

Public Cloud Server partition based on location and Load

Megha BIRTHARE*¹, Prof. Rasna Sharma²

¹Research Scholar, CIIT, Indore, M.P, India.

²Assistant Professor, CIIT, Indore, M.P, India.

*Department of Computer Science & Engineering.

Abstract:

With growth of cloud computing load balancing is important impact on performance. Cloud computing efficiency depends on good load balancer. Many type of situation occur that time cloud partitioning is done by load balancer. Different type of situation needed different type of strategies for public cloud partitioning using load balancer. In this paper we work on, partition of public cloud using two type of situation first is load status evaluation and second is cloud division rules. Load status evaluation is measure in number of cloudlets arrives at datacenter and cloud divisions rules are based on cloudlet come from which geographical location. On the basis of geographical location we partition public cloud and improve performance of load balancing in cloud computing. We implement proposed system with help of cloudsim3.0 simulator.

Keywords: Load Balancing, Public Cloud, Cloud Partition, Round Robin.

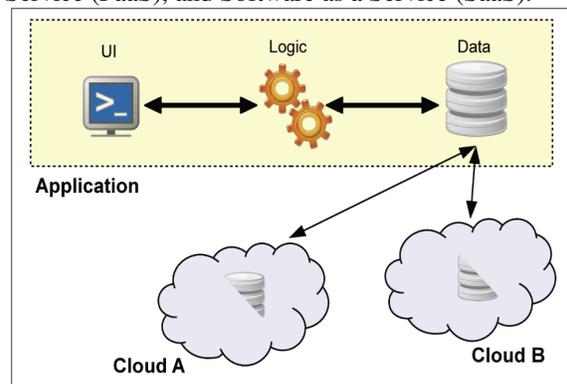
1. Introduction:

Cloud Computing has become the most viable solution for the problems that are computation intensive. Cloud provides the method of sharing resources and services to user on demand. Virtualized resource and services can be used without the knowledge of geographical differences. Cloud computing provides mechanisms that concentrate on run time demand of computing resources, like storage, availability, software etc.

The available Cloud platforms distinguish among the service type, the cost, the Quality of Service (QoS) as well as performance. This fact brings Cloud customers the flexibility of freely selecting target architecture from broad range of Cloud platforms. However at the same time, this raises the issue of the interoperability among the different Clouds [13].

Development of efficient service provisioning policies is the major issues in Cloud research. Modern Clouds exists

in an open world characterized by constant changes occurring autonomously and unpredictably. In this context, game theoretic methods allow in-depth analytical understanding of the service-provisioning problem [2]. The cloud computing mainly offers three types of services viz: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).



In a cloud-model there are four main participants:

- **Cloud Provider:** A cloud provider (service provider) is an entity that is responsible for everything required for making a cloud service available.
- **Cloud Consumer:** A cloud consumer is either a cloud service owner or a cloud service consumer. Cloud service owner is the individual or organization who subscribes for a cloud service. If there is any charge associated with the service, the cloud service owner will be responsible for the bills. Cloud service consumer is an individual or application who accesses a cloud service.
- **Cloud Broker:** A cloud broker is an entity that mediates between cloud providers and cloud consumers. The goal of a service broker is to provide the cloud consumer a service that is more suitable for its needs. This can be done by simplifying and improving the service and

contract, aggregating multiple cloud services or providing value-added services. One can consider cloud brokers as a special cloud provider.

- Cloud Auditor: A cloud auditor is an independent party who examines a cloud service stack to provide an assessment on security, privacy and availability level of the corresponding cloud services and ensures that the corresponding SLAs (Service Level Agreement) are fulfilled. The details and scope of auditing process is normally specified in the service contract

2. Related work:

[1]. Gaochao Xu, Junjie Pang, and Xiaodong Fu “ A Load Balancing Model Based on Cloud Partitioning for the Public Cloud”[*Tsinghua Science and Technology, February* 2013]. Load balancing in the cloud computing environment has an important impact on the performance. Good load balancing makes cloud computing more efficient and improves user satisfaction. This article introduces a better load balance model for the public cloud based on the cloud-partitioning concept with a switch mechanism to choose different strategies for different situations.

