Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 83 (2016) 870 – 877

Procedia Computer Science

The 6th International Conference on Sustainable Energy Information Technology (SEIT 2016)

EATS: Energy-Aware Tasks Scheduling in Cloud Computing Systems

Leila Ismail[a,*], Abbas Fardoun[b]

[a]*College of Information Technology, UAE University, Al-Ain, UAE*
[b]*College of Engineering, UAE University, Al-Ain, UAE*

Abstract

The increasing cost in power consumption in data centers, and the corresponding environmental threats have raised a growing demand in energy-efficient computing. Despite its importance, little work was done on introducing models to manage the consumption efficiently. With the growing use of Cloud Computing, this issue becomes very crucial. In a Cloud Computing, the services run in a data center on a set of clusters that are managed by the Cloud computing environment. The services are provided in the form of a Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). The amount of energy consumed by the underutilized and overloaded computing systems may be substantial. Therefore, there is a need for scheduling algorithms to take into account the power consumption of the Cloud for energy-efficient resource utilization. On the other hand, Cloud computing is seen as crucial for high performance computing; for instance for the purpose of Big Data processing, and that should not be much compromised for the sake of reducing energy consumption. In this work, we derive an energy-aware tasks scheduling (EATS) model, which divides and schedules a big data in the Cloud. The main goal of EATS is to increase the application efficiency and reduce the energy consumption of the underlying resources. The power consumption of a computing server was measured under different working load conditions. Experiments show that the ratio of energy consumption at peak performance compared to an idle state is 1.3. This shows that resources must be utilized correctly without scarifying performance. The results of the proposed approach are very promising and encouraging. Hence, the adoption of such strategies by the cloud providers result in energy saving for data centers.

Peer-review under responsibility of the Conference Program Chairs

Keywords: Green Computing; Energy Efficiency; Energy Management; Scheduling; Cloud Computing; Performance

1. Introduction

Cloud computing[1,2,3] is an emerging technology for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. It is

* Corresponding author:
E-mail address: leila@uaeu.ac.ae

Peer-review under responsibility of the Conference Program Chairs
doi:10.1016/j.procs.2016.04.178

composed of three service models: Cloud Software as a Service (SaaS), Cloud Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). The Cloud SaaS provides the users with applications that they can run and get results. The Cloud PaaS provides the users with the possibility to deploy applications onto the cloud. The Cloud IaaS provides the users with the capability to provision processing, storage, and networks for running their applications. The services run in a data center on a set of clusters that are managed by the Cloud Computing environment. The Cloud computing is promising for high performance computing[4,5] of many scientific and engineering applications; for instance for the purpose of Big Data processing. Both performance and quality of services are essential. Many researchers were aiming at increasing the performance of applications by selecting adequate resources without paying attention to energy consumption. However, the energy consumption of the data center is equally a crucial issue for the both the entity hosting the data center as the corresponding bill increases in cost, and for the environment. For instance, Data centers in European Union are estimated to be about 1% in 2005 (including cooling) and it is estimated that it has reached 2.8% in the US. They are estimated to use 3% of the total electricity production in Europe, hence responsible for the same percentage of $CO2$ emission[6].

The data centers power consumption is expected to reach 100 TWh in year 2020[6]. The cost of the power consumed by a server during its lifetime could surpass its cost[7]. When a data center is built, a provision of power energy consumption is needed for building the infrastructure; i.e., the facilities, such as the cooling system and the electrical facilities. The facilities' cost is linearly dependent on the maximum utilization of the computing nodes; i.e., the servers operating at their peak performance. This cost is divided into a facility cost and a power consumption cost. Therefore, in order to amortize that cost, it is important to maximize the utilization of the computing machines. Underutilized resources become costly as a fraction of the total cost of ownership[8].

In Cloud computing, the hardware allocation is hidden to users. However, the distribution of users' applications, which is part of a Cloud Computing environment should take into consideration the energy efficiency of the cloud. Cloud server machines should not be overloaded, at risk of high power consumption, and execution inefficiency. They should not be underloaded, as the energy consumption of the computing facilities, designed for an efficiently-used data center, increases compared to the usage pattern[9]. This problem becomes crucial with the emergence of Big Data analysis in the Cloud[10], as infrastructure should be used in an energy-optimal way.

The goal of this work is to distribute a Big data for distributed processing in the Cloud in a way to decrease the overall energy consumption of the Cloud, without scarifying the application's performance. Therefore, we develop an energy-aware task scheduler (EATS) whose aim is to reduce both the power consumption of the utilized resources and the processing time of an application. We are considering applications of type divisible load applications[11]. Divisible load applications are a class of applications that can be divided into independent tasks and assigned to distributed resources with no synchronization and inter-tasks communication. Divisible load occurs in many scientific and engineering applications to analyze and process Big data, such as search for a pattern, compression, generic search applications, multimedia and video processing, image processing and data mining. Distribution of processing is meant to increase the performance of the distributed application compared to its sequential execution. In our previous works[11,12], we developed a scheduler to increase the performance of divisible load applications in a Cloud computing environment. The scheduler follows a linear-model approach and did not take into account the energy consumption of the cloud. In this work, we develop a non-linear-programming based scheduling model which aims at optimizing both the performance of the application and the energy consumption of the underlying resources. Our experiments show that the ratio of the energy consumption of a fully utilized server to its power consumption when it is idle is 1.3. This shows that idle, and underutilized resources consume a large amount of energy compared to that of full use. Our model takes into consideration this observation to optimize energy consumption during distributed computing.

The rest of the paper is structured as follows. Section 2 provides the system model of our scheduling approach. The scheduling algorithm is presented in section 3. Section 4 overviews related works and compare them to our approach. Our experiments and their evaluations are described in section 5. Section 6 concludes our work and highlights future works.