International Journal of Research in Advent Technology, Vol.7, No.5S, May 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

Cloud Computing : Research Aspects, Strategy and Measures

[1]Shilpa Kukreja, [2]Surjeet Dalal
Department of CSE, SRM University, Delhi-NCR, Sonipat, Haryana

Abstract: Making a service available to the client is becoming an important job for cloud providers. Failure in providing the promised service decreases the system performance. Due to the increasing demand of cloud resources, the system has to manage heavily loaded servers with the customer's demand. To achieve this, Load balancing is playing a big role in cloud computing, which improves the performance by evenly distributing the load across the participating nodes. Several load balancing algorithms have been applied to manage the incoming demands to make the system balanced. The algorithms like round robin, cat swarm algorithm, ant colony algorithm, etc. are used to manage the load effectively. In this paper, we have discussed how cloud computing came into existence, its adoption by the industries, its increased demand, research aspects of cloud computing.

Key Attributes: Cloud computing, research aspects, load balancing.

1. INTRODUCTION

Cloud computing has deeply influenced IT industries to adopt cloud services in their business. A strong thirst of using cloud computing is to realize their capabilities which can be accessed over the internet. Cloud computing is a new technology which leverages virtualized services to IT consumers with the promised quality of service parameters. Cloud architecture is useful and capable of providing the development, deployment and management of client applications on cloud platform. It reduces overhead of handling the resources for the users, removes the need for preplanning of resources to be used providing the increased demand of resources whenever needed [1]. Cloud computing follows some principle providing the benefits:

- Takes total control of managing the resources
- Strict payment on the basis of usage
- Offering service on demand
- Hosting client's applications
- Always service available

Cloud computing has penetrated IT professionals to add cloud services in their enterprise. Cloud has already received a great response from their customers, influenced by its amazing features. Only cloud is a big place where user's can deploy and manage their applications accordingly. Now, we can understand from the above features, why even large enterprises also have integrated their firms with Cloud which in turn increased their market value. Here, we have Gartner's Hype Cycle for Emerging Technologies of 2011 when Cloud computing became most popular among the others. We all know that Cloud computing came into market and gained popularity in two to three of arrival that is in 2011 shown in Fig. 1. From this, we can see that cloud computing has been adopted and the expectations of enterprise have also been increased.

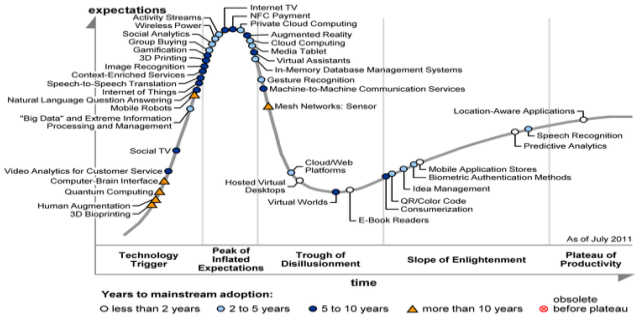

Fig. 1: Gartner's 2011 Hype Cycle of Emerging Technologies [2]

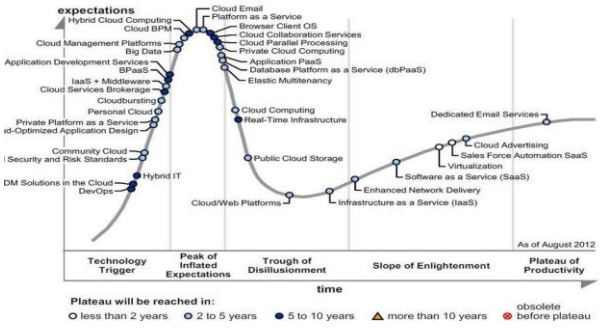

Fig. 2: Hype Cycle of Cloud Computing [2]

The business enterprises identified the potential of cloud computing by its huge infrastructure, capabilities, flexibility, reliable services, availability required to own a successful business. There is Gartner's Hype Cycle shown in Fig 1 and 2 which gives emergence of technologies, their adoption, expectations and their impact on business firms. Figure 2 shows all about cloud technology, how it has grown up and the organizations has started providing cloud services within very shortly. Due to the facilities provided to customers, cloud became popular and gained a huge success among the other technologies as shown in fig 1.

