A Heuristic Meta Scheduler for Optimal Resource Utilization and Improved QoS in Cloud Computing Environment

R. Jeyarani, Coimbatore Institute of Technology, India
N. Nagaveni, Coimbatore Institute of Technology, India

ABSTRACT

This paper presents a novel Meta scheduler algorithm using Particle Swarm Optimization (PSO) for cloud computing environment that focuses on fulfilling deadline requirements of the resource consumers as well as energy conservation requirement of the resource provider contributing towards green IT. PSO is a population-based heuristic method which can be used to solve NP-hard problems. The nature of jobs is considered to be independent, non-pre-emptive, parallel and time critical. In order to execute jobs in a cloud, primarily Virtual Machine (VM) instances are launched in appropriate physical servers available in a data-center. The number of VM instances to be created across different servers to complete the time critical jobs successfully, is identified using PSO by exploiting the idle resources in powered-on servers. The scheduler postpones the power-up/activation of new servers/hosts for launching enqueued VM requests, as long as it is possible to meet the deadline requirements of the user. The Meta Scheduler also incorporates Backfilling Strategy which improves makespan. The results conclude that the proposed novel Meta scheduler gives optimization in terms of number of jobs meeting their deadlines (QoS) and utilization of computing resources, helping both cloud service consumer as well as cloud service provider.

Keyword: Cloud Computing, Backfilling, Meta Scheduling, Particle Swarm Optimization, QoS Based Scheduling

1. INTRODUCTION

Cloud Computing is emerging as a promising distributed computing paradigm. Cloud computing delivers a wide range of services like Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). These services are made available on a subscription basis using pay-as-you-use model to customers, regardless of their location. Thus CPU cycles and bytes have become the new commodity now. Cloud computing is a cost-effective model for provisioning services and it makes IT management easier and more responsive to the changing needs of the business.
In a cloud computing system, the workload significantly moves away from the end user as the jobs are fired into the cloud. The local machine of the user has the interface installed to communicate with the cloud, make requests for resources and dispatch jobs. Clouds are clearly next-generation data centers with nodes "virtualized" through hypervisor technologies, dynamically provisioned on demand as a personalized resource collection to meet a specific Service Level