# A Load Balancing Strategy for Cloud Computing Environment

Raza Abbas Haidri[1], C. P. Katti[2], P. C. Saxena[3]
School of Computer & System Sciences
Jawaharlal Nehru University
New Delhi, India
[1]razaabbas.amu@gmail.com, [2]cpkatti@yahoo.com, [3]premchand_saxena@yahoo.com

***Abstract*—Information and communication technology has been growing rapidly for the last few decades. In today's age of information and globalization massive computing power is desired to generate business insights and competitive advantage [1]. Cloud Computing is an emerging technology and has attracted a lot of attention in both commercial and academic spheres. It is growing very fast and provides an alternative to conventional computing. There are cloud service providers who provide services in a flexible manner. The main goal of cloud service provider is to establish an efficient load balancing algorithm which ensures a fair distribution of loads among virtual machines and better resource utilization. In this paper we present a heuristic based load balanced scheduling model for efficient execution of tasks. The proposed model balances the loads coming from several users among datacenters and hence it offers better resource utilization and high availability in the form of improved response time and turnaround time. The proposed algorithm is implemented using CloudSim simulator and the result shows that the proposed algorithm outperforms to existing algorithms on similar objectives.**

**Keywords- Cloud Computing, Load Balancing, Turnaround Time, Response Time, CloudSim**

## I. INTRODUCTION

Some decades ago small and medium enterprises (SMEs) could not perform high performance computing (HPC) because they were unable to afford huge up-front cost of dedicated supercomputers. However, with the advent of Cloud Computing the cost of HPC has reduced. The fundamental principle of Cloud Computing is to shift the computing from traditional desktop to the internet that is moving computation, services and data off-site to an external, internal, location transparent centralized contractor. The Cloud Computing model is often referred as [2-3] "pay as you go model" that is users essentially rent virtual resources and pay for what they use. According to the abstraction level [3-5] of the services cloud delivery models are of three types as Infrastructure as a Service (IaaS) in which in which services are offered to users in the form of hardware platform where user can deploy their Virtual Machines (VMs), software platforms to support their application and the application itself, Platform as a Service (PaaS) which is a software platform for hosting application is already installed in an infrastructure and user uses this platform to develop their specific application [6], and Software as a Service (SaaS) is last level in which actual application is offered to users. A Cloud Deployment Model [3, 5] is of four kinds. Public cloud is situated on the premises of the cloud provider, Private cloud is dedicated to particular organization, Community clouds offers services to organizations that have common functions and purposes and Hybrid cloud is a Combination of public, private and community clouds forms.

There are several challenges in Cloud Computing that need to be resolved before exploiting the features this technology [7]. Some challenges include security issues [8, 9] legal and compliant issues [10], performance and QoS [10], interoperability issues [10], load balancing [11], data management issues [12]. Load Balancing is one of the primary concerns in Cloud Computing. As we know cloud platform can be quickly scaled up and down at any point of time. So the numbers of user's requests can join to and leave from the cloud during the execution of the applications. In this dynamic environment an efficient load balancing is required to minimize the response time, lower network congestion, avoid the interruption of services, limited energy consumption and provides high availability which means continuity of services when components becomes non-responsive. Load Balancing is used to implement failover mechanism across different datacenters to improve response time, turnaround time and maintains the system stability and performance.

Rest of the paper is organized as follows. The various load balancing algorithms are being discussed in section II. Section III describes problem formulation and parameter estimation of proposed model. Section IV presents algorithm and illustration. Section V discusses simulation framework and simulation results. The paper ends with conclusions and future scopes covered in section VI.

## II. RELATED WORK

In Load Balanced Scheduling is an attempt to balance loads among the nodes of distributed systems in a way the progress of all processors proceeds at approximate same

978-1-4799-3140-8/14/$31.00 ©2014 IEEE