There were lots of challenges and reasons for adding cloud services in the business. We are underlying some reasons that we have studied during review of research articles.

Reasons for Adopting Cloud Computing	
Adoption Reasons	**% of Study**
Wide range of solutions	90.22
Reduced infrastructure cost	88.67
Availability of service	89.75
Assured QoS parameters	89.30
Reducing the overhead of resource ownership	88.25
Unlimited data storage	89.65
On-demand provisioning	87.34
Backup and recovery	83.21
Quick resource management	87.45

Fig 3: Cloud Computing Adoption Reasons and Study Remarks

1.1 Recent Research Aspects in Cloud Computing

The reasons behind the adoption of cloud are due to its success in providing a wide range of solutions to the organizations. Still, we have examined some research issues left to be solved, which may affect the performance in the market. Basically, there are two research aspects of cloud computing provided in fig 4:

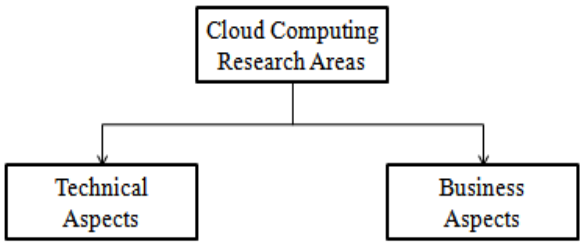

Fig 4: Research Categories in Cloud Computing [3]

Cloud delivers resources to the end users in a distributed manner to accomplish the request needed for the client. The components and subcomponents of the system must cooperate with the others to fulfill the client's request so that the correct number of resources can be provisioned to the customer at the time of need. Due to the huge increasing number of requests, some resources may be overloaded and some may be underloaded which in turn makes the system imbalance. Imbalanced load decreases system performance and also the efficiency of the algorithm will also be decreased. Load Balancing provides the solutions for related problems to balance the load effectively among all the nodes in the system. Load Balancing is an important technical aspect in cloud computing as you can see in fig 5.

Technical Aspects	**Business Aspects**
Load Balancing	Quality of Service
Availability	Cost Optimization
Virtualization	Resource Management
Storage Technologies	Standards
Security	Constraints
Dynamic Provisioning	Service Level Agreement
Backup and recovery	Data Management
Multi-tenant	Energy Management

Fig 5. Technical and Business Aspects in Cloud Computing [3]

2. LOAD BALANCING IN CLOUD COMPUTING

The objective of load balancing is to provide even distribution of load between all participating components so that resources can be used optimally. "Load Balancing is the process of improving the system's performance by evenly distributing the load so that no node will remain overloaded and underloaded." Many researchers explained their ideas for efficient load balancing algorithm to enhance the quality of service parameters.

Static Load Balancing Algorithms: Static algorithm collects node's information like capacity, capability and performance, but does not consider the current load status of the machine. No change can be made after the execution of the process [4]. Only earlier knowledge is required in this required in the process.

Dynamic Load Balancing Algorithms: This algorithm manages load changes in the system and also reassigns the load dynamically. It also requires some prior information to start the process, but also monitors regular load changes in the system. Dynamic algorithms follow some strategy in managing the load [5]:

Transfer Strategy: The strategy is responsible for transferring load from heavily loaded node to underloaded node.

Location Strategy: It determines the available node to transfer the task and selects the best one. Location strategy has three sub-policies: probing, random and negotiation.

Information Strategy: The task of this strategy is to collect workload related information of every node so that transfer and location strategy can get an exact knowledge of the node.

Triggering Strategy: This strategy depicts the time of starting the load balancing process. It works with two sub-processes: sender-initiated, receiver-initiated.

