International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-4, Issue-5)

Research Article | **May 2015**

Study on Cloud Computing and Different Load Balancing Algorithms in Cloud Computing

Prof. Bhavani. S, Ankit Hatwal, Utsav Mittal
SITE, VIT
Vellore, Tamil Nadu, India

***Abstract*—**

Today most of the organizations are adopting cloud computing techniques as it offers various services like pay as you go, massive scalability, unlimited data storage, elasticity, multitenancy, virtualization and many applications as a utility over the internet through SOA(Service Oriented Architecture) based architecture. In spite of these services there are numerous issues in the cloud and load balancing is one of them. To balance the load in a cloud, the workload must be distributed to nodes or computers or resources. By balancing the load we can reduce response time and optimize the resource utilization. There are various static and dynamic load balancing algorithms which try to solve several workload issues. In this paper, we made a study on cloud computing and suitable algorithms for load balancing such as round robin scheduling, MapReduce algorithm, ACO, and honeybee.

Keywords*— *Cloud computing, load balancing, load balancing algorithm, ACO, MapReduce, Honey Bee, Round Robin

I. INTRODUCTION

Cloud computing is a technique of using thousands of computers and servers provided by the service provider that are hosted on the internet for storing, managing and processing of the data rather than storing on the local computer. The technique and the infrastructure surrounding the cloud is invisible to the users. Cloud comprises of multiple companies, various server providers, different servers and multiple networks. There are basically three layers in cloud architecture that provides different services based on the consumer requirement. These layers are also called as service layer which consist of Software as a service (SaaS), Platform as a service (PaaS) and Infrastructure as a service (IaaS). Cloud infrastructure can be operated in one of the following deployment models: public cloud, private cloud, community cloud or hybrid cloud. The differences are generally based on how solely the cloud resources are made available to a cloud consumer.

Load balancing is a technique of distributing the workload on a single node to the multiple nodes. Here nodes can be virtual computers, virtual servers or virtual networks. Whenever work load increases on a single node, different problems may arise which degrade the service performance provided to customers. To overcome this problem load is distributed to the multiple nodes instead of a single node at a same time that increases the reliability through redundancy. The main aim of load balancing is to make optimum use of the resources with reduce in response time and avoiding the overload in a single node which in-turn provides a reliable services to the customers. There are two types of load balancing algorithms static and dynamic load balancing algorithms. Static load balancing algorithms balance the load prior to the execution and are mostly appropriate to the homogeneous and stable environments. They are not stable for the environment where the attributes changes dynamically during execution time. Dynamic load balancing algorithm balance the load prior as well as during the execution time [1].

In this paper we present a study on different types load balancing algorithm that are well suited to the cloud computing environment. Round robin scheduling, MapReduce, ACO and honey bee are the load balancing algorithm that are studied.

II. CLOUD ARCHITECTURE

There are two kinds of cloud, one is a single-site cloud and another one is geographical distributed cloud. Single-site cloud which is also known as datacenter consist of computer nodes which are grouped into racks, switches that connect the racks, a network topology, storage node and software services. A geographical distributed cloud consist of multiple single-site cloud and each site with probably different structure and services. Cloud architecture consist of five main components i.e. Cloud consumer, Cloud provider, Cloud auditor, Cloud broker and Cloud carrier as shown in fig 1 [2].

A person or the organization that uses the cloud service provided by service provider is known as cloud consumer. Cloud provider can be a person or an organization that provides the service to the customer. The services that are provide to the customer by the service provider are transported through cloud carrier that act as an intermediate and connects cloud consumer and cloud provider. Cloud broker act as an organization that manages how the cloud services is delivered to and used by the customer and whether customer is satisfied with the service or not. It also negotiates the relationship between consumer and service provider. Cloud auditor performs the individual analysis of the various services provided to the customer and its performance. Cloud service is delivered in three different service models based on the service requirement of the cloud customer. The three service models are

