Adaptive Data Replication in Cloud to Improve Availability of Popular Data

[1] T.K.Akalya Maheswari,[2] S.Deepika,[3] S.Karunakaran
Department of Computer Technology (UG)
Kongu Engineering College
Perundurai, Erode.

Abstract— *Data replication is a method to improve the performance of the data access in distributed systems. Dynamic replication is a replication that adapts replication configuration with the change of user behavior during the time to ensure the benefits of replication. To improve the system availability, replicate the popular data to multiple suitable locations, as users can access the data from a nearby site. It decide a reasonable number and right locations for replicas has become a challenge in the cloud computing. A dynamic data replication strategy suitable for distributed computing environments. It includes, analyzing and modeling the relationship between system availability and the number of replicas, evaluating and identifying the popular data and triggering a replication operation when the popularity data passes a dynamic threshold, calculating a suitable number of copies to meet a reasonable system byte effective rate requirement and placing replicas among data nodes in a balanced way, designing the dynamic data replication algorithm in a cloud. As a result of the proposed method, increase the availability, performance, reduce user waiting time and also reduce the execution time of the system.*

Keywords- *cloud computing, dynamic data replication, system availability, fault tolerance and temporal locality.*

I. INTRODUCTION

Cloud computing is large scale distributed computing paradigm. In cloud, an application is accessible from anywhere and anytime becomes infinite for all intents and purposes. The users or clients can access the powerful applications, platforms and services delivered over internet.

Cloud computing services are delivered on demand to external customers over high-speed internet with the "X as a service (XaaS)" computing architecture. The cloud computing refers to the application delivered via the software as a services (SaaS) via an infrastructure as a service (IaaS) and via platform as a service (PaaS) and also via hardware as a service (HaaS).

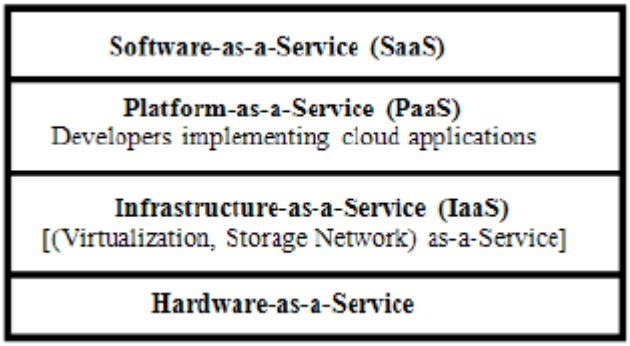

Fig. 1 cloud architecture

With cloud computing growing in IT enterprise. Cloud computing is a paradigm that changes the idea of local computers to a cloud of computers that contains server pool providing different services to many clients at the same time. In cloud computing there are multiple copies from the same application, all copies are updated regularly. Clients can be sharing not only the software but also hardware without being aware of the sharing methods and techniques.

Basically cloud can be defined by three deployment models. *Private cloud:* It's known as internal clouds. Private clouds are mainly used by a single organization. A private cloud it may be built and managed by the organization or by external provider. A private cloud provides the highest degree of control over the reliability, security and performance. *Public cloud:* A cloud in which service providers offer their resources as services to the general public. The public clouds offer several key benefits to service providers. That includes no initial capital investment on infrastructure and shifting of risks to infrastructure providers. *Community cloud:* The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns. *Hybrid cloud:* A hybrid cloud is a combination of public and private cloud models that tries to address the limitations of each approach. a hybrid cloud, a part of the service infrastructure runs in private clouds while the remaining part runs in public cloud. A hybrid cloud offer more flexibility than both public and private cloud.

The cloud can differ from other large-scale distributed computing platforms can be summarized as rapid elasticity, broad network access, measured services, on-demand self-service and resource pooling. And also cloud can be a highly available, much more elastic and reliable more cheaper and scalable computing environment

compared to supercomputers, other large-scale distributed computing environments and grids.

Cloud computing having some benefits is reduction in upfront capital expenditure on hardware and software deployment. Consumption is usually billed on a utility (like phone bills) or subscription (like magazines) model, Location independence, so long as there is access to the internet and allows the enterprise to focus on its core business.

