RESEARCH ARTICLE OPEN ACCESS

Algorithm of Load Balance in Cloud Computing

Jenia Sagar[1], Lekha Bhambhu[2]
Research Scholar[1], HOD[2]
Department of Computer Science and Engineering
Jan Nayak Ch. Devilal College of Engineering, Sirsa
Haryana – India

ABSTRACT
In the cloud computing, the process works on response and request generate by user and fulfil by server. The cloud computing has different types such IaaS, PaaS and SaaS used for providing the services to the users. The mail goal of Cloud computing is to provide services correctly and on right time. We need a load balancer for the cloud computing because random and concurrent access of services load the server which result to failure so to overcome this problem we need load balancers. Load balancing is an issue in Cloud Computing. In this, Virtual Server's load balancing algorithm has been proposed and implemented in order to achieve better response time and cost.
Keywords:- IaaS, PaaS, SaaS, Concurrent, load balancer

I. INTRODUCTION

Cloud Computing is a paradigm in which information is permanently stored in servers on the internet and cached temporarily on clients that include desktops, entertainment centres, table computers, notebooks, wall computers, hand-held's, sensors, monitors, etc.

Types of Cloud Services
Cloud computing offers a variety of ways for businesses to increase their IT capacity or functionality without having to add infrastructure, personnel, and software. There are different types of cloud computing and offer to businesses:

- **Web-based cloud services**

These services provide the web service functionality, rather than using fully developed applications. For example, it might include an API for Google Maps, or for a service such as one involving payroll or credit card processing.

- **IaaS (Infrastructure as a Service)**

The cloud provider provides the cloud consumer with essentially a virtual machine. The cloud consumer has the ability to provision processing, storage, networks, etc., and to deploy and run arbitrary software supported by the operating system run by the virtual machine. This service provides access to computing resource in a virtualized environment across a public connection through the internet. In the case of IaaS the computing resource provided is specifically that of virtualized hardware i.e. computing infrastructure. This includes such offerings as virtual server space, network connections, bandwidth, IP addresses and load balancers. Physically, the pool of hardware resource is pulled from a multitude of servers and networks usually distributed across numerous data centers. The client is given access to the virtualized components in order to build their own IT platforms. IaaS can be utilized by enterprise customers to create cost effective and easily scalable IT solutions where the complexities and expenses of managing the underlying hardware are outsourced to the cloud provider. If the scale of a business customer's operations fluctuate, or they are looking to expand, they can tap into the cloud resource as and when they need it rather than purchase, install and integrate hardware themselves.

- **SaaS (Software as a Service)**

This provides a given application to multiple tenants (Users), typically using the browser. SaaS solutions are common in sales, HR, and ERP. SaaS is accessed by users using a thin client via a web browser. SaaS has become a common delivery model for many business applications, including office & messaging software, DBMS software, management software, CAD software, Development software, virtualization, accounting, collaboration, customer relationship management (CRM), management information systems (MIS), enterprise resource planning (ERP), invoicing, human resource management (HRM), content management (CM) and service desk management.

- **PaaS (Platform as a Service)**

This is a variant of SaaS. This is very important service. The users can build their own application but implemented on the cloud provider's infrastructure.
In this PaaS, the consumer creates the software using tools and libraries from the provider. The consumer controls software deployment and configuration settings. The provider provides the networks, servers, storage, and other

services that are required to host the consumer's application.

PaaS offerings facilitate the deployment of applications without the cost and complexity of buying and managing the underlying hardware and software and provisioning hosting capabilities. There are various types of PaaS vendors.

PaaS offerings include facilities for application design, application development, testing, and deployment as well as services such as team collaboration, web service integration, database integration, security, scalability, storage, persistence, state management.