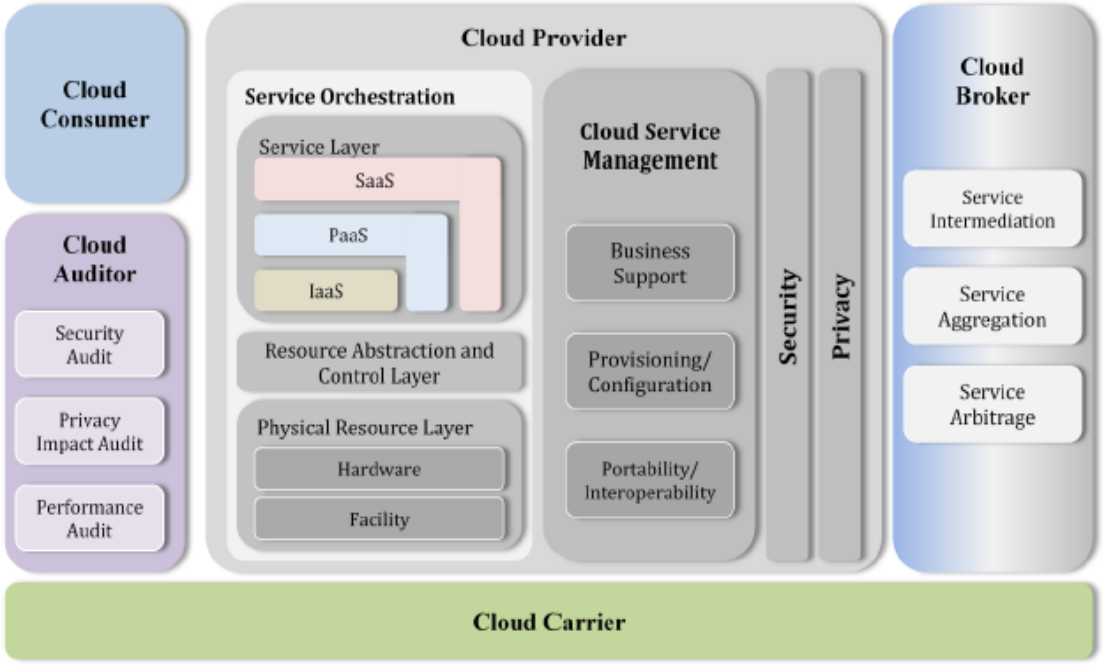

Fig 1. Cloud architecture

A. Software as a Service (SaaS) model

This service allows the customer to access the software or the applications that are hosted by the service providers. Using this service customer does not have to worry about the purchasing or the installation of the software rather rents it for use as pay per as use model. Customer can access the service through interface such as web browsers at anytime and anywhere. Most of the administrative responsibilities of the software or the application are handled by the service provider and consumer have very limited control over the administrative responsibilities. Example of SaaS model applications are google docs, Microsoft office, Evernote etc.

B. Platform as a Service (PaaS) model

This service model provides the access to the flexible computing and storage infrastructure that is coupled with a software platform [3]. This service model is consolidated form of IaaS. Customer does not have access to virtual machines but they can write the code in the highly integrated software platform. The toolkit, execution resources and standards for development are provided by the service provider. The development tools are hosted in the cloud and can be accessed through web browser. Cloud consumers of PaaS model are generally application developer who design and implements application software. Google's app engine, Salesforce etc. are some of the examples.

C. Infrastructure as a Service (IaaS) model

IaaS model provides the infrastructure to the organizations to deliver custom business solutions. Developers get the full access to the storage and computing resources. Virtualization technique is used to provide the access to the resources. The infrastructure is built in such a way that it can handle the peaks and troughs of its customer demand and add new capacity as the demand increases. Cloud providers controls the physical hardware and cloud software that makes the provision of these infrastructure services possible. Scalability and pay as you go are some of the features provided by the IaaS service model. Amazon web services (EC2 and S3), Microsoft azure etc. are examples of IaaS service model.

To provide the cloud services cloud is deployed in different deployment models. Each deployment model have their different ownership, size and access. There are four deployment model associated with cloud computing that are discussed below.