In order to meet the high availability, high fault tolerance and high efficiency requirement, it is necessary to dynamically adjust the popular data files, the number of data replicas and the sites to place the new replicas according to the current cloud environments. In order to achieve the dynamic data replication, there are three important problems that must be solved. 1) Which data should be replicated and when to replicate in the cloud systems to meet the users' requirements on waiting time reduction and data access speeding up are important issues for further research, as the wrongly selected and too early replicated data will not reduce the waiting time or speed up data access. 2) How many suitable new replicas should be created in the cloud to meet a reasonable system availability requirement is another important issue to be thoroughly investigated. With the number of new replicas increasing, the system maintenance cost will significantly increase too many replicas are not increase availability and bring unnecessary spending instead. 3) Where the new replicas should be placed to meet the system task successful execution rate and bandwidth consumption requirements is also an important issue.

Our work is originally motivated by the fact that a more recently accessed data will be accessed again in the near future according to the current data access pattern, which is called temporal locality. With the fact of temporal locality, a popular data is determined by analyzing the users' access to the data. When the popularity of the data passes a dynamic threshold, the replication operation will be triggered. The number of replicas will be determined based on the system availability and failure probability. New replica will be created on near-by locations for users who generate the most requests for the data.

II. Related Work

This section presents two broad categories of related work. The first category discusses cloud data storage, and the second category presents the related work to the cloud data replication.

A. Cloud Data Storage

Cloud computing technology moved computation and data storage away from the end user and onto servers located in data centers, thereby relieving users of the burdens of application provisioning and management. As a result, software can then be thought of as purely a service that is delivered and consumed over the Internet, offering users the flexibility to choose applications on-demand and allowing providers to scale out their capacity accordingly. Many large institutions have set up data centers and cloud computing platforms, such as Google, Amazon, IBM. Compared with traditional large scale storage systems, the clouds which are sensitive to workloads and user behaviours focus on providing and publishing storage service on Internet.

The key components of the cloud are distributed file systems, such as The Google File System GFS, the Hadoop distributed file system HDFS. In the GFS [6], there are three components a single master server, multiple clients and multiple chunk servers. Files are stripped into one or many fixed size chunks, and these chunks are stored in the data centers, which are managed by the chunk servers. Chunks are stored in plain Linux files which are replicated on multiple nodes to provide high-availability and improve performance. The master server maintains all the metadata of the file system including the namespace, the access control information mapping from files to chunks and the current locations of chunks. Clients interact with the master for metadata operations, but all data available in communication goes directly to the chunk servers. Secondary name servers provide backup for the master node.

In a multi-cluster system, each cluster is a complete GFS cluster and with its own master, and each master maintains the metadata of its own file system. Different masters can share the metadata by the namespace, which describes how the log data is partitioned across multiple clusters. Compared with a single cluster, in a multi-cluster system, the performance of the cloud system and the size of the cloud data storage can be improved significantly. The mechanism of HDFS is similar to that of GFS, but it is light-weighted and open-source [8]. HDFS also follows a master/slave architecture which consists of a single master server that manages the distributed file system namespace and regulates access to files by clients called the Name node. In addition, there are multiple data nodes, one per node in the cluster, which manages the disk storage attached to the nodes and assigned to Hadoop. The Name node determines the mapping of blocks to data nodes.

B. Cloud Data Replication

Replication technology is one of the useful techniques in distributed systems for improving availability and reliability. In Cloud computing, replication is used for reducing user waiting time, increasing data availability and minimizing cloud system bandwidth consumption by offering the user multiple replicas of a specific service on different nodes.

Data replication can be classified into two categories: static replication [1] and dynamic replication algorithms [3, 4, 7]. In a static replication, the number of replicas and their locations are predetermined. On the other hand, dynamic replication dynamically creates and deletes replicas according to changing environment load conditions. There has been an interesting number of works for data replication in the Cloud computing. For example, in [1], a static distributed cloud data replication algorithm is proposed.

