



2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)

Efficient Scheduling Algorithm for Cloud  
Ellendula Madhukara, Thirumalaisamy Raguathanb

*a*Sreenidhi Institute of Science and Technology, Hyderabad, India  
*b*ACE Engineering College, Hyderabad, India

---

Abstract

In the emerging e-commerce scenario, information captured by enterprises through their web-based applications has been exploding. Extensive use of applications based on multimedia and social media effects in exponential growth of data. Cloud computing system (or simply cloud) provides the required infrastructure for storing the large data generated by these applications. Efficient task scheduling in the cloud will reduce the task completion time and hence improve the performance of the cloud. In this paper, we have proposed an efficient cloud scheduling algorithm based on the results of the analysis performed on the log which consists of previous scheduling information for improving the performance of the cloud.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of scientific committee of 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)

*Keywords:* Distributed System; Speculation; Performance;

---

1. Introduction

In the emerging e-commerce scenario, information captured by enterprises through their web-based applications has been exploding. Extensive use of applications based on multimedia and social media effects in exponential growth of data. Cloud computing system (or simply cloud) provides the required infrastructure for storing the large data generated by these applications. According to [3], a cloud is defined as "A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort of service provider interaction". Efficient task scheduling in the cloud will reduce the task completion time so that efficient services can be provided to the users. Ant colony- and simulated annealing-based algorithms, genetic algorithms and hybrid algorithms are the ones proposed in the literature for improving the performance of the cloud.

In this paper, we have assumed that a cloud consists of a controller node and one or more compute nodes. The controller node does the control of compute nodes present in the cloud including task scheduling. The compute node store the data and carry out the execution of tasks. Virtual machines (VMs) are created in the compute nodes to execute the user tasks. We have assumed that a fixed number of VMs are available (or created) in the compute nodes and different execution environments are maintained in these VMs. Note that, in a cloud environment VMs carry out the execution of tasks. In this paper, we are proposing that a log

---

*Email addresses:* [emadhukar@gmail.com](mailto:emadhukar@gmail.com) (Ellendula Madhukar), [ragu\\_savi@yahoo.com](mailto:ragu_savi@yahoo.com) (Thirumalaisamy Raguathan)

is maintained in each compute node and all scheduling information relevant to individual VMs are captured in the log.

In this paper, we have proposed an efficient cloud scheduling algorithm which can improve the performance of the cloud. The proposed algorithm mines the log maintained in the compute nodes to come out with the some scheduling patterns for VMs present in the compute nodes of the cloud. The incoming task to the cloud (controller node) will be scheduled to a particular compute node if the task requirements matches one of scheduling patterns of VMs present in those nodes.

Characteristics	Symbols used
User	U
Executable	E
Number of Nodes	N
Type (Batch, Interactive Serial, Parallel)	T
Queue	Q
Class	C
Load leveler Script	S
Arguments	P
Network Adaptor	Na
Start time	St
Run time	Rt

Fig. 1: Job Characteristics Proposed by Smith

## 2. Related Work

In this section, we discuss how job characterization is done in the cloud. Gibson [2] proposes a template model and that has been modified by Smith [4]. The job (or task) characteristics used by Smith is shown in the Figure 1.

Ariel Goyeneche [1] proves that using dynamic templates without normalized data will lead to inaccurate predictions. The authors have used weighted templates. In this method, characteristics of jobs will be considered based on the different weights allocated. Important characteristics will have higher priority and they are mandatory for job performance estimations. The prediction accuracy is based on the weights allocated to characteristics. Autopsy of all templates can be used to fix the weights of templates. In this method, characteristics are divided into two sets. First set comprises of User Id, Job Id, VMnum , Job size, VM ram, VMos(VM Operating system). Second set consists of all the remaining characteristics. The characteristics specified in the first set will have more weights than characteristics specified in the secondset

Download English Version:

<https://daneshyari.com/en/article/489878>

Download Persian Version:

<https://daneshyari.com/article/489878>

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