[2]. Abhijeet G Purohit et. ”Load balancing in public cloud by division of cloud based on the geographical location”. Load balancing is a method of controlling the traffic in a cloud environment. Cloud applications look for resources for execution. The resources can be storage, processing, bandwidth, etc. Allocation of these resources efficiently to all the competing jobs is called as load balancing. In this paper, we describe load balancing in a public cloud by partitioning the cloud into several sub-clouds. This division of public cloud into several sub-clouds is done based on the geographical location. In this approach we use a central controlling system that monitors all the sub clouds. Here, every sub cloud has a balancer system which monitors the resources in its sub cloud and allocates the available resources to the competing jobs. These balancer systems also communicate with the central controlling system about the status of the respective sub cloud. Based on this information the central controlling system selects the optimal sub cloud.

[3]. Ms. Shilpa D. More et. ”Reviews of Load Balancing Based on Partitioning in Cloud Computing” [International Journal of Computer Science and Information

Technologies, Vol. 5 (3) , 2014]. Load Balancing Model Based on Cloud Partitioning for the Public Cloud environment has an important impact on the performance of network load. A cloud computing system, which does not use load balancing, has numerous drawbacks. Now days the usage of Internet and related resources has increased widely. Due to this there is tremendous increase in workload. So there is uneven distribution of this workload, which results in server overloading and may crash. In such systems the resources are not optimally used. Due to this the performance degrades and efficiency reduces. Cloud computing efficient and improves user satisfaction. This article introduces a better load balance model for public cloud based on the cloud-partitioning concept with a switch mechanism to choose different strategies for different situations. The algorithm applies the game theory for load balancing strategy to improve the efficiency in the public cloud environment.

3. Existing System:

Since the job arrival pattern is not predictable and the capacities of each node in the cloud differ, for load balancing problem, workload control is crucial to improve system performance and maintain stability. Load balancing schemes depending on whether the system dynamics are important can be either static and dynamic. Static schemes do not use the system information and are less complex while dynamic schemes will bring additional costs for the system but can change as the system status changes. A dynamic scheme is used here for its flexibility.

- Cloud computing environment is a very complex problem with load balancing receiving.
- The job arrival pattern is not predictable and the capacities of each node in the cloud differ, for load balancing problem, workload control is crucial to improve system performance and maintain stability.

4. Load Balancing Algorithm:

There are many simple load balancing algorithm methods such as the First Come First Served (FCFS), Round Robin algorithm, Equally spread current execution algorithm and Throttled algorithm. The FCFS and Throttled algorithms are used here for their simplicity and also they provide good response time compared to other algorithms.

5. Proposed System:

When the cloud partition is idle, many computing resources are available and relatively few jobs are arriving. In this situation, this cloud partition has the ability to process jobs as quickly as possible so a simple load

balancing method can be used.

There are many simple load balance algorithm methods such as the Random algorithm, the Weight Round Robin, and the Dynamic Round Robin[12]. The Round Robin algorithm is used here for its simplicity.

Throttled Algorithm:

1. *The Main Controller (Admin) maintains an index table of job requests.*
2. *The job requests are stored in the table based on the arrival time.*
3. *The Main Controller (Admin) scans the index table from top to bottom.*
4. *The Main Controller (Admin) grants the permission to post jobs and changes the REQUEST_NEED flag to GRANTED.*
5. *The HR receives the response to the request sent and then posts jobs by providing details about the interview.*
6. *In this way only one job interview details is posted by a company at a time and if a Company HR wants to post another job then he should send job request again.*

First Come First Served

1. *Main Controller (Admin) maintains an index table of job requests.*
2. *The job requests are stored in the table on the basis of their arrival time. The Main Controller (Admin) scans the index table from top to bottom.*
3. *The first job request according to the arrival time is allocated the grant by the Main Controller (Admin).*
4. *The HR receives the response to the request*