II. LOAD BALANCING

The Load Balancer systems allow creating an infrastructure able to distribute the workload balancing it between two or more Cloud Servers. It is easy to shape the infrastructure to allow it to sustain activity peaks, optimize the allocation of resources and ensure a minimal response time. Efficient provisioning of resources and scheduling of resources as well as tasks and goals will ensure:

- Improve the performance
- Maintain system stability
- Build fault tolerance system
- Accommodate future modification.
- Resources are efficiently utilized under condition of high/low load.
- Energy is saved in case of low load i.e. when usage of cloud resources is below certain threshold.
- Cost of using resources is reduced.

III. CLOUD ENVIRONMENT

The cloud environment can of two types and cloud components designed according to the environment of Cloud. Cloud computing can have either static or dynamic environment based upon how developer configures the cloud demanded by the cloud provider. As per the Environment, the algorithm will be implemented. E.g. Round Robin Algorithm is implemented in Static Environment.

a. Static Environment

In static environment the cloud provider installs homogeneous resources. The resources in the cloud are not flexible when environment is made static. In this scenario, the cloud requires prior knowledge of nodes capacity, processing power, memory, performance and statistics of user requirements. These user requirements are not subjected to any change at run-time. Algorithms need to be implemented to achieve load balancing in static environment cannot adapt to the run time changes in load. Static environment is easier to simulate but is not well suited for heterogeneous cloud environment. In this the resources are provisioned to the task on first-cum-first-serve (FCFS- i.e. the task that entered first will be first allocated the resource) basis and scheduled in time sharing manner. The resource which is least loaded (the node with least number of connections) is allocated to the task.

b. Dynamic Environment

In dynamic environment the cloud provider installs heterogeneous resources. The resources are flexible in dynamic environment. In this scenario cloud cannot rely on the prior knowledge whereas it takes into account run-time statistics. The requirements of the users are granted flexibility (i.e. they may change at run-time). Algorithm for achieve load balancing in dynamic environment can easily adapt to run time changes in load. Dynamic environment is difficult to be simulated but is highly adaptable with cloud computing environment. In dynamic algorithm the lightest server in the whole network or system is searched and preferred for balancing a load. For this real time communication with network is needed which can increase the traffic in the system. The load balancing algorithms allocates the resource with least weight to a task and takes into account node capabilities.

IV. CHALLENGES FOR LOAD BALANCING

There are some qualitative metrics that can be improved for better load balancing in cloud computing. The load balancers load the balance if the request exceeds the expected load. After the load balancing, the performance of the system should be maintained so that it revert proper information on right time. The below are the factors need to be considered in Cloud:

- **Throughput**

It is the total number of tasks that have completed execution for a given scale of time. It is required to have high through put for better performance of the system.

- **Associated Overhead**

It describes the amount of overhead during the implementation of the load balancing algorithm. It is a composition of movement of tasks, inter process communication and inter processor. For load balancing technique to work properly, minimum overhead should be there.

- **Fault tolerant**

It is the ability to perform load balancing by the appropriate algorithm without arbitrary link or node failure. Every load balancing algorithm should have good fault tolerance approach.

- **Migration time**

It is the amount of time for a process to be transferred from one system node to another node for execution. For better performance of the system this time should be always less.

- **Response time**

In Distributed system, it is the time taken by a particular load balancing technique to respond. This time should be minimized for better performance.

- **Resource Utilization**

It is the parameter which gives the information within which extant the resource is utilized. For efficient load balancing in system, optimum resource should be utilized.

- **Scalability**

It is the ability of load balancing algorithm for a system with any finite number of processor and machines. This parameter can be improved for better system performance.

- **Performance**

It is the overall efficiency of the system. If all the parameters are improved then the overall system performance can be improved.

V. PROBLEM FORMULATION

In the cloud computing environment, the resource allocation and workload management is the main aspect for provide the better performance to the end-users. The cloud is the distributed environment with distributed resources for provide the better response time to the end users and request is managed by the distributed system but the resource should be allocate to the efficient processor for high availability of the services. The process is not handling by the single system if number of request exceeds and processing will not be handling by the processors. There is Scalability issue which can be achieved through a combination of multi-user service hosting, application virtualization, and both time and CPU multiplexing and load balancing. The task is to manage the running request by distribute the load in its Node processors. The number of load balancing algorithm can be implemented for the better response time to the end-users.

