ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

Cloud Computing: An Overview

Divya garg[1], Urvashi Saxena[2]

M.Tech Student, Dept. of C.S.E, JSS Academy of Technical Education, Noida, U.P.,India[1]

Assistant Professor, Dept. of C.S.E., JSS Academy of Technical Education, Noida, U.P.,India[2]

ABSTRACT: Cloud computing, one of the emerging topic in the field of information technology, is the development of parallel computing, distributed computing and grid computing. As Cloud Computing is growing rapidly and clients are demanding more services and better results, load balancing for the Cloud has become a very interesting and important research area. Many algorithms were suggested to provide efficient mechanisms and algorithms to enhance the overall performance of the Cloud and provide the user more satisfying and efficient service. In this paper, we investigate many
Algorithms proposed to resolve the issue of load balancing in cloud computing and also compare these algorithm on the basis of their properties.

KEYWORDS:- Cloud computing, Load balancing, Cloud Service Models, Replications.

I. INTRODUCTION

Cloud computing has been envisioned as the next generation computing model for its major advantages in on demand self-service, ubiquitous network access, location independent resource pooling and transference of risk. Cloud Computing is a catagory of distributed computing where Different massively scalable IT related resources are provided to a number of external users as a service using internet. It is an on demand service in which shared resources, information, software and other devices are provided according to the clients requirement at specific time.
The main feature of Cloud Computing is that it makes all the resources available at one place in the form of a cluster and the resources are allocated to the users according to their requests. This cluster based approach helps in achieving the maximum CPU utilization and reduces the effort of users to access the cloud resources.
The main part of Cloud Computing is the so called "cloud". Cloud is a group of computers – personal computers or servers- which are interconnected together. Cloud is the network that provides resources to the clients.
In 2011, NIST defined Cloud Computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable (e.g. , networks, applications, storage, services and servers) that can be rapidly provisioned and released with minimal management effort or service provider interaction. According to **NIST** [1] there are three service models, four deployment models and five characteristics.

Services of cloud computing is divided into Infrastructure-as-a-service (IaaS), Platform-as-a-service (PaaS) and Software-as-a-service (SaaS).
Four deployment model in cloud computing includes- Public Cloud, Private Cloud, Hybrid Cloud and Community Cloud.

The Five main characteristics of Cloud Computing are:
1. **On-Demand self service**: In this a user can use the computing capabilities on-demand automatically without requiring any interaction from service provider.
2. Rapid Elasticity: It provides resources and services which can be rapidly and elastically provisioned.
3. Broad Network Access: In this resources and services are available over the network and can be accessed using heterogeneous thick and thin client (Laptops, PDAs, Mobile phones).
4. Resource Pooling: It provides a pool of resources which can be accessed using a multi-tenant model.
5. Measured Services: It works on pay-as-per use model.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

II. CURRENT CLOUD COMPUTING PLATFORMS

1. AbiCloud Platform: AbiCloud is a cloud computing platform developed by Abiquo. It can be used to build, integrate and manage public as well as private cloud in the homogeneous environments. [2]User can easily and automatically deploy and manage the server, storage system, network, virtual devices and applications etc.

2. Eucalyptus Platform: Eucalyptus is an open source implementation of Amazon EC2 and compatible with business interfaces. It also implements virtualization depending on Linux and Xen as EC2.[3]Eucalyptus is an elastic computing structure that can be used to connect the users' programs to the useful systems, it is an open source infrastructure using clusters implementation of elastic, utility, cloud computing and a popular computing standard based on service level protocol that permit users lease network for computing capability. Currently, Eucalyptus is compatible with EC2 from Amazon.

3. Nimbus Platform: Nimbus is an open tool set and also a cloud computing solution providing IaaS. It permits users lease remote resources and build the required computing environment through the deployment of virtual machines. All the functional components can be classified as three kinds.[4] One kind is client supported modules which are used to support all kinds of cloud clients. The second kind of component is mainly service supported modules of cloud platform, providing all kinds of cloud services.

4. OpenNebula Platform: Like nimbus, it is also an open source cloud service framework. It allows user deploy and manage virtual machines on physical resources and it can set user's data centers or clusters to flexible virtual infrastructure that can automatically adapt to the change of the service load. The main difference of OpenNebula [5]and Nimbus is that nimbus implements remote interface based on EC2 or Web Service Resource Framework (WSRF) through which user can process all security related issues, while OpenNebula does not.