The Round Robin algorithm is one of the simplest load balancing algorithms, which passes each new request to the next server in the queue. The algorithm does not record the status of each connection so it has no status information. In the regular Round Robin algorithm, every node has an equal opportunity to be chosen. However, in a public cloud, the configuration and the performance of each node will be not the same; thus, this method may overload some nodes. Thus an improved Round Robin algorithm is used, which called "Round Robin based on the load degree evaluation"

(1) Cloud division rules: Cloud division is not a simple problem. Thus, the framework needs a detailed cloud division methodology. For example, nodes in a cluster may be far from other nodes or there will be some clusters in the same geographic area that are still far apart. The division rule should simply be based on the geographic location.

(2) How to set the refresh period for data statistics analysis, the main controller and the cloud partition balancers need to refresh the information at a fixed period. If the period is too short, the high frequency will influence the system performance. If the period is too much long, the information will be too old to make good decision. Thus, tests and statistical tools are needed to set reasonable refresh periods.

(3) A load status evaluation: A good algorithm is needed to set Load degree high and Load degree low, and the evaluation mechanism comprehensive.

There are several cloud computing categories with this work focused on a public cloud. A public cloud is based on the standard cloud computing model, with service provided by a service provider [10]. A large public cloud will include many nodes and the nodes in different geographical locations. Cloud partitioning is used to manage this large cloud. A cloud partition is a subarea of the public cloud with divisions based on the geographic locations.

When job i arrives at the system, the main controller (Admin) decides to which partition the job should be assigned. If this is the last updated job, then the job is assigned to Partition1. If it is an upcoming job, then it is assigned to Partition2. If it's a currently running job then it

is assigned to Partition3. If it is an outdated job then it is assigned to Partition4.

5:Result:

6. Conclusion:

The several strategies lack efficient scheduling and load balancing resource allocation techniques leading to increased operational cost and give less customer satisfaction. Load balancing in the cloud-computing environment has an important impact on the performance. Good load balancing makes cloud computing more efficient and improves user satisfaction. In this paper we have proposed a better load balance model for the Job Seeker's Web Portal based on the cloud-partitioning concept with a switch mechanism to choose different strategies for different situations. Thus, this model divides the public cloud into several cloud partitions. When the environment is very large and complex, these divisions simplify the load balancing. The cloud has a main controller that chooses the suitable partitions for arriving jobs based on arrival date. Thus with cloud partitioning concept it is possible to provide good load balancing and hence improving the overall performance of cloud environment and user satisfaction.

7. References:

[1].Gaochao Xu, Junjie Pang, and Xiaodong Fu “ A Load Balancing Model Based on Cloud Partitioning for the Public Cloud”[*Tsinghua Science and Technology*, February 2013].

[2]. Abhijeet G Purohit et. ”Load balancing in public cloud by division of cloud based on the geographical location”[*International Journal of Research in Engineering and Technology*] Volume:03/2014 <http://www.ijret.org>

[3]. Ms.Shilpa D.More et.”Reviews of Load Balancing Based on Partitioning in Cloud Computing” [*International Journal of Computer Science and Information Technologies*, Vol. 5 (3) , 2014].

8. Yogita, DurgaToshniwala, “A Framework for Outlier Detection in Evolving Data Streams by Weighting Attributes in Clustering,” 2ndInternational Conference on Communication, Computing & Security, pp. 214–222, (ICCCS-2012).

9. Yogita,DurgaToshniwal, “Unsupervised Outlier Detection in Streaming Data Using Weighted Clustering,” *World Academy of Science, Engineering and Technology*, Vol:6, Nov 2012.

10. F. Angiull, F. Fassetti, “Distance-based outlier queries in data streams: the novel task and algorithms,” *Data Mining and Knowledge Discovery*, 20(2),2010, pp. 290–324.

11. DimitriosGeorgiadis, Maria Kontaki, AnastasiosGounaris, Apostolos Papadopoulos, Kostas Tsichlas, YannisManolopoulos, “Continuous