and on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with very less management effort [1]. The most effective benefit of using cloud computing is higher availability of services with lower cost and easy scalability. Cloud provides different type of services and resources effectively but some challenges are still present in it. Out of which security concerns, synchronization, scalability, load balancing and replication are important issues.

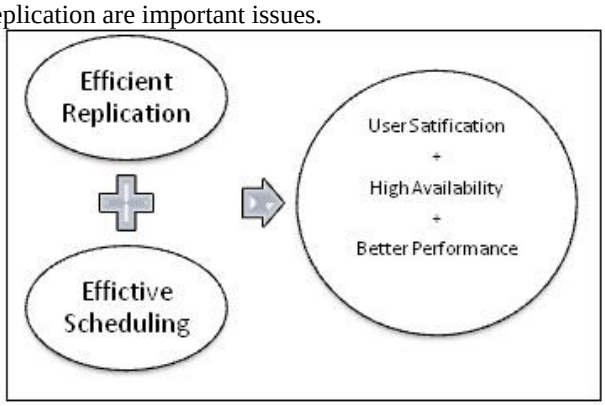

Fig 1 Relationship between Various Aspects

A. Cloud Data Stores

Cloud Storage system, also known as DAAS (Data storage as a service), is the abstract of storage behind an interface where storage can be administered on demand. Cloud data stores works on distributed file systems because of its ability to handle a vast volume of data effectively. Storage can be local or remote. Cloud computing is cost effective, efficient and scalable but managing the load of unpredictable job arrivals is a difficult task. Virtualization is one of the important characteristics supported by cloud data stores. A single physical node can be presented as a set of multiple virtual machines for processing user requests. A set of multiple virtual nodes helps in maintaining the availability as per demand [2].

B. Availability

Data availability means data is accessible whenever it is requested. Availability of data increases with increment in number of replicas of data. But after reaching a specific level of replication, there occurs no improvement in availability. So it is better to find an optimum level of replication. Availability and replication ratio also depends on node failure ratio. If failure probability is high, more number of replicas of that data is required. So if node failure ratio is less, less replica number is required for maximum file availability [3].

C. Data Replication

Data replication means maintaining multiple copies of same data on same server or on different servers. In connection with cloud computing data replication can be said as storing multiple copies of same data on different locations (servers), locally or at remote sites. If data is present at one site only, then it will be very difficult to handle the requests for accessing the data. Server will face a heavy load situation and system performance may degrade.
Also if that site fails, all that data will be lost, this is also a serious concern. Replication is necessary for maintaining the availability, performance level, backing up the data and also for balancing load.

D. Load Balancing

Load balancing is a technology to distribute the workload across multiple resources. Workload balancing is an important aspect of system performance and user satisfaction. Cloud computing can be made more efficient by balancing load. A controller is required for managing the status of load on sites [4].
Need of load balancing is

- To spread the load among a number of machines
- For provide redundancy in case one server fails
- For manage resources effectively & better user satisfaction
- For achieving application response time

II. DATA ACCESS FRAMEWORK FOR CLOUD STORAGE

Clients access storage servers through Cloud Storage Clients whenever they try to access data. User interacts with web server through user interface (as web browser), which further interacts with storage clients for processing. Storage client interacts with name node on storage servers for completing the process.

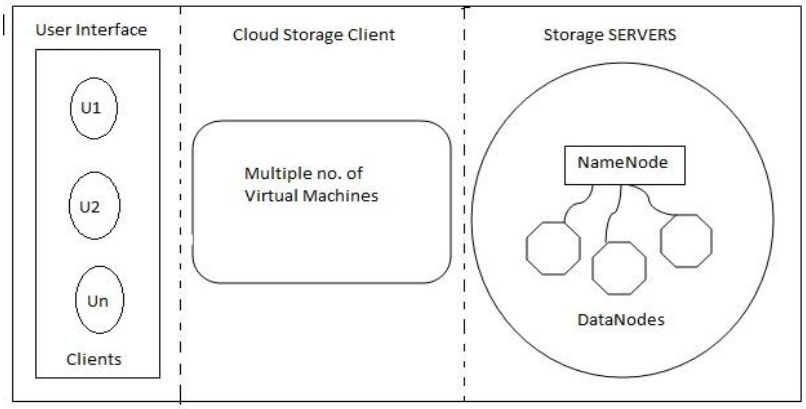

Fig2 Cloud Storage Access Process

One Name Node can handle multiple Data Nodes in a cluster. Name node is main node having complete details about all other nodes (as properties, storage, data present on node) [5].
Whenever request arrives for accessing a specific data, then it is the Name Node which provides details about where to go for that data. If request is for placement of data on server, it provides the detail of nodes most preferable for placement of that data. Job scheduling algorithm and Replica placement algorithms are the two most important decisions which affect overall performance.

