Computers and Electrical Engineering 47 (2015) 186–203

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Resource management in cloud computing: Taxonomy, prospects, and challenges ☆

Saad Mustafa [a], Babar Nazir [a,*], Amir Hayat [b], Atta ur Rehman Khan [a], Sajjad A. Madani [b]

[a] Department of Computer Science, COMSATS Institute of Information Technology, Abbottabad, Pakistan
[b] Department of Computer Science, COMSATS Institute of Information Technology, Islamabad, Pakistan

ARTICLE INFO

Article history:
Received 28 October 2014
Received in revised form 29 July 2015
Accepted 29 July 2015
Available online 12 August 2015

Keywords:
Cloud computing
Resource management
Resource allocation
Virtual machine migration
Hybrid cloud
Mobile cloud computing

ABSTRACT

Cloud computing has emerged as a popular computing paradigm for hosting large computing systems and services. Recently, significant research is carried out on Resource Management (RM) techniques that focus on the efficient sharing of cloud resources among multiple users. RM techniques in cloud are designed for computing and workload intensive applications that have different optimization parameters. This study presents a comprehensive review of RM techniques and elaborates their extensive taxonomy based on the distinct features. It highlights evaluation parameters and platforms that are used to evaluate RM techniques. Moreover, it presents design goals and research challenges that should be considered while proposing novel RM techniques.

1. Introduction

During the past few years, the computational world has evolved tremendously due to constant increase in demand for high-end computational devices [1]. This evolution resulted in the emergence of new computational paradigms, such as cluster computing, grid computing, and cloud computing. Among these paradigms, cloud computing has gained significant popularity [2]. Cloud computing is offered in three forms, namely: (c) public cloud, (b) private cloud, and (c) hybrid cloud [1,3,4]. Moreover, different set of services are provided that can be broadly placed into three categories, namely: (a) Software as a Service (SaaS), (b) Infrastructure as a Service (IaaS), and (c) Platform as a Service (PaaS) [1]. Cloud computing is based on Service-Oriented Architecture (SOA) that uses concepts of virtualization and distributed computing [1]. In cloud computing SOA, access to shared pool of resources is provided via network and resources on hand can be configured according to users' demand [3].

One of the key aspect of cloud computing and virtualization is Resource Management (RM). RM is a process that deals with the procurement and release of resources [2]. Virtualization techniques are used for flexible and on-demand resource provisioning [3]. To do so, for each received task, either a new VM is created or it is placed on the existing VM of the same user [2]. Once the task is completed, all the acquired resources are released which become parts of the free resource pool. Resource assignment is performed on the basis of Service Level Agreement (SLA) that is agreed between the service provider

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. D. Gomes.

* Corresponding author.
E-mail address: babarnazir@gmail.com (B. Nazir).

http://dx.doi.org/10.1016/j.compeleceng.2015.07.021
0045-7906/

and the customer. SLA contains details of the service level that is required by a tenant. Moreover, it contains information about the payment process and SLA violation penalty [5].

As the cloud computing paradigm is market-oriented, the traditional system-centric RM architectures are unsuitable for such systems [5]. System-centric architectures do not provide any incentives to the service providers and deal all requests with equal importance. Consequently, explicit market-oriented techniques are devised for cloud environments that are capable to cater on-demand resource provisioning and resource sharing among users. Such RM techniques provide economic incentives to both the service providers and customers. Service providers share their resources among multiple tenants on pay as you use basis.

In this work, we intend to provide an insight, and highlight issues in resource management techniques that need to be improved. To the best of our knowledge, currently, there exist only one survey paper [6] on this topic that lacks taxonomy of RM techniques, detailed working of each technique, critical discussions, overview of evaluation parameters, design goals, and research challenges that should be considered while designing RM techniques. Although, a few related surveys [7,8] highlight research issues and design goals with reference to a single aspect of RM techniques, there is a need to conduct a comprehensive study. This study presents taxonomy of RM techniques, their detailed working, critical discussions, evaluation parameters, evaluation platforms, design goals, and research challenges. For this study, we studied over 250 research papers, and selected over 100 papers. The selected papers provide a representative sample of the most significant work. Mixed-method systematic review strategy was used for the selection of paper, and at-least five papers were selected for each category. Moreover, only those papers were considered, which are published in high reputed journals and conferences. The keywords and strings used for searching of the papers are, ''resource management + cloud computing'', ''resource allocation + cloud computing'', ''energy-aware + resource management + clouds'', ''SLA-aware + resource management + clouds'', ''network load + resource management + clouds'', ''mobile cloud + resource management'', and ''hybrid cloud + resource management''.

The major contributions of this study are as follows:

1. It presents a taxonomy of RM techniques that is based on the major RM metrics. Research metrics that are used to provide RM solutions include energy efficiency, SLA-awareness, network load minimization, load balancing, profit maximization, hybrid cloud computing, and mobile cloud computing.
2. It presents detailed working of the existing techniques and highlights the research challenges that are addressed in each technique. Moreover, it highlights the issues that are left unconsidered while devising the techniques.
3. It provides discussion on various performance evaluation parameters and provides the definitions and equations of commonly used parameters. In addition, it provides a review of the parameters that are used by the researchers to evaluate their RM techniques.
4. It highlights various evaluation platforms that are used for the evaluation of RM techniques.
5. It provides detailed discussions/recommendations on various design goals that must be considered while designing a new RM technique. It also highlights the importance of each design goal, and discusses its importance for the better performance of RM technique.
6. Lastly, it highlights open research issues, which are of high importance and requires attention.

The rest of the paper is organized as follows. Section 2 presents the brief overview of existing surveys. Section 3 illustrates taxonomy of RM techniques. Section 4 presents the major performance evaluation parameters. Section 5 provides an overview of the evaluation parameters and platforms. Section 6 provides discussions on design goals. Section 7 highlights various research challenges, and Section 8 presents conclusions.

2. Existing surveys

In the past few years, RM techniques for cloud environment received great attention from the researchers. Various RM techniques are devised that consider research challenges that can hinder smooth provisioning and maintenance of cloud resources. However, very limited survey/review articles are available to facilitate new researchers in understanding the key concepts, and working of the existing techniques. In this section, we discuss various studies that somehow intend to provide insight on RM in cloud environments.

Jennings and Stadler [6] provide a survey on RM in cloud computing that discusses the scope of RM, various types of resources, enabling technologies, functions of RM, and various RM techniques. The authors also discuss workload management and some research challenges. In [7], Lin discusses working of various resource scheduling algorithms, and classifies the algorithms based on multiple factors, such as time, cost, and energy. Their main idea is to assist users in selecting suitable scheduling algorithm based on the type of service they want to use. Rygielski and Kounev [9] discuss various network virtualization techniques and their impact on QoS-aware RM. This study presents the impact of virtualization techniques on the performance of data centre networks. Various research issues are conferred that are faced during the performance modeling and designing of the RM techniques. In [10], the authors present architecture, principle, and an algorithm for energy-efficient cloud environments. In addition, it provides a survey and research challenges of energy efficient resource allocation techniques. Marojevic et al. [8] discuss numerous RM techniques that manage resources in Software Defined Radio (SDR)