Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 50 (2015) 427 – 432

Procedia Computer Science

2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)

An in-depth analysis and study of Load balancing techniques in the cloud computing environment.

Geethu Gopinath P P[1], Shriram K Vasudevan[2]

Department of Computer Science and Engineering [1]
Amrita School of Engineering,
Amrita Vishwa Vidyapeetham (University),
Ettimadai, Coimbatore—641112,
geetkrish90@gmail.com

Asst professor (senior)[2]
Department of Computer Science and Engineering,
Amrita School of Engineering,
Amrita Vishwa Vidyapeetham (University),
Ettimadai, Coimbatore—641112,
Kv_shriram@cb.amrita.edu *ph: 08939918562*

Abstract

In the cloud computing paradigm, load balancing is one of the challenges, With Tremendous increase in the users and their demand of different services on the cloud computing platform, fruitful or efficient usage of resources in the cloud environment became a critical concern. Load balancing is playing a vital role in maintaining the rhythm of Cloud computing. The performance metrics of load balancing algorithms in cloud are response time and waiting time. In this paper we mainly focus on two load balancing algorithms in cloud, Min-Min and Max-Min algorithm.

Peer-review under responsibility of scientific committee of 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)

Keywords: Min-Min, Max-Min, Response time, Load Balancing, Batch mode scheduling, Immediate mode scheduling, Make span, Minimum completion time, Minimum execution time

Peer-review under responsibility of scientific committee of 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)
doi:10.1016/j.procs.2015.04.009

1. Introduction

Load balancing [1] is a method that distributes the workload among diverse nodes in the given environment such that it ensures no node in the system is over loaded or sits idle for any instant of time. An efficient load balancing algorithm will make sure that every node in the system does more or less same volume of work. The responsibility of load balancing algorithm is that to map the jobs which are set forth to the cloud domain to the unoccupied resources so that the overall available response time is improved as well as it provides efficient resource utilization. Balancing the load became one of the crucial concerns in cloud computing since we cannot predict the number of requests that are issued at each second in cloud environment. The unpredictability is due to the ever changing behaviour of the cloud. The main focus of load balancing in the cloud domain is in allocating the load dynamically among the nodes in order to satisfy the user requirements and to provide maximum resource utilization by assorting the overall available load to distinct nodes.

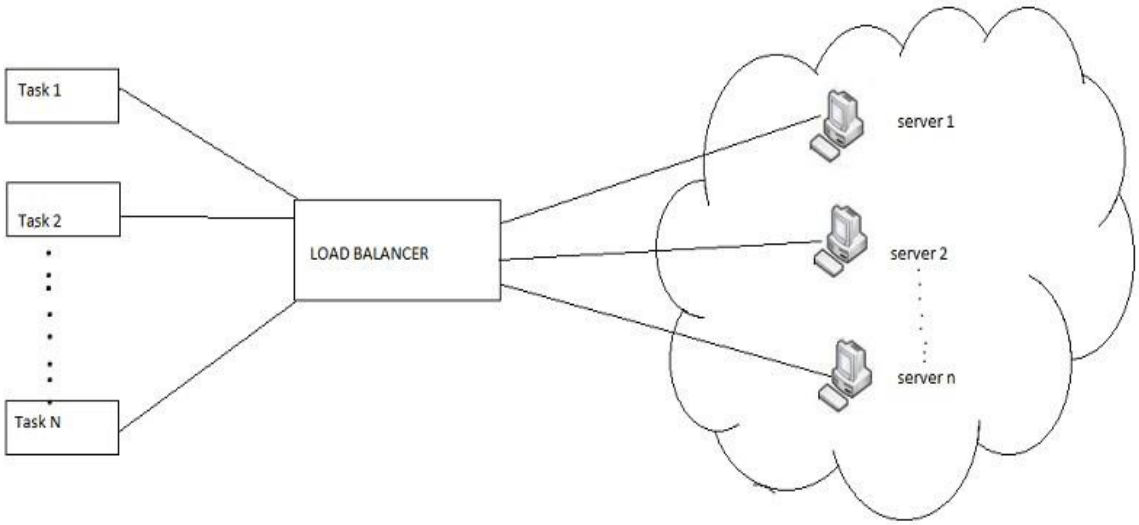

Fig 1. Diagram for load balancing

2. Demand of Load Balancing in Cloud Environment

Load balancing [1] is a method such that it assigns the workload equally among all the available nodes which are present in the system. Higher user satisfaction is the motto behind load balancing. As the number of user's as well as their demands are increasing day by day, the clouds should provide the services to the customers with their at most satisfaction.

An appropriate or an ideal load balancing algorithm help in making use of the available resources most favourably, thereby ensuring no node is over loaded or under loaded. Load balancing enables scalability, avoids bottlenecks and also reduces time taken to give the respond. Many load balancing algorithm [2] have been designed in order to schedule the load among various machines. But so far there is no such ideal load balancing algorithm has been developed which will allocate the load evenly across the system. It has been proved that allocating the tasks evenly across the system is considered to be an NP complete problem [7].