A. Private clouds

It is used by a single organization that uses the cloud infrastructure and computing resources exclusively. It can be managed internally or by a third-party and can be hosted on organization premises or on the third-party premises. Here organization itself takes the responsibility of their own data. Organization can enable pooling and sharing of computing resources across different applications, departments or business units.

B. Public clouds

The computing resources of this deployment model are accessed by the public customers. Public cloud is hosted, operated and managed by the third party cloud provider from one or more data centers. The cloud provider is responsible for the security management and operations in the public cloud. An individual customer is not aware of their neighbours which arises the security and data protection concerns.

C. Community clouds

Community clouds are used by the organizations that have shared concerns such as objectives, security, privacy and compliance. Similar to private clouds it can be managed by organization or by a third-party. The community benefits from public cloud capabilities but they also know who their neighbours are that reduces the security and data protection concerns.

D. Hybrid clouds

Hybrid clouds consist of multiple off-site or in-site community, private or public clouds. Each cloud is distinct but are bound together by standard technology. In this deployment model public clouds are used for general computing while customer data is stored within private cloud or community cloud. Its architecture is developed is such a way so that it allows interface with different management systems [4].

III. LOAD BALANCING IN CLOUD

To understand the concept of load balancing we need to first understand the concept of clustering or cluster of computers. Clustering is basically a concept in which we use multiple servers for accessing a web application or information instead of a single server. For this to work, we need to install same or different operating systems like windows 2008, some version of Linux etc. on each of the servers we are using for clustering. Next, we need to have an application installed on each of these servers, which cannot be clustered. For this we mostly use database applications like MYSQL which is used in web programming.

When we form a cluster of these computers, replication occurs such that each of the server contains the same information as the other. So when user is trying to access the data base of one server, they basically hits the cluster rather than the particular server. The cluster figures out which server it should be directed to, which means that it will directed to a server which is not fully loaded with the hardware connections. But if one server already has enough connections, that is, it has reached the load limit, then it automatically gets rerouted to the next server in the cluster. If the next server is also loaded, then it is again rerouted till it finds a server which has available connections with it and is therefore not loaded. Here all the servers contains the same information, therefore switching of servers is not a matter of concern.

This concept in which we direct the incoming users to a server which has the least load on it is called load balancing**.** It is a very important concept in the present web world. Load balancing certainly has some advantages that cannot be ignored. If one server fails, then the cluster realizes the failure and it does not send the user to that particular server. This means that the failure of one system does not affect the system, instead a different server with available connections substitutes for it. So along with optimality, it's also a safer option. Secondly we can have different types of hardware like Pentium, Celeron, Xeon, AMD etc. processors in servers connected to each other for load balancing. They may not necessarily have the same processors.

Load balancing is of two types:

1. Static Load balancing

In contrast to dynamic load balancing it requires a predefined knowledge about the system that we are working on like the processing power, memory etc. It is a technique that doesn't depend on the present state of the system. Even though the static environment is much simpler to simulate, it is difficult to simulate for heterogeneous cloud environment [5].It basically distributes the load on the servers equally which can also be referred to as round robin algorithm. But there are several drawbacks in the algorithm which leads to the introduction of the concept of weighted round robin algorithm in which each server assigns a particular weight and the connections are received accordingly, i.e. , higher the weight greater the number of connections.

2. Dynamic load balancing

It is a load balancing algorithm that doesn't need any prior information about the system. In-fact these load boundary decisions is exclusively based on the existing or current status of the system. It is difficult to simulate dynamic environment but they are mostly suited with cloud computing environment [6].

IV. LOAD BALANCING ALGORITHMS

1. MapReduce Algorithm

Map reduce is a property of load balancing concept that computes large amounts of obtained information in parallel configuration. It involves dividing a set of the input key pair and obtaining a set of immediate key pairs. As the name suggests, it is a concept that involves two basic operations namely mapping and reducing.The map operation is used for taking the input pairs to produce output key pairs. What it basically does is that, it takes a set of inputs, identifies the set of similar key pairs and maps them into a single set. Now this single set containing the similar information or key pairs is reduced to a single pair using the reduce algorithm. So on a whole, the reduce operation takes in the immediate key and the values that are associated with that key, merges these values and forms a smaller set of values. The output is obtained in binary form.