In the GFS, a single master considers three factors when making decisions on data chunk replications: 1) to place the new replicas on chunk servers with below-average disk space utilization; 2) to limit the number of "recent" creations on each chunk server; 3) to spread replicas of a chunk across racks. A data chuck is replicated when the number of replicas falls below a limit specified by the users.

The differences between the mentioned replication algorithms and our proposed strategy lie in the following aspects. 1) A heuristic is proposed based on a formal model that describes the relationship between the data files availability and the number of replicas. 2) The popular data is identified according to the history of the user access to the data. When the popularity of a data file passes a dynamic threshold, the replication operation will be activated. 3) Replicas are placed among data nodes in a balanced way.

III. PRPPOSED METHOD

The paper proposes the dynamic data replication strategy, that strategy solves the above mentioned problems and explained that strategies as follows:

A. Dynamic Data Replication

The proposed dynamic data replication has three important phases: 1) which data file should be replicated and when to replicate in the cloud system to meet users' requirements such as waiting time reduction and data access speeding up; 2) how many suitable new replicas should be created in the cloud system to meet a given availability requirement; 3) where the new replicas should be placed to meet the system task successful execution rate and bandwidth consumption requirements. The proposed architecture typically consists of the replica selector, replica manager replica broker and data centers, as shown in Fig. 2.

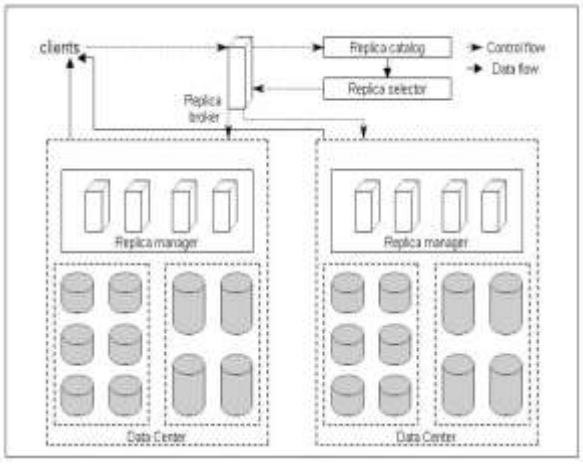

Fig. 2 The Cloud Data Server Architecture

In this paper proposes the dynamic data replication strategies, the proposed work has the following components that are described as follows:

a) Replica Broker

User sends the task into replica broker through the user interface. The replica broker schedules them to the appropriate cloud data server sites. In replica broker, a task characterized by the task identification, generation rate, deadline time and the number of required file of the task respectively. For simplicity, it assumes that the tasks are non-preemptable and non-interruptable.

It means that a task cannot be broken into smaller sub-tasks and it has to be executed as a whole on a single processor with given resources. In addition, a task starts its execution on a processor, that task cannot be interrupted and that task occupies the processor until its execution completes successfully or a failure occurs. The replica broker schedules the user task based on the replica selector information. The replica broker sends control to the replica manager in the appropriate cloud data server site.

b) Replica Catalog

Replica catalog maintains the number of replicas in the data center, availability and number of access the each file at certain time period. When each site store a new replica, it send a file register request to replica catalog and then replica catalog add this site to the list of sites that holds the replica.

c) Replica Selector

Replica selector is identifying the popular data based on the dynamic threshold value. The popular data identified using the starting time, present time, number of access at certain time period and time based forgetting function. Replica selector uses the dynamic data replication strategy. The replication strategy satisfies the three considerations of data replication.

The three considerations are which and when data should be replicate, decide how many new replicas create, and where the new replicas placed. And replica selector sends replication information to the replica broker.

d) Data Center

Data center composed of n data nodes, which are running virtual machines on physical machine. The data node characterized by the data node identifier, request arrival rate, average service time, data failure probability and network bandwidth. A data file which is stripped into number of blocks according to its length. The popular data identified by the replica selector. If the user requested data place to the users nearby site. And the users get the data directly from the data storage server.

e) Replica Manager

Replica manager hold the general information about the replica location in data center. When the replica broker send the user task to the replica manager, it point out the available replicas location in the data center. The used can then directly get the data from the data storage server.