III. RELATED WORK

A large number of replication strategies for management of replicas have been proposed in literature. As a result of replication, data replicas are stored on different data nodes for high reliability and availability. Replication factor for each data block and replica placement sites need to be decided at first.
A replication strategy [6] for region based framework based on the demand of files over a geographically distributed Grid environment. Access frequency of each file is calculated and on that basis it is determined that in which region the replicas need to be placed and the number of replicas need to be placed. When a file is created, the access frequency is calculated for each region and replicas are placed in the regions with the large order of the access frequency. Number of requests and the response time are considered as main points for deciding in which site within the region the file has to be placed. Therefore, their strategy increases the data availability and also reduces the number of unnecessary replications.
Another replication strategy proposed on the basis of a dynamic file assignment according to access load on servers [7]. At first, files sorted on basis of size are assigned to disks in a round-robin scheme. It is done to distribute the load of all files evenly across all disks. Then creation and deletion of replicas is performed according to the load of all files and the load on each disk. From the experiments, load balancing can be achieved in an environment where user access patterns change significantly.
Two effective replica reconstruction schemes are proposed to balance the workloads of each Data Node by properly choosing source and destination Data Nodes [8]. Nodes are kept in ring arrangement. One directional ring structure block transfer (biased data transfer) is applied for minimizing the difference of amount of data transfer of each node.

Heuristic scheme and optimization schemes are compared out of which heuristic performed better because it shown better scalability as compared to other.

A Dynamic Data Replication scheme is proposed in [9] to handle dynamic workload situations for better data access. It is very difficult to handle unpredictably changing data access patterns with traditional methods. Algorithm proposed here detects changes in user access pattern and according to that dynamically applies an optimal replication strategy. Dynamic Data Replication scheme can work well for cloud data services as it can effectively handle any pattern of data access.

Replication and various algorithms used for replication are discussed and compared in [10]. Replication is directly related to performance, availability, fault tolerance, load balancing and reliability of system. Servers handle replicas in such a way that all the replicas of same data have latest updated information (Consistency or synchronization of data).It is stated that on the basis of system replication, model will be selected for better results because different algorithms perform better in different situations. Optimistic approach is generally used for replication.

Table I Comparison of Algorithms Used

Title of Paper	Main Consideration	Replication Algorithm Used
Improving Data Availability in Cloud Storage with Efficient Replication[5]	Availability & Storage	Dynamic Replication strategy based on Access Frequency
Effective Replica Reconstruction Schemes at Node Deletion for HDFS[8]	Throughput	Heuristic scheme for replication given better results as compared to traditional
Dynamic Data Replication Scheme in Cloud Computing Environment[9]	Access Time & Network Usage	LRU: In case of sequential access patterns LDRS: In case of random access patterns

IV. CONCLUSION

Replication can be used for maintaining availability in presence any load conditions or failure situations. By improving the technique of replication, performance and availability of system can be improved. But excessive replication can also adverse effects like high storage cost or degradation in systems overall performance due to excessive use of bandwidth. Dynamic data replication scheme is better to use because it can understand random user data access patterns. Round robin scheduling can provide better results in case when system is in ideal state. It is generally used when requests are of similar nature and distributed equally. A selective replication method must be applied for overall better results. We need a better scheduling algorithm for jobs and an effective replication strategy for data blocks to improve overall system availability.

REFERENCES

[1] Peter Mell, Timothy Grance(2011, Sept), NIST Definition of cloud computing, *NIST special publications 800-145*

[2] Furht, A. Escalante, *Handbook of Cloud Computing*, ISBN 9781-4419-6523-3, Springer Science + Business Media, 2010

[3] Sage A.W., Scott A.B., Ethan, Darrel (2006, Nov), Ceph: A Scalable, High Performance Distributed File System, *Proceedings of 7th Conference on Operating System Design And Implementation* (OSDI 06)

[4] Pragathi M, swapna Addamani, Venkata Ravana Nayak(2014, April), Resource Monitoring and Workload Balancing Model for Public Cloud, *International Journal Of Science And Research Publications*, vol. 4, Issue 4

[5] Julia Myint, Thinn Thu Naing(2012, Feb), Improving Data Availability in Cloud Storage with Efficient Replication, *10th International Conference on Computer Application, ICCA* , Yangon, Myanmar

[6] K.Sashi, Antony S.Thanamani (2010), A New Replica Creation and Placement Algorithm for Data Grid Environment, *International , Conference on Data Storage and Data Engineering*, pp.265-269

[7] J. Tjioe, R. Widjaja, A.Lee, T.Xie (2009), DORA:A Dynamic File Assignment Strategy with Replication, *International Conference on Parallel Processing*

[8] Asami H., Atsuko T., Hidemoto N., Masato Oguchi (2014, May), A Study of Effective Replica Reconstruction Schemes at Node Deletion for HDFS, IEEE, 978-1-4799-2784

[9] M. Jeon, K. Lim, H.Ahn, B. Lee (2012, Sept), Dynamic Data Replication Scheme in Cloud Computing Environment, *IEEE*

[10] Nikhil Chaturvedi, Dinesh Chandra (2012, May), Analysis of Replication and Replication Algorithms in Distributed System, *International Journal of Advanced Research in Computer Science and Software Engineering* , Volume 2, Issue 5, Available online at: www.ijarcsse.com

