An Approach for Load Balancing in Cloud Computing Using JAYA Algorithm

Subhadarshini Mohanty, Department of Computer Science and Engineering, Siksha O Anusandhan University, Bhubaneswar, India
Prashanta Kumar Patra, Department of Computer Science and Engineering, College of Engineering and Technology, Bhubaneswar, India
Mitrabinda Ray, Department of Computer Science and Engineering, Siksha O Anusandhan University, Bhubaneswar, India
Subasish Mohapatra, Department of Computer Science and Engineering, College of Engineering and Technology, Bhubaneswar, India

ABSTRACT
Cloud computing is a jargon in the era of information technology. It acts as a metaphor for the internet. Still, it possesses several challenges related to automated resource provisioning, security, event content dissemination, server consolidation, virtual machine migration. Load balancing is one of the critical challenges in the cloud that are faced by the enterprises. Here the basic objective of load balancing is to minimize response time, data center service request time, and improve the overall performance of the system. In this article, a JAYA algorithm is used for load balancing in a cloud which uses less control parameters and provides a better optimized result. Comparisons are made with other evolutionary approaches to observe the efficiency of the proposed algorithm.

KEYWORDS
Cloud Computing, Data Center, JAYA Algorithm, Load Balancing, Virtual Machine

1. INTRODUCTION
From the past decades, internet demands are increasing exponentially. Users access the internet to store and retrieve their data. Apart from data storage, users require different services on demand basis with better quality of service. Cloud Computing concept is developed in which cloud acts as a metaphor for the internet (Hwang et al., 2013) and computing acts as utilizing the internet technology to complete the task. It offers services such as server, storage and application on requirement basis. Cloud offers three types of services such as SaaS, PaaS, IaaS. In SaaS, on demand basis cloud provider provides their services to the end user. Similarly, in PaaS the users get the liberty to develop its own application, because it offers different computing platforms like operating system, web server, database, programming language. IaaS provider provides infrastructure and resources in a virtualized manner (Zhang et al., 2010). The cloud consists of three major components that is end users, broker and cloud provider. Here the end user request for the services, after getting the request broker, who acts as an interface between end user and cloud provider negotiate with different cloud vendors by considering the need of end user. Recently the IT industries are focusing to enhance the quality of service in cloud computing by optimizing the time and cost in various measures due to the flexible cloud architecture.

DOI: 10.4018/IJITWE.2019010102

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
Cloud computing architecture consists of four layers. The inner layer is the hardware layer; next to the inner layer is the infrastructure layer. Platform layer is above the infrastructure layer and application layer is the outermost layer. The characteristics of each layer are as flows:

1. **Hardware Layer**: This layer handles the physical resources of the cloud, including physical servers, routers, switches, power, and cooling systems. In practice, the hardware layer is typically implemented in data centers. A data center usually contains thousands of servers that are organized in racks and interconnected through switches and routers. Hardware configuration, fault tolerance, traffic management, and power management are typical the issues in the hardware layer.

2. **Infrastructure Layer**: The infrastructure layer is an essential component of cloud architecture. It is also known as the virtualization layer. This layer creates a pool of storage and computing resources by using virtualization technologies such as Xen, Kernel-based Virtual Machine (KVM), and VMware. They are helpful in dynamic resource assignment in a cloud system.

3. **Platform Layer**: This layer is on the top of the infrastructure layer. It consists of operating systems and application frameworks. The purpose of the platform layer is to minimize the burden of deploying applications directly into VM containers. For example, Google App Engine runs on the platform layer to provide application programming interface (API) support for implementing storage, database, and business logic of typical web applications.

4. **Application Layer**: Cloud application layer is different from traditional applications, and it can leverage the automatic-scaling feature to achieve better performance, availability, at a lower operating cost. Each layer is loosely coupled with the layers above and below, allowing each layer to evolve separately. The architectural modularity allows cloud computing to support a broad range of application requirements while reducing management and maintenance overhead.

Although cloud computing is the mega trend in recent days, still the issues in this computing paradigm include automated service provisioning, resource management, task scheduling, Virtual Machine Migration, Server Management, Privacy and Security etc.

### 1.1. Load Balancing

Load balancing is buzz word among the software professional at the enterprise level. Thousands of people access the servers for different purpose day in and day out. Before introduction of the concept of load balancing, it was observed that with capacity utilization of one server the load was transferred to another server, that caused congestion, consequent delay and degraded the performance of the systems and over all, the customer satisfaction was less. The main focus of cloud computing is user’s satisfaction. So, cloud computing introduced a load balancing technique where load balancer, a device is used to efficiently distribute the workload across a group of backend servers, as shown in Figure 1. Usually load balancing includes the dedicated hardware or software. Load balancing provides a distinct internet service from a collective group of servers. The objective of load balancing is to minimize the makespan, response time and maximize the resource utilization and performance of the system.

The primary opportunities of load balancing include location independence, availability, reliability, and optimized cost. To achieve these objectives, the tasks need to be scheduled properly. The scheduling principle should aim to reduce the amount of data transfer with minimum cost and ensure balanced distribution of tasks as per processing capability. Thus, proper task scheduling is an important issue for which various optimization techniques have been experimented. Task scheduling based on load balancing is essential in preventing bottleneck of the system which may occur due to load imbalance. Load balancing helps in providing quality of service by optimizing the response time. It ensures that every computing resource is distributed efficiently and equitably. Cloud computing system achieves dynamic load balancing using virtualization technology. It is possible to remap virtual machine and physical resources dynamically with flexible resource allocation and reallocation using VM migration. VM migration is achieved by moving the live VMs on execution from one
Related Content

Image Mining: A Case for Clustering Shoe prints
[www.igi-global.com/article/image-mining-case-clustering-shoe/2642?camid=4v1a](www.igi-global.com/article/image-mining-case-clustering-shoe/2642?camid=4v1a)

Outsourcing Issues in Web Development
[www.igi-global.com/chapter/outourcing-issues-web-development/29586?camid=4v1a](www.igi-global.com/chapter/outourcing-issues-web-development/29586?camid=4v1a)

A Framework Towards Semantic Web Service Composition Based on Multi-Agent System
[www.igi-global.com/article/framework-towards-semantic-web-service/2652?camid=4v1a](www.igi-global.com/article/framework-towards-semantic-web-service/2652?camid=4v1a)