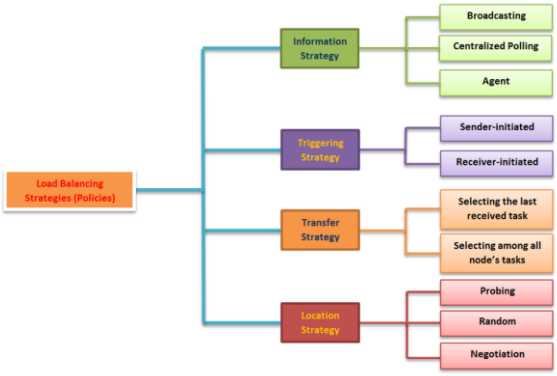

Fig 6: Strategies of Dynamic Load Balancer [5]

2.1 Challenges of Load Balancing in Cloud Computing

Before proceeding to the literature part of load balancing algorithms, we are underlying the issues and challenges that an algorithm considered in finding the solution for load balancing problem. Some main challenges are here:

A. *Central Node Failure*:

Some load balancing algorithms follows centralized approach in which central has the charge of distribution of load between the nodes and has the overall control of the system. In case, central node failure occurs, then the entire nodes will not respond and downs the system performance [6].

B. *Availability*:

Availability implies that clients can access the service at a time when needed. Load balancing can improve and maintain the availability of the resources. Cloud computing must provide the resources available whenever the user needs to execute the request.

C. *Virtual Machine Migration*:

Using virtualization, load of the heavily loaded virtual machine can be shifted to the other machine on the same physical machine. This can lead to the better utilization of resource and can use the underutilized machines optimally [5].

D. *Resource Utilization:*

Load distribution among the resources should be done according to a good allocation algorithm to achieve the best performance. Utilizing a resource depends on the distribution of tasks divided in the processors. In load balancing, even task distribution is a best way to achieve better resource utilization.

2.2 Review of Load Balancing Algorithm

Dinesh Babu et al. [1] Provided a honey bee inspired load balancing algorithm for nonpreemptive independent tasks. The algorithm maximized the throughput, reduced execution time and kept the tasks of higher priority in the queue with minimum waiting time.

A resource allocation mechanism has been provided in [2] to minimize the skewness metric and also tried well to manage the overload of the system. One more algorithm was also designed to predict the use of resources in the future. A trust based load balancing model was proposed by Gupta et al. [3] Which distributed the load effectively to increase the quality of service parameters. The algorithm used initiation time, fault rate and MIPS, to calculate the trust value.

A centralized load balancer was proposed in [4] that reduced the power consumption and made the efficient utilization of resources. The balancer skipped a VM when there was minimum load and extends the system the number of VM when the load was increased in the system.

The authors in [5] used a power-aware load balancing technique that is artificial bee colony approach migrate the load of heavy loaded to underloaded virtual machine. The approach found the overloaded machines and searched for low loaded machines to transfer the load.

Fei Ma et al. Presented a distributed load balancing model. The model assigns the virtual machine monitor to each data center to determine the resource utilization of each resource to depict the system performance [6].

2.3 Planning for Load Balancing Approach

We have studied lots of load balancing approaches in previous research papers. Each algorithm has its positive and negative impact on the system performance. We have prepared some points that will help in designing an efficient approach for load balancing. Previous algorithms specified their goals, but some of the goals left to fulfill, so I am considering the goals to make the overall load balanced.

Goals of Load Balancing

- Enhances the system performance
- Balances workload of the system
- Maintain system reliability
- Improve and maintain Availability
- Manage optimal utilization of resources
- Enhance Scalability and Flexibility

2.4 Measuring Efficiency of Load Balancing Algorithm

We are providing some parameters that will be consider in future to design a load balancing algorithm [7]:

Reliability (S): Reliability of an algorithm transfers the process from one location to the other. Static algorithms are less reliable than dynamic. But reliability increases waiting time to transfer the process and in turn dissatisfy the customer. Here, E denotes the number of successful runs and T denotes total runs available.

$$\text{Reliability (S)} = E/T$$

Response Time (S): Response time is the time when the system starts responding to the problem. Response time of static algorithms is always shorter than the dynamic algorithms.