III. CHALLANGES IN CLOUD COMPUTING

Before we could review the current load balancing approaches for Cloud Computing, we have to identify the issues and challenges[6] that could affects the algorithm. So we discuss the challenges to be addressed when attempting to propose an optimal solution to the issue of load balancing in Cloud Computing. These challenges are-

1. **Security and Privacy**: In cloud computing, a data center holds information that end-users would more traditionally have stored on their computers. This raise concerns regarding user privacy protection because users must outsource their data.
2. **Data Transfer Bottlenecks**: Applications continue to become more data-intensive. If we assume applications may be "pulled apart" across the boundaries of clouds, this may complicate data placement and transport.
3. **Replication time and costs**: How fast can the data be replicated is important for data resiliency.
4. **Reliability**: Servers in the cloud can have the same problems as the organization's resident servers. Downtimes and slowdowns can occur with cloud servers too.
5. **Freedom**: Cloud computing does not allow users to physically possess the storage of the data, leaving the data storage and control in the hands of cloud providers.
6. **Open Standard**: Open standards are critical to the growth of cloud computing. Most cloud providers expose APIs which are typically well-documented but also unique to their implementation and thus not interoperable.
7. **Network transmission problem**: Cloud computing services depend on the low-level network. At present wan rate is low and unstable, it has a direct impact on the cloud computing capability.

IV. LOAD BALANCING

Load balancing is one of the central issues in cloud computing. The load can be CPU load, memory capacity, delay or network load. Load balancing is the process of distributing the load among various nodes of a distributed system to improve both resource utilization and job response time. 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.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

Need of Load Balancing in Cloud Computing- Load balancing in clouds is a mechanism that maintain the load to each processing element such that all the processing elements become neither overloaded nor idle. It is used to achieve a high user satisfaction and resource utilization ratio. It also helps in implementing fail-over, enabling scalability, avoiding bottlenecks and over-provisioning, reducing response time etc . By load balancing strategy it is possible to make every processor equally busy and to finish the works approximately at the same time. The aim of load balancing algorithm is dynamic in nature which does not consider the previous state .
Goals of Load balancing are:

- To improve the performance substantially,
- To have a backup plan in case the system fails even partially,
- To maintain the system stability,
- To accommodate future modification in the system.

V. LOAD BALANCING ALGORITHMS REVIEW

Here we discuss the literature for load balancing in Cloud Computing. The load balancing algorithms are of two types: static algorithms and dynamic algorithms.

1. **Static Load balancing algorithms –** In this, assign the tasks to the nodes based only on the ability of the node to process new requests. [14]The process is based solely on prior knowledge of the nodes' properties and capabilities like node's processing power, memory and storage capacity, and communication performance. These algorithms cannot adapt to load changes during run-time. Examples-

- **Round Robin Algorithm**: In this algorithm, the processes are divided between all processors. The process allocation order is maintained locally independent of the allocations from remote processors. In Round Robin, it send the requests to the node with the least number of connections, so at any point of time some node may be heavily loaded and other remain idle .
- **Central Load Balancing Decision Model (CLBDM):** CLBDM is an improvement of Round-Robin algorithm. It is based on session switching at the application layer[7]. The improvement is that, in the cloud it calculated the connection time between the client and the node, and if that connection time exceeds a threshold then connection will be terminated and task will be forwarded To another node using the regular Round Robin rules.
- **Map Reduce Algorithm**: MapReduce is a computing model and generating large datasets . Map task and reduce task two main task in this model. Map takes an input pair and produces a set of intermediate value pair and Reduce task accepts an intermediate key and a set of values for that key and merges these values to form a smaller set of value. Map task read entities in parallel and process them, this will cause the Reduce task to be overloaded.[8]
- **Ant colony optimization (ACO)**: In ACO [9] algorithm when the request in initiated the ant start its movement. Movement of ant is of two ways*:*
  Forward Movement : Forward Movement means the ant in continuously moving from one overloaded node to another node and check it is overloaded or under loaded ,if ant find an over loaded node it will continuously moving in the forward direction.
  Backward Movement: If an ant find an over loaded node the ant will use the back ward movement to get to the previous node, in the algorithm if ant finds the target node then ant will commit suicide, this algorithm reduced the unnecessary back ward movement, overcome heterogeneity, is excellent in fault tolerance.

2. Dynamic Load Balancing Algorithms- These algorithms rely on a combination of knowledge based on prior gathered information about the nodes in the Cloud and run-time properties. Such algorithms require constant monitoring of the nodes and are usually harder to implement.[10] They are more accurate and could result in more efficient load balancing. Examples-

- **Index Name Server Algorithm (INS):** The goal of INS algorithm is to find an algorithm to minimize the data duplication and redundancy.[11] INS have some parameters which help in calculating the optimum selection point, the position of the server, the transition quality, the maximum bandwidth. Another calculation point whether the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

connection can handle additional nodes or not. They classified the busy levels B(a),B(b), and B(c). B(a) means that connection is very busy and cannot handle any additional connection. B(b) means connections is not busy and can handle additional connections. B(c) means That the connection is limited.

