Chapter 28
Comparative Study for Different Provisioning Policies for Load Balancing in Cloudsim

Layla Albdour
University of Jordan, Amman, Jordan

ABSTRACT

Distributing application requests across applications located in different datacenters with in cloud equally must be provided by cloud load balancing. In this paper, we compare different provisioning policies within cloud for virtual machines and workloads, where we are focusing on how to distribute the processing power between virtual machines and how to distribute workload among virtual machines. Cloudsim is the simulation plate form used to test the different distributions scenarios to check the performance on makespan, average turnaround time, bandwidth utilization and CPU utilization. Result showed the difference in performance between the three tested provisioning schemes, where the space-shared gives better readings for the selected performance metrics.

INTRODUCTION

A Cloud is a logical structure consists of managed pool of computing resources, thus cloud computing is the process of delivering on demand and convenient access over the internet to that pool (Vaquero et al., 2008). It is an effective scalable computation resources such as hardware, platform and software, which are provided to the user on-demand (Zhou et al., 2012). Due to the rapid development in the internet speed along with the more security guarantees, cloud computing became the tool of choice for many enterprises and IT industry (Jouini & Rabai, 2016). A big number of organizations are using cloud computing to accomplish their needs, for example on line games, advertising, viral marketing activities and many others (Chow et al., 2009). With the increasing using mobile devices a new branch of cloud computing arises which is, Mobile Cloud Computing. Where we transparently using cloud services from smart phones and any mobile device (Dinh et al. 2013).

DOI: 10.4018/978-1-7998-5339-8.ch028

Due to the high number of cloud computing users. Cloud computing providers are highly challenging to provide reliable services with a guaranteed Quality of Service (QoS) level that is defined in SLA. SLA is the Service Level Agreement between a customer and provider, it defines the expected level of services and some Quality of Service attributes such as: delay and response time (Patel et al., 2009).

Load balancing is a critical issue, because a cloud provider easily increases the processing power of its cloud for instance, but how to balance load between a cloud users? Is a main concern and an open research question (Randles et al., 2010). Amazon for instance lately incorporated the Zeus (Khiyaita et al., 2012) Simple Load Balancer in EC2 that aims to balance load between users. The main goals of load balancing in cloud computing is (Rimal et al., 2009):

1. Dividing workloads and computing resources in a cloud computing environment.
2. Manage application or workload demands by allocating resources among multiple computers, networks or servers.
3. Guarantee a fair service provisioning, with no starvation.
4. Prevent the situation where some processors are high loaded and others are low loaded.

Business and technical goals for load balancing in cloud computing are studied in [1], cloud load balancing parameters is the same as load balancing global server parameters that are:

1. Response time for applications
2. User location
3. Service availability
4. Time of day
5. Data center capacity, the current and the total capacity

This paper investigates load balancing methods for cloud computing. The authors are using Cloudsim (Calheiros et al., 2011) as the research platform, because Cloudsim provides different load balancing levels for cloud components. Cloudsim is a new simulation tool that provide a high flexibility for providing space-shared and time-shared allocation for processing power between virtualized virtual machines (services).

This paper is organized as follows: in section two we review some related work on load balancing, section three we introduce simulation in Cloudsim. In section four we introduce provisioning policies for load balancing in Cloudsim, section five introduce the simulation results, where in the last section we conclude the paper.

RELATED WORK

Load balancing have different implementation policies (Karatza,2001): (1) Information policy that defines what, when and from where work load should be collected; (2) Triggering policy that defines the appropriate time period to start a load balancing; (3) Resource kind policy that makes load balancing on a specified resource such as server or receiver of tasks;(4)Location policy that select the suitable server or partner;(5) selection policy that determines the migrating tasks from an overloaded server or resource to an idle one.