Research Methodology

1) Study of existing load balancing Algorithms in Cloud.
2) Identify and analyze the benefits of load balancing techniques.
3) Research on the load balancing real time issues.
4) Choose the reliable load balancing technique in Cloud.
5) Flow Development of new research and its Implementation in any of the language for making it understandable steps.
6) Analyze the results.

VI. OBJECTIVES

In the research scenario, the load balancing algorithm will be implemented. There are number of load balancing algorithms in cloud environment for resource allocation but the selection of best algorithm is difficult task. Load balancing is a core and challenging issue in Cloud Computing. for the high availability of the services to the end-user, there is need to be manage Cloud computing resources efficiently and gain the maximum profits with efficient load balancing algorithm. The load balancing provides the facility to handle the faults that appear in typical Cloud computing infrastructures and faults impact user's applications.

- Theses faults can be handled in an efficient and cost-effective manner.
- Analysis of the cloud computing service and resource issues.
- Study and select the best algorithm for load balancing in cloud Environment.
- Implement the algorithm in any of language.
- Analyze results.

VII. CONCLUSION

This paper has been proposed that the server implies that the virtual machine or virtual server which will serve and process the requests to the users. The multiple requests will be generated and processed in queue. The load balancer will identify the load on each server. If the server's resources are free, then the request will be allocate to the particular server. The prime condition is that the request should be concurrent because the load balancing aim is to balance the load on particular condition otherwise this will handle the request without considering the other pending request. The Proposed algorithm finds the expected response time of each resource (VM) and sends the ID of virtual machine or severs having minimum response time to the data center controller for allocation to the new request. This has been conclude that if the algorithm selects an efficient virtual machine then it affects the overall performance of the cloud Environment and also decreases the average response time.

REFERENCES

[1] Lei Lu et al.(2013), "Predictive VM consolidation on multiple resources: Beyond load balancing",IEEE, Page(s): 1 – 10.
[2] Huankai Chen (2013), "User-priority guided Min-Min scheduling algorithm for loadbalancing in cloud computing",IEEE, Page(s): 1 – 8

[3] Ajit, M. ; Vidya, G (2013), "VM level load balancing in cloud environment",IEEE, Page(s): 1 – 5
[4] Al-Rayis, E. ; Kurdi, H. (2013), "Performance Analysis of Load Balancing Architectures in Cloud Computing",IEEE,Page(s): 520 – 524.
[5] P.G.J. ; Sharmila, J.(2013), "Live VM migration techniques in cloud environment — A survey Leelipushpam", IEEE, Page(s): 408 – 413
[6] Tushar Desai, Jignesh Prajapati (2013), "A Survey of Various Load Balancing Techniques And Challenges In Cloud Computing",IJST.
[7] Rajesh George Rajan, V.Jeyakrishnan (2013), "A Survey on Load Balancing in Cloud Computing Environments", (IJARCCE).
[8] MR.Manan D. Shah, MR.Amit A. Kariyani, MR.Dipak L. Agrawal (2013),"Allocation Of Virtual Machines In Cloud Computing Using Load Balancing Algorithm", (IJCSITS).
[9] Achar, R. et al. (2013), "Load balancing in cloud based on live migration of virtual machines", IEEE, Page(s): 1 – 5.
[10] Sahu, Y. (2013), "Cloud Server Optimization with Load Balancing and Green Computing Techniques Using Dynamic Compare and Balance Algorithm", IEEE Page(s): 527- 531.
[11] Al-Rayis, E. ; Kurdi, H. (2013), "Performance Analysis of Load Balancing Architectures in Cloud Computing", IEEE, Page(s): 520- 524
[12] Prabavathy et al (2013), "A load balancing algorithm for private cloud storage", IEEE, Page(s): 1 – 6
[13] Arora, V. ; Tyagi, S.S. (2014), "Performance evaluation of load balancing policies across virtual machines in a data center", Page(s): 384 – 387.