A Comparative Study of Load Balancing Algorithm in Cloud Computing

Author: **Mr. Vinod KishanLal beg,**
Lecturer, Neville Wadia Institute of Management Studies & Research, Pune

Under the guidance of **Dr. Minesh Ade,**
(PhD Guide)

Abstract:

Cloud computing is the hottest technology in the field of computer science. The Cloud computing technology is changing our life by providing users with new types of services viz. PaaS, SaaS and IaaS as the major types. The term "Cloud Computing" involves virtualization, distributed computing,networking, software and web services. A cloud consists of several elements s clients, datacenter and distributed servers. Cloud Computing also includes fault tolerance, high availability,scalability, flexibility, reduced overhead for users, reduced cost of ownership, ondemand services etc.

As more and more organizations are moving to avail the benefits of cloud computing, it has become important for the service providers to address the main issue related to cloud computing. Issues with the performance of business applications can cause detritions of an organization business performance. Recent research studies have indicated that many organizations have lost significant part of their revenues due to lack of efficient scheduling and load balancing resource allocation techniques. It is important to reduce the operational cost and give customer satisfaction by using the appropriate and efficient load balancing algorithm. Central to all these issues laid the establishment of an effective load balancing algorithm.The load can be CPU load, memory capacity, delay or network load. Loadbalancing is the process of distributing the load among various nodes of a distributedsystem to improve both resource utilization and job response time while also avoidinga situation where some of the nodes are heavily loaded while other nodes are idle ordoing very little work. Load balancing ensures that all the processor in the system orevery node in the network does approximately the equal amount of work at any instantof time. This technique can be sender initiated, receiver initiated or symmetric type (combination of sender initiated and receiver initiated types). In this paper we study round robin and honeybee foraging algorithms to estimate response time, processing time which has the major impact on cost. The result of this paper to suggest which load balancing technique should be used to improve application performance to improve revenue growth, cost saving and reputation. To test the performance of the algorithms the researcher intends to use java codes and Cloud Analyst as simulator.

***Keywords**: Cloud computing, Load balancing algorithm, Cloud Analyst*

Introduction:

There has been much discussion in industry as to what cloud computing actually means. The term cloud computing seems to originate from computer network diagrams that represent the internet as a cloud. Most of the major IT companies and market research firms such as IBM, Sun Microsystems, Gartner and Forrester Research have produced whitepapers that attempt to define the meaning of this term. The US National Institute of Standards and Technology (NIST) has developed a working definition that covers the commonly agreed aspects of cloud computing. The NIST defines cloud computing as:

"A model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction"[1]

Gartner has defined Cloud Computing as *a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service using Internet technologies.* [2]

Cloud computing is a general term for anything that involves delivering hosted services over the Internet. These services are broadly divided into three categories: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). The name cloud computing was inspired by the cloud symbol that's often used to represent the Internet in flowcharts and diagrams.

There are three types of services namely public, private and hybrid. A Public Cloud sells services to anyone on the Internet. E.g. Amazon Web Services (EC2). Whereas a Private Cloud is a proprietary network or a data centre that supplies hosted services to a limited number of people. E.g. Google (GOOG) – Apps Engine, Salesforce.com (CRM) etc. When a service provider uses public cloud resources to create their private cloud, the result is called a virtual private cloud. Private or public, the goal of cloud computing is to provide easy, scalable access to computing resources and IT services.