Typically consists of different tiers of data centers with different regions and sizes. The super data centers in tier 0 will handle the data analysis in the intra domain and

exchange data information among the inter domains. The main data centers are in tier 1, ordinary data centers are in tier 2, and users are in tier 3. The architecture minimizes the data access time and network load by creating and spreading replicas from the super data centers to main data centers, or to ordinary data centers. The super data centers periodically collect and broadcast the global information.

IV. Result Analysis

The part of the proposed work has been simulated in CloudSim 3.0. The CloudSim is a framework for modeling and simulation of cloud computing infrastructures and services. Main features of cloudsim are it supports for modeling and simulation of large scale Cloud computing data centers, it support for modeling and simulation of virtualized server hosts, with customizable policy for provisioning host resources to virtual machines and support for modeling and simulation of energy-aware computational resources.

The figure 3 describes the creation of virtual machines and data centers. First starts the replication broker that broker receives the cloud resource list and broker provided in cloudsim only submits a list of virtual machines to be created and schedules cloudlets sequently on them. It responsible for allocating these virtual machines and resource consumption. Each virtual machine has its own configuration that consists of its hypervisor. Data center behaves like an Infrastructure as a Service provider, it receives requests for virtual machines from brokers and creates virtual machines in hosts.

Fig.3 Creation of Virtual Machines and Data Centers

The figure 4 describes the replicated data in data center. Replication broker replicated the data in to data center. The availability increases in the system, reduce bandwidth consumption and also reduce the user waiting time. Dynamically replicated data eliminated from the data center, when that data having less replica factor value.

Fig.4 Replicate Data in Data Center

The figure 5 represents increasing number of requests and also increasing replicas.

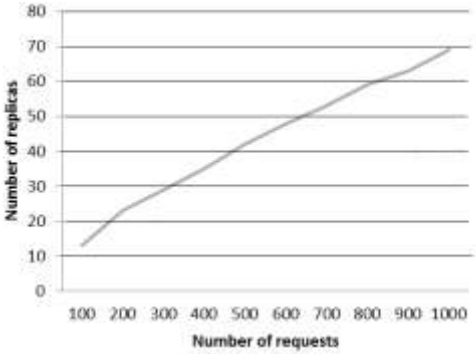

Fig.5 Increasing Number of Requests corresponding Increase Replicas

The figure 6 represents the identification of popular data. The popular data placed in datacenter DC1, then replicate the data from the datacenter DC1 and placed the data in nearby user site and within the region.

Fig.6 Identifying the popular Data

Table.1 Response Time

Usertase	Avg (ms)	Min (ms)	Max (ms)
UB10	79.283	48.54	99.511
UB11	136.227	53.273	177.284
UB12	83.134	49.759	106.776
UB1	74.447	52.215	93.635
UB2	89.03	68.786	99.274
UB3	96.282	48.525	80.757
UB4	82.006	51.362	102.77
UB5	136.566	49.039	176.791
UB6	82.759	47.76	102.756
UB7	69.061	43.91	85.009
UB8	70.776	50.527	79.527
UB9	60.666	49.479	70.018

Simulation Complete

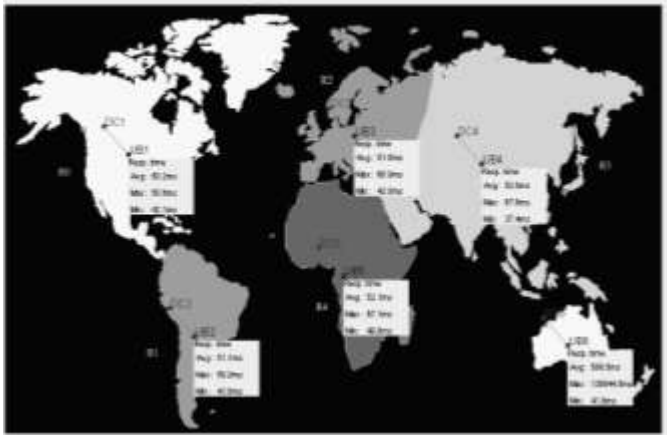

Fig. 7 Replicated the Popular Data in nearby Site

The figure 7 represents replicated the Popular Data in nearby Site. Then increase availability because reduce the repose time, processing time, datacenter load and reduce the virtual machine and data transfer cost. The table 1 displaying the response time of user request. The figure 6 and 7 represents the response time for each user.

