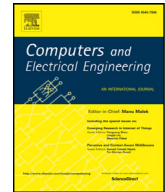




Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

A virtual machine scheduler based on CPU and I/O-bound features for energy-aware in high performance computing clouds[☆]

Felipe Fernandes^{a,b,*}, David Beserra^c, Edward David Moreno^c, Bruno Schulze^b, Raquel Coelho Gomes Pinto^a

^aInstituto Militar de Engenharia (IME), Rio de Janeiro, RJ, Brazil

^bLaboratório Nacional de Computação Científica (LNCC), Petrópolis, RJ, Brazil

^cUniversidade Federal de Sergipe (UFS), Aracaju, SE, Brazil

ARTICLE INFO

Article history:

Received 31 August 2015

Revised 4 August 2016

Accepted 1 September 2016

Available online xxx

Keywords:

Green cloud computing

HPC

VM scheduler

Energy-aware

ABSTRACT

A Virtual Machine (VM) scheduler for homogeneous High Performance Computing (HPC) cloud environments is presented in this paper. This scheduler considers each VM workload type (CPU or I/O-bound) to decide on its allocation. Scheduler is able to reduce energy consumption, as well as SLA violations on this cloud environment, avoiding performance losses by allocating simultaneously VMs which run different types of tasks.

The scheduler method was validated through simulations conducted with the CloudSim framework. Two synthetic benchmarks representing both workload types were run previously in VMs in order to obtain basic data to design the scheduler. Thus, it was possible to implement a scheduling method that employs a VM allocation policy based on features of each application. Results showed that knowing the specifications and characteristics of environment may contribute to a better usage of resources, leading to an increased level of services availability and, finally, reducing problems caused by competition in resource usage.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

A currently well-known advantage of cloud computing is the ability to provide HPC infrastructure to people and organizations with no physical space, financial resources or technical and operational capacity to implement and maintain their own infrastructure aimed at HPC.

This enables even small start-ups to have access to a large computing power with no need of a complex datacenter environment, which demands high power consumption, a higher burden on the company and also causes damage to the environment. By using hosted cloud infrastructure, this responsibility can be transferred to specialized providers.

On the other hand, infrastructure providers are also organizations that face the challenge of reducing their expenses but keeping the quality of services and attractive prices. In this context, datacenters are focus of organizations in terms of

[☆] Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. F. Xhafa.

* Corresponding author.

E-mail addresses: felipejf@lncc.br (F. Fernandes), dw.beserra@gmail.com (D. Beserra), edwdavid@gmail.com (E.D. Moreno), schulze@lncc.br (B. Schulze), raquel@ime.eb.br (R.C.G. Pinto).

<http://dx.doi.org/10.1016/j.compeleceng.2016.09.003>

0045-7906/© 2016 Elsevier Ltd. All rights reserved.

infrastructure and optimization efforts of resource usage [1]. Energy cost of these environments is high, requiring large computational workload to keep them operating efficiently avoiding idle periods that generate unnecessary energy consumption [2]. Poor managing of datacenters is the greatest barrier to achieve sustainability. One example is the traditional datacenter that keeps web applications. These applications commonly have high availability and sporadic use, causing energy waste due to idleness. However, sometimes peak load may occur and demands for resources may exceed the available amount. On these occasions, there is a reduction in application performance and a resulting increase in energy consumption, affecting both final consumers and service providers.

In contrast to traditional datacenters, cloud computing model allows greater energy efficiency to be achieved by using techniques such as virtualization and consolidation workload, providing better control of computing resources in relation to their demand, and avoiding their waste [3]. In the context of a cloud computing environment, scheduling and fixing VMs process across physical servers gained prominence, and a good scheduler can provide a significant energetic consumption reduction in these environments [2,4].

Thus, in this work we present a VM scheduling algorithm and its development process for a homogeneous cloud environment focused particularly on resource provision for HPC applications. This algorithm improves performance and energy-consumption ratio for computer applications of specific patterns, as well as reduces the occurrences of SLA violations, considered one of the main concerns of cloud computing service providers [5]. Our algorithm was developed based on energy consumption of servers in consequence of workload type running at the allocated VMs.

This work is structured as follows: in Section 2 we explain some fundamental concepts for understanding this work. In Section 3 we present a selection of related works. In Section 4 we describe the adopted experimental methodology. In Section 5 we show the scheduling method developed and in Section 6 we show our algorithm validated by simulation and its implementation outlooks. Finally, in Section 7 we present conclusions and ideas for future works.

2. High performance computing, green clouds and scheduling

HPC is a generic term to applications that are, in nature, computationally intensive or data intensive [6]. Traditionally, HPC applications run in dedicated infrastructures in which resources are locally owned, with private access and composing complex structures with quasi-static configurations. This type of in-house infrastructure is expensive and difficult to setup, maintain, and operate [7].

As a low-cost and flexible alternative, HPC users have utilized cloud infrastructures instead of in-house clusters; with this, they can enjoy cloud benefits, such as high availability (HA), operating system (OS) customization, lower wait queueing time, and maintenance cost reduction [8]. Some cloud providers, such as Amazon, already offer HPC cloud services; allowing building clusters over its infrastructure. To offer and manage their resources in a practical way, cloud providers commonly rely on virtualization technology [6].

A growing energetic consumption is expected of datacenters due demand for cloud services, generating an additional amount of carbon dioxide (CO₂) emissions in atmosphere [2]. On the other hand, this increase can be reduced with utilization of computational clouds which can be run under virtualized infrastructures [9]. This is possible because virtualization allows consolidation of workloads of many users simultaneously and in the same physical space, implying in minimal spatial utilization and heat emission; minimizing the use of cooling system on the environment [10]. These clouds focused in reducing the CO₂ emissions of their host datacenters are known as Green Clouds.

However, energetic consumption reduction cannot be accompanied with a corresponding performance reduction in the environment, as well in the service level agreed with customers [11]. Therefore, other aspects should be considered in a green cloud project. Good examples of these aspects are the network infrastructure that connects cloud servers [12] and the size of the files manipulated in these environments [9]. Facing the challenge of the implementation of green clouds is a colossal task, given the number of components and environments involved. To correctly analyze all challenges and treat them in an effective manner, it is primordial to establish metrics and benchmark sets that allow to relate the performance issues to environmental issues; both in levels of energy consumption and in greenhouse gas emission forecast [11].

Choice of metrics must take into account workloads that will be performed more frequently in the environment and what types of computational resources these applications will predominantly use. In environments destined for HPC purposes, workloads intensively use only one type of resource during its runtime (See Fig. 1). Most common workload used in these environments are CPU-bound workloads which intensively use CPU or I/O bound workloads which most intensively use some I/O device.

A misallocation of resources can lead to situations in which many processes of the same type are competing for a single available resource, while other types of resources remain idle. This simultaneously generates overload and underutilization situations, increasing waiting time, power consumption and causing SLA violations. To ensure a better workload distribution to available resources, schedulers are employed to manage access to resources of their consumers [13].

There are two main types of schedulers in cloud computing environments: cloud manager schedulers and virtualization tool schedulers. The first one operates in scheduling and allocation of VM instances on servers while the second one acts directly on the OS of individual servers, sharing resources of an individual server among all VMs allocated in it [14]. In this paper, we have described a cloud manager scheduler.

Download English Version:

<https://daneshyari.com/en/article/4955354>

Download Persian Version:

<https://daneshyari.com/article/4955354>

[Daneshyari.com](https://daneshyari.com)