This concept also involves an object called iterator object that supplies the user's reduce operation and enables the user to go through all the elements in the collection. Hadoop is the unique programming model that provides an organized procedure to execute this programming paradigm. Since this model reduces these given set of values into a smaller set of values, therefore it allows the user to quickly write and test distributed systems. [7].

Given an example:
Function map(file_name, file_contents)
For each w in contents
Emit(w,1)
Function reduce(words, partialcounts)
Sum=0;
For each value in partialcounts:
Sum=sum+ parseint(pc)
Emit(words, sum)
Mapreduce algorithm increases efficiency of throughput for large data sets. But when these sets are smaller, there is no significant increase in the efficiency of the algorithm.

2. Round robin algorithm

Round robin algorithm is a wide IP level load balancing algorithm. It is a simple and beneficial algorithm at a global level scale, but it may not be the best load balancing algorithm available. As the name suggests, in a round robin algorithm the contents are being accessed on a rotational basis. Here the first available data server is granted the first request, second available data server is granted to the second request and so on. When the server IP address has been allocated an IP address, it is therefore moved at the back of the present IP address list, and it gradually comes back to the top of the list, and therefore can be accessed again. This algorithm based on the rotational of the IP address is called round robin. So the frequency of a particular IP address coming to the top of the available IP address list is dependent on the number of IP address available on that list. The servers that are distributed on various geographic locations are the best catered by applying DNS load balancing round robin server.

But it shall be noted that it is certainly not the best load balancing algorithm. Even though the algorithm follows a simplistic approach, it has certain disadvantages as well. This algorithm might result in unpredictability and may even corrupt the DNS tables of the best network load balancer. Round robin algorithm can be seen as a simple algorithm, easy to execute and involves easy to maintenance which can produce a more than satisfactory results in some situations, however it is not the optimal algorithm and has a certain unavoidable disadvantages and is by no means the best algorithm.

3. Artificial Bee algorithm

Some of the load balancing techniques are complex and are inefficient, so to solve these disadvantages there is a honey bee algorithm. Artificial bee mechanism introduces an optimization method based on the gathering behaviour of honey bees.

For this we need to understand that how does a bee get its food. There are two types of bees. One is employed bees and other is unemployed bees. The leader of the bee clan comes under the employed category. The leader accesses the food source, gets information of food and that has been forged, then comes back to the hive and shares the information with its other fellow bee mates through a dance called waggle dance. The profit and feasibility is determined by the duration of the dance. The second category of bees, i.e., the unemployed category, has two types within themselves. Detectors and followers. Detectors have a duty of searching for new food sources for the leader to pay a visit to. The followers on the other hand wait for them at the beehive, use the available information and analyse the profit and decide the source.

There are three types of decision they work upon.

1) Abandon the food source
2) Continue dancing and getting information
3) Continue gathering honey

The honey bee gathering mechanism is used in the search algorithms as follows.

1) While computing profit of certain requests , profit limits can be set accordingly if we normally calculate the requests, it takes certain amount of CPU time as well as waiting time.
2) Recording the leader bee's profit into the database which equals the dancing area. The remaining part of the request searches for data from the database and therefore comparisons can be easily made. It makes it into search servers that provide high amounts of profits using lightly loaded nodes.

4. Ant Colony Optimization

Ant colony optimization algorithm is an algorithm that is inspired from the social characteristic behaviour of ants. Although individually ants are pretty ordinary creatures in terms of finding the optimal path on their own. They have a limited memory and exhibit random behaviour. But when these ants work in a group, they are able to perform a variety of complex tasks. These complex behaviour of ants have been observed by people and scientists who are now trying to solve different computational tasks with optimization technique used by them. It is a technique used to decipher the shortest path out of the given paths and is now considered as a field of ant colony optimization.

