# An improved Resource Scheduling Algorithm for Segregating User Requests in Cloud Computing

Sreenivasa Prasad Kakumani
x18146686
MSc in Cloud Computing

29th July 2019

**Abstract**

Resource Management has become key for cloud providers for maintaining trust, enriching the user experience by meeting the customer demands within the specified Service Level Agreement (SLA). To overcome issues like non-uniform distribution of loads and utilizing the resources effectively an efficient resource provisioning algorithm has to be implemented. In this paper, an improved resource scheduling algorithm is developed to segregate the user requests to the virtual machine in such a way that all the long-running tasks will be assigned to one virtual machine and the rest of the tasks to a different machine. The segregation helps both user and provider for identifying potential threats like Denial of Service (DoS) attacks on high volume tasks. These vulnerable tasks can be terminated without impacting the other processes that are in the running state. The proposed model is able to segregate user requests to different virtual machines considering the execution time of the tasks and a similar computing experiment will be carried on a cloud environment.

## Contents

# 1 Introduction

Managing cloud resources is a process of covering distinct steps starting from request submission until request execution phase in the cloud. It has gained its importance because of the increased demand from the end-user. This process aims at meeting customer expectations through proper distribution of cloud resources. It has become important to maintain these resources effectively because of the heterogeneity nature in physical machines, humongous fluctuations in workload distributions due to unpredictable loads which lead to interdependent issues in the cloud environment. The different problems in resource management process can be solved by three cloud models - Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). Moreover efficient resource management strategies help in achieving the distinctive advantage of reduced power consumption which is another burning factor in cloud computing. In simple, it is a process of effectively and efficiently managing cloud resources by considering user requirements, agreed SLA etc. for improving Quality of Service (QoS) [1].

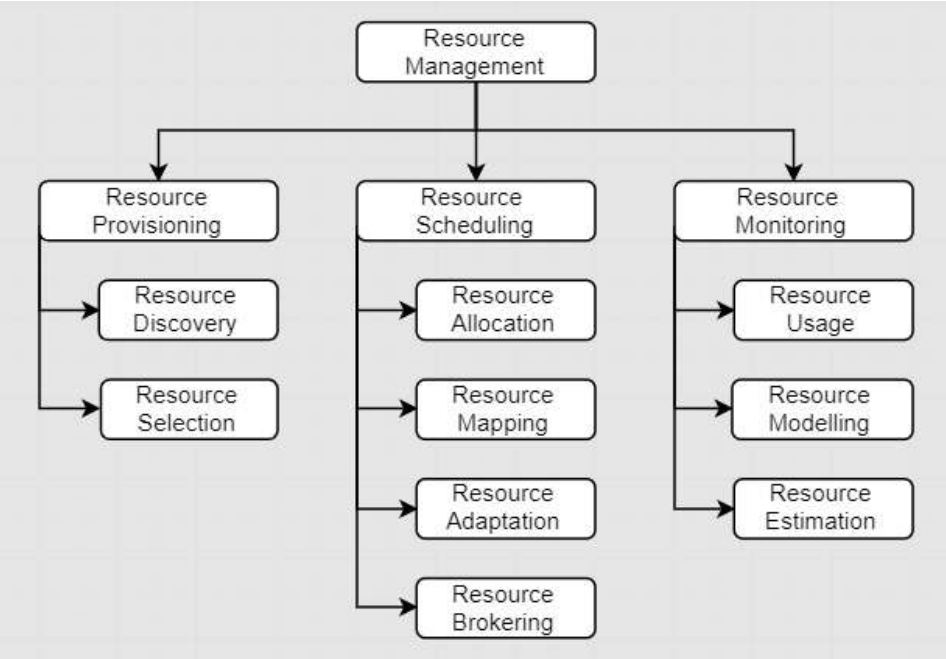


Figure 1: Resource Management Process classification based on [1]

There are three terms that we often come across during resource management process and these are further classified into several terms as shown in Figure 1. Each of this individual module is equally important for efficient resource management. Resource provisioning is a procedure of discovering the resources and distributing services to end-users. On the other hand, resource monitoring deals with measuring the physical resource usage for predicting resource demands in the future. The key part of the entire process is Resource scheduling because of various reasons like unpredictable loads, price models and last being the distribution of resources geographically. There are many challenges which are not solved by existing algorithms in resource scheduling module because of dispersion, uncertainty nature of tasks. The objective is to discover the best resource for the input tasks so that the algorithm used can enhance the QoS parameters. We often met with two major problems while scheduling the resources which are unequal load distribution and another being not able to utilize the resources efficiently. To overcome the above stated two problems we need to develop a capable and efficient algorithm which can serve this purpose [2].

There are plenty of such algorithms like First Come First Serve (FCFS), non-preemptive and preemptive Shortest Job First (SJF), Longest Job First (LJF) etc. which have their own limitations and advantages. Still, there is a wide research gap open in this module because of the criticality and our aim is to implement an improved algorithm with the help of existing solutions to improve QoS parameters [3]. We have organized rest our paper into several sections where Section 2 is focused on explaining the research question and its objective, Section 3 illustrates related work done in our chosen topic, an overview of our research method and specification is discussed in Section 4, Section 5 describes a detailed approach of our model, implementation of our model is depicted with the help of flowchart in Section 6, theoretical evaluations and an idea on different parameters we are considering are shown in Section 7, a detailed research timeline for our proposal is shown in section 8 using a Gantt chart, summary of our proposal is provided in Section 9.

## 2 Research Question and Rational

1. What is the impact/outcome of separating the long-running and short running user tasks to different Virtual Machines (VMs)?

2. How to develop a resource scheduling algorithm that can segregate the tasks keeping security features in mind?

The primary objective of this study is to separate customer requests to different VMs based on the execution time required by allotting distinct queues for both short running and long-running jobs to improve system efficiency, reduce waiting times, average response time and identifying potential threats such as DoS assaults primarily on high-volume requests. We will be employing a queuing technique with the help of existing models like FCFS, SJF, LJF etc. that can be able to separate high volume and low volume requests to achieve our goal. User requests are then routed to different VMs based on the pre-defined parameters like execution time, arrival time and the volume of the tasks received.

## 3 Literature Review

A brief study on different scheduling techniques was performed by M.Kumar et al., 2019 [2] and Arunarani et al., 2018 [4]. These two surveys clearly explain the various scheduling techniques for efficient Resource scheduling in the Cloud computing paradigm. The review also includes different solutions proposed to further enhance the performance of existing Algorithms. In cloud computing, Resource Allocation (RA) is the process of distributing available resources to the requested cloud applications through the internet. Resource Scheduling often is done in two ways; one is based on demand scheduling in which cloud provider assign the services required randomly that is there exists a huge chance of allocating tasks to a single machine and the other being the reserved resources i.e. there is a problem of being idle without performing any task for a long time. In order to effectively and efficiently manage these problems, a solution was proposed by Singh and Chana, in 2016 [5].

The requirement of Resource Provisioning with Scheduling (RPS) is provisioning VMs to end-users while not violating the predefined and agreed SLA. Provisioning these VMs to distribute the workload efficiently is a crucial part of scheduling. There are several other parameters that are to be considered like cost, energy consumption, makespan time, response times, etc. [3] [2]. All the scheduling Algorithms in cloud computing are categorized into two types; static scheduling like FCFS, SJF, round robin, LJF etc. and dynamic scheduling algorithms like Particle Swarm Optimization (PSO), Heterogeneous Earliest Finish Time (HEFT) etc. Load