One of the appealing aspects of cloud computing is that it hides the complexity of IT technology from users and developers. No need to know details of how a service is generated – it is the service provider's job to provide a corresponding abstraction layer. Some of the important technologies on which cloud computing depends are virtualization, service-oriented architectures (SOA), and Web services. Cloud computing is efficient and scalable but maintaining the stability of processing so many jobs in the cloud computing environment is a very complex problem with load balancing being one of the important concerns. Load Balancing is a method to distribute workload across one or more servers, network interfaces, hard drives or other computing resources. Typical datacenter implementations rely on large, powerful computing hardware and network infrastructure, which are subject to the usual risks associated with any physical device, including hardware failure, power and/or network interruptions, and resource limitations in times of high demand. Load balancing in the cloud differs from classical thinking on load-balancing architecture and implementation by using commodity servers to perform the load balancing. Load balancing is used to make sure that none of your existing resources are idle while others are being utilized. To balance load distribution, you can migrate the load from the source nodes (which have surplus workload) to the comparatively lightly loaded destination nodes. This provides for new opportunities and economies-of-scale, as well as presenting its own unique set of challenges. [3] 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 or dynamic. [4] Load balancing at runtime is called dynamic load balancing and it is based on the current state of the system. No prior knowledge is needed. Whereas static load balancing does not depend on the current state of the system and it requires prior knowledge of the system. Load balancing ensures that all the processor in the system or every node in the network does approximately the equal amount of work at any instant of time. This technique can be sender initiated, receiver initiated or symmetric type (combination of sender initiated and receiver initiated types). The important things to consider while developing such algorithm are: estimation of load, comparison of load, stability of different system, performance of system, interaction between the nodes, nature of work to be transferred, selecting of nodes and many other ones. [5] This load considered can be in terms of CPU load, amount of memory used, delay or Network load. There are various load balancing algorithms which are used to improve the performance of the load balancers in the market. We will study Round Robin and honeybee – based load balancing techniques and check the performance time and cost.

Distributed Load Balancing Techniques:

Round Robin Algorithm

Round-robin (RR) is one of the algorithms employed by process and network schedulers in computing. As the term is generally used, time slices are assigned to each process in equal portions and in circular order, handling all processes without priority (also known as cyclic executive). Round-robin scheduling is

simple, easy to implement, and starvation-free. Round-robin scheduling can also be applied to other scheduling problems, such as data packet scheduling in computer networks. It is an Operating System concept. The name of the algorithm comes from the round-robin principle known from other fields, where each person takes an equal share of something in turn.

Honeybee- foraging Algorithm

The main idea behind the Honeybee Foraging algorithm is derived from the behavior of honeybees. There are two kinds of honeybees: finders and reapers. The finder honeybees first goes outside of the honey comb and find the honey sources. After finding the source, they return to the honey comb and do a waggle dance indicating the quality and quantity of honey available. Then, reapers go outside and reap the honey from those sources. After collecting, they return to beehive and does a waggle dance. This dance indicates how much food is left. M. Randles proposed a decentralized honeybee based algorithm for self-organization. In this case, the servers are grouped as virtual server and each virtual server have a process queue. Each server, after processing a request from its queue, calculates the profit which is analogous to the quality that the bees show in their waggle dance. If profit is high, the server stays else, it returns to the forage. This algorithm requires that each node to maintain a separate queue. This computation of profit on each node causes additional overhead.

The disadvantage of this algorithm is that, it does not show any significant improvement in throughput, which is due to the additional queue and the computation overhead. [7]

Result Summary:

The simulation was done using CloudAnalyst 1.0 beta version and the summary of the simulation is given below:

Response Time of Honeybee Foraging Algorithm and RR Algorithm for 5 servers

Table 1:

Server No.	Response Time(in ms)	
	Honey Bee Foraging Behavior	Round Robin
1	201.87	155.76
2	171.35	124.58
3	124.12	77.97
4	107.89	60.68
5	90.78	44.77

Figure 1:

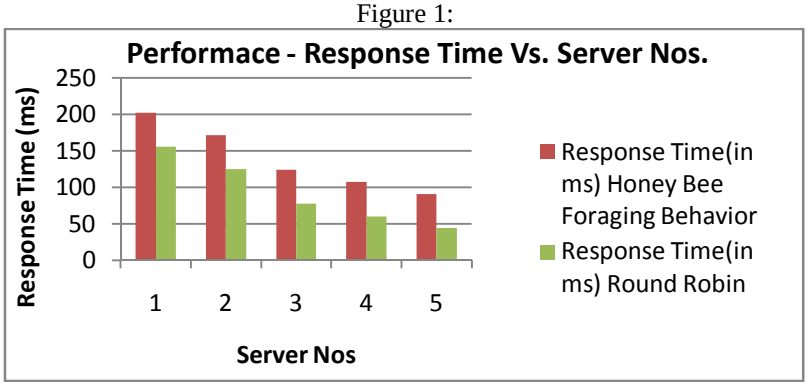
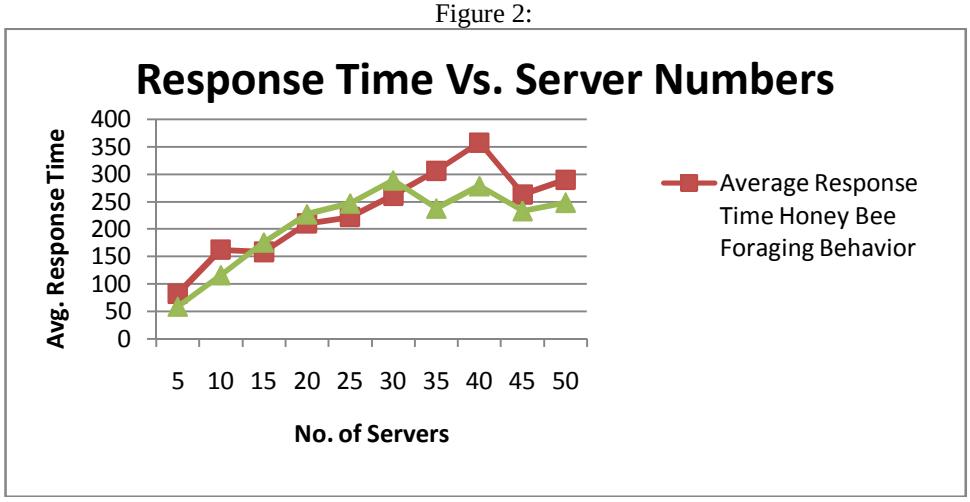


Table 2:

No. of Servers	Average Response Time	
	Honey Bee Foraging Behavior	Round Robin
5	81.7	58.3
10	162.2	115.6
15	158.25	175.5
20	209.7	227.4
25	221.7	246.7
30	260.5	288.9
35	306.3	237.8
40	357.4	277.99
45	262.8	233.1
50	289.4	248.56

Figure 2:



Conclusion: Cloud Computing an internet based computing provides resources to the users on demand. The user pays for the service he uses to the provider.Thus response time and process time reduction has become one of the very important concern to both the user and the provider while building a cloud system. Distributed Load Balancing Algorithms of cloud computing are analyzed using CloudAnalyst. From the analysis and comparison, it is concluded that Round Robin algorithm takes lesser time for Execution than Honeybee Foraging Algorithm. Thus, Round Robin Algorithm can be used compared to Honeybee Foraging Algorithm to have a better response time and give customer satisfaction, which can further reduce operational cost.

References:

[1] Mell, P. a. (2009). Draft NIST Working Definition of Cloud Computing.

[2] www.gartner.com/it-glossary/cloud-computing

[3] Ratan Mishra and AnantJaiswal, (2012) "Ant Colony Optimization: A Solution of Load balancing in Cloud", IJWesT, *vol.*-3, No.2, pg.33-50

[4] N.G. Shivaratri, P. A. (1992). Load Distributing for Locally Distributed Systems. *Computer*, 33-44.

[5] Alakeel, A. M. (2010, June). A Guide to Dynamic Load Balancing in Distributed Computer Systems. *IJCSNS International Journal of Computer Science and Network Security, 10*(6).

******