V. Conclution And Future Work

This paper proposes dynamic replication strategy in the cloud environment. The strategy investigates the availability and efficient access of each file in the data center, and studies how to improve the reliability of the data files based on prediction of the user access to the blocks of each file. The strategy identifies the files which are popular file for replication based on analyzing the recent history of the data access to the files using HLES time series. Once a replication factor based on the popularity of the files is less than a specific threshold, the replication signal will be triggered. Hence, the strategy identifies the best replication location based on a heuristic search for the best replication factor of each file. The experimental evaluation demonstrates the efficiency of the proposed dynamic replication strategy in the cloud environment.

In the future work, further reducing the user waiting time, speeding up data access, and further increasing data availability. In addition, the replication strategy will be deployed and tested on a real cloud computing platform. It is also planned to make data replication strategy as a part of cloud computing services to satisfy the special demands of cloud computing, and finally, to develop a complete dynamic data replication framework based on the proposed model.

References

[1] Armbrust M, Fox A, Grffith R, Joseph A D, Katz R, Konwin-ski A, Lee G, Patterson D, Rabkin A, Stoica I, Zaharia M, "A view of cloud computing. Communications of the ACM," Vol.53, Issue.4. pp.50-58, 2010.

[2] Bhaskar Prasad Rimal, Eunmi Choi and Ian Lumb , "A Taxonomy and Survey of Cloud Computing System," Fifth Internationational Joint Conference on INC, IMS and IDC, pp.44-51, 2009.

[3] Da-Wei sun, Gui-Ran Chang and Shang Gao , "Modeling a Dynamic Data Replication Strategy to increase System Availability in cloud computing environment," Journal of computer science and technology, Vol.27, Issue.2, pp.256-271, 2012.

[4] Mohamed-K Hussein, "A Light-Weight Data Replication for Cloud Data Centers Environment," International Journal of Engineering and Innovative Technology, Vol.1, Issue.6, pp.196-175, 2012.

[5] Qi Zhang , Lu Cheng and Raouf Boutaba, "Cloud Computing:State-of-art and research challenges," J Internet Serv Appl. pp.1-18, 2010.

[6] Sanjay Ghemawat, Sharma.A.K and Vishnu Swaroop, "Google File System," In proceedings of the 19th ACM podium on operating system principles, pp.29-43, 2003.

[7] Sashi.K, "Dynamic Replica Management for data Grid," International Journal of Engineering and Technology, Vol.2, No.4, pp.329-333, 2010.

[8] WangDi, Konstantin Shvachko, Harirong, Sanjoy Radia and Robert Chansler, "Hadoop Distributed File System," sun Microsystems, pp.1-10, 2010.

[9] Somayeh Abdi and Somayeh Mohamadi, "The Impact of Data Replication on Job Scheduling Performance in Hierarical Data Grid" International Journal on applications of graph theory in wireless a hoc networks and sensor networks, Vol.2, No.3, pp.15-25, 2011.

[10] Sharrukh Zaman and Daniel Grosu, "A Distributed Algorithm for the replication Placement Problem" IEEE Transactions on Parallel and Distributed Systems, Vol.22, No.9, pp.1455-1468, 2011.

[11] Sepahvand.R, Horri.A and Dastghaibyfrad.Gh, "Replication and Scheduling Methods based on Prediction in Data Grid" Australian Journal of basic and Applied Sciences, Vol.5, Issue.11, pp.1485-1496, 2011.

[12] Sanjay kumar Tiwari, "Issues in Replicated data for Distributed Real-Time database System" International Journal of Computer Science and Informational Technologies, Vol.2, Issue.4, pp.1364-1371, 2011.

[13] Leyli Mohammad Khanli, Ayaz Isazadeh and Tahmuras Shishavan.S, "PHFS: A Dynamic replication method, to decrease access latency in the multi-tire data grid" Journal on future generation computer systems, Vol.27, issue.3, pp.233-244, 2010.