This technique is arguably the most successful and widely organized algorithm based on ant behaviour. While travelling from their homes to food source or vice versa, ants lay pheromone trails. These shortest path recognition algorithm is dependent on the strength of these trails. So when ants move, they often stop in between for a while and drop this

pheromone on the ground, so when the next ant moves in the same direction, it considers the strength of the trail left by the previous ants and observe assess the time when it was laid and hence choose the path with the freshest pheromone trail on it and hence in such a way are able to figure out the shortest path of the trail.
Ant colony optimization has several applications in scheduling such as job shop problem, single machine total weighted tardiness problem etc. In this algorithm, the information on resource is dynamically refreshed for every individual moment which can be considered as a major advantage over other algorithms. Load balancing systems is based on multiple ant colonies data.

V. OBSERVATION

In this paper we have studied about cloud computing and different types of load balancing algorithms used in cloud computing. We have mainly focused on four different algorithms i.e. round robin load balancing algorithm, mapreduce load balancing algorithm, honey bee algorithm and ACO algorithm. During the study many observations were noted down that are given in table 1.

Table 1- Comparison of load balancing algorithm

Algorithm	**Static Environment**	**Dynamic Environment**	**Scalability**	**VM Load distribution**	**Service provider cost**
MapReduce	Yes	No	Yes	Fast	Low
Round Robin	Yes	No	Yes	Slow	High
Honey Bee	No	Yes	Yes	Slow	High
Ant Colony	No	Yes	Yes	Fast	Low

Table 2 contains the remarks about the different algorithm that are discussed above.

Table 2- remarks on algorithms

Algorithms	**Remarks**
MapReduce	• MapReduce uses parallelization and aggregation to schedule applications across clusters • Efficiency of throughput increases for large volume of data • MapReduce load balancing works more efficiently in clusters
Round Robin	• Round Robin algorithm is simple to implement but is not very efficient way for load balancing as compare to other algorithms • It uses centralized load balancing technique • Round Robin load balancing has fast execution time but low throughput
Honey Bee	• Based on behavior of honeybees to find optimal solution • It takes more execution time as compared to other algorithms • Throughput does not increases as the system size increases
Ant Colony	• It requires minimum time to find overloaded node • It uses distributed load balancing technique

VI. CONCLUSIONS

This paper has presented a study on cloud computing and various algorithm for load balancing in cloud computing. No doubt cloud computing is one of the most emerging technology in IT but it also have some issues and load balancing in cloud is one of the major issues of the cloud. This issue can be resolved by using various load balancing algorithm that balances the workload. This paper gives study on some of the different load balancing algorithms like Mapreduce, Round Robin, Honey Bee and Ant Colony Optimization and also gives comparison between them on different properties. According to the study, Ant Colony Optimization is better load balancing algorithm as compared to other algorithms.

REFERENCES

[1] Klaithem Al Nuaimi, Nader Mohamed, Mariam Al Nuaimi and Jameela Al-Jaroodi, "A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms", 2012 IEEE Second Symposium on Network Cloud Computing and Applications.
[2] Fang Liu, Jin Tong, Jian Mao, Robert Bohn, John Messina, Lee Badger and Dawn Leaf, NIST, cloud computing reference architecture.

[3] Indranil Gupta, "Introduction to cloud computing", CS425/ECE428 Distributed systems, University of Illinois at urbana-champaign.
[4] Ramgovind S, Eloff MM, Smith E, The Management of Security in Cloud Computing, Information Security for South Africa (ISSA), IEEE, 2010.
[5] Mayanka Katyal, Atul Mishra, "A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment", International Journal of Distributed and Cloud Computing, Volume 1 Issue 2 December 2013.
[6] Nikita Haryani, Dhanamma Jagli, "Dynamic Method for Load Balancing in Cloud Computing", IOSR Journal of Computer Engineering (IOSR-JCE).
[7] The IBM website. [Online]. Available: http://www.ibm.com/developerworks/cloud/library/cl-mapreduce.