Response Time (S) = Transfer time + Time to start running

Waiting time: It is the time spent waiting for start of execution of the process.

Adaptability: Adaptability checks the algorithm for its adaptation towards variable changes made in the system. Static algorithms are not adaptable since they do not consider changes in the system. Dynamic algorithms are adaptable.

Fault tolerance: In case any fault occurs, system must continue to work. The system must run so that the faults can not affect the user's running application.

3. CONCLUSION

In this paper, we have discussed cloud computing, principles, goals and reasons to adopt cloud in enterprise businesses. Some technical and business aspects have also been discussed which shows that how cloud is providing benefits to the organization and gaining control of resources of the technical firms. The popularity of cloud computing has also been discussed as we can see in Gartner's hype cycle provided in fig 1 and 2. From the research aspects, we have identified that load balancing is the basic objective that

must provide the solution for it. So, we have planned to give future enhancements of load balancing solutions based on the literature of our study. We have also detected some cons from our survey, which lead to the system performance decrement.we have decided to provide an optimal load balancing technique that will help to use the resources optimally and also will optimize the resource utilization. Our future work will be to provide load balancing algorithm.

REFERENCES

[1] R. Buyya, S. Pandey, C. Vecchiola, "Cloudbus toolkit for market-oriented cloud computing", CloudCom, LNCS, Vol. 5931, pp 24-44, 2009.
[2] L.Qian, Z. Luo,Y. Du and L. Guo, "Cloudcomputing: An Overview", CloudCom, LNCS, pp 626-631, 2009.
[3] A.C. Adamuthe, V.D. Salunke, S.H. Patil, "Cloud Computing-A market perspective and research directions", International Journal of Information Technology and Computer Science, Vol. 10, pp 42-53, 2015.
[4] Z. Chaczko, V. Mahadevan, S. Aslanzadeh and C. Medermid, "Availability and Load balancing in cloud computing", IPCSIT, Vol. 4, pp 134-140, 2011.
[5] M. Mesbahi, A.M. Rahmani, "Load Balancing in Cloud Computing: A state of art survey", International Journal of modern Education and Computer Science, Vol. 3, pp 64-78, 2016.
[6] A. khiyaita, M. Zbakh, H. El. Bakkali, D. El. Kettani, "Load Balancing Cloud Computing: State of art", Network Security and Syatems, pp 106-109, 2012.
[7] A. Singh, D. Juneja, M. Malhotra, "autonomous Agent based load balancing algorithm in cloud computing", Prrocedia Computer Science, Elsevier, Vol. 45, pp 832-841, 2015.
[8] L.D. Dinesh Babu, P. Venkata Krishna, "Honey Bee Behaviour Inspired Load Balancing of Tasks in Cloud Computing Environment", Applied Soft Computing Journal, Vol. 13, No. 5, pp 2292-2303, 2013.
[9] Z. Xiao, W. Song and Q. Chen, "Dynamic Resource Allocation using Virtual Machine for Cloud Computing Environment", IEEE transactions on parallel and distributed systems, Vol. 24 no. 6, pp 1107-1117, 2013.
[10] P. Gupta, M.K. Goyal and P. Kumar, "Trust and Reliability based load balancing algorithm for Cloud IaaS", 3rd IEEE International Conference on Advance Computing, 2013.
[11] M. Simjanoska, "L3B: Low level load balancer in the cloud, EUROCON, 2013.
[12] S.M. Gafari et al., "Bee-MMT: A Load balancing method for power consumption managementin cloud computing, 6th IEEE Conference on Contemporary Computing (IC3), 2013.
[13] F. Ma, F. Liu and Z. Liu, "Distributed load balancing Allocation of VM in cloud data center, 3rd IEEE conference on software engineering and service science (ICSESS), 2013.