- **Load Balancing Min-Min Algorithm (LBMM***):* LBMM [12] has a three level load balancing framework. In first level LBMM architecture is the request manager which is responsible for receiving the task and assigning it to service manager, when the service manager receives the request; it divides it into subtask and assigns the subtask to a service node based on node availability, remaining memory and the transmission rate which is responsible for execution the task.
- **Dual Direction Downloading Algorithm (DDFTP):** DDFTP is a dual direction downloading algorithm from FTP server [13]. This is a fast and efficient concurrent technique for downloading large files from FTP server in a cloud environment. DDFTP uses the concept of processing the files for transfer from two different directions. The algorithm reduces the network communication between the client and nodes and network overhead.

VI. COMPARISON OF ALGORITHMS

In this section, we compare different algorithms which were discussed in section V based on the challenges discussed in Section III. Table I shows a comparison among the reviewed algorithms. The comparison shows the positives and negative points of each algorithm.

TABLE I. PROS AND CONS OF LOAD BALANCING ALGORITHMS

	PROS	CONS
INS	Initially proved to handle some sort of load balancing	.• Complicated in terms of implementation. • Only certain parameters are considered such as distance and time
CLDBM	• Solves issues of Round Robin Algorithm •Automated tasks forwarding reduces the need for a human administrator	• Inherits Round Robin issues such as not taking into consideration node Capabilities • Single point of failure • The threshold might not be applied to all cases.
ANT COLONY	• underloaded node is found at beginning of the search • Decentralized, no single point of failure • Ants can collect the information faster	• Network overhead because of the large number of ants • Points of initiation of ants and number of ants are not clear • Nodes status change after ants visits to them is not taken into account
ENHANCED MAP REDUCE	• Less overhead for the reduce task	• High processing time • Reduce tasks capabilities are not taken into consideration
DDFTP	• Fast • Reliable download of files	• Full replication of data files that requires high storage in all nodes.
LBMM	• Reliable tasks assignment to nodes	• Slower than other algorithms because Work must pass through three layers to be processed.

VII. CONCLUSION AND FUTURE WORK

This paper presents all about of cloud computing which is a new emerging technology in present world. We surveyed multiple algorithms with their advantages and disadvantages in cloud computing. Although there is a comparison between the existing algorithms based on the challenges we discussed. In cloud computing load balancing is the main

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

issue. Load balancing is required to distribute the excess dynamic local workload evenly to the entire node in the whole cloud to achieve a high user satisfaction and resource utilization ratio.

On the basis of this survey, in future we can Categorize the algorithms as decentralize or non-decentralize instead of static and dynamic. In particular DDFTP [12] algorithm which provides acceptable performance over other algorithms, lacks in storage utilization as it rely on full replication. So there is a scope of enhancement in this technique. By solving the existing issues and challenges, we will enter a new era of next generation computing through cloud computing technology.

REFERENCES

1. Wei Zhao, Yong Peng, Feng Xie, Zhonghua Dai "Modeling and Simulation of Cloud Computing: A Review", 2012 IEEE Asia Pacific Cloud Computing Congress (APCloudCC)
2. Cloud Computing Platform: AbiCloud; http://www.abiquo.com/
3. Cloud Computing Platform: Eucalyptus; http://www.eucalyptus.com/
4. Cloud Computing Platform: Nimbus; http://www.nimbusproject.org/
5. Cloud Computing Platform: OpenNebula;http://opennebula.org/
6. Pradeep K.Sinha, "Distributed operating Systems Concepts and Design" IEEE Computer Society Press.
7. Radojevic, B. and M. Zagar, "Analysis of issues with load balancing algorithms in hosted (cloud) environments." In proc.34th International Convention on MIPRO, IEEE,2011
8. Kolb, L., A. Thor, and E. Rahm, E, "Load Balancing for MapReduce-based Entity Resolution," in proc. 28th International Conference on Data Engineering (ICDE), IEEE, pp: 618-629,2012
9. Nishant, K. P. Sharma, V. Krishna, C. Gupta, KP. Singh, N. Nitin and R. Rastogi, "Load Balancing of Nodes in Cloud Using Ant Colony Optimization." In proc. 14th International Conference on Computer Modelling and Simulation (UKSim), IEEE, pp: 3-8, March 2012
10. T-Y., W-T. Lee, Y-S. Lin, Y-S. Lin, H-L. Chan and J-S. Huang, "Dynamic load balancing mechanism based on cloud storage" in proc. Computing, Communications and Applications Conference (ComComAp), IEEE, January 2012.
11. Wang, S-C., K-Q. Yan, W-P. Liao and S-S. Wang, "Towards a load balancing in a three-level cloud computing network," in proc. 3rd International Conference on. Computer Science and Information Technology (ICCSIT), IEEE, Vol. 1, pp: 108-113, July 2010
12. Al-Jaroodi, J. and N. Mohamed. "DDFTP: Dual-Direction FTP," in proc. 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), IEEE, pp:504-503, May 2011\
13. Mohamed, N. and J. Al-Jaroodi, "Delay-tolerant dynamic load balancing," in proc. 13th International Conference on High Performance Computing and Communications (HPCC), pp:237-245, September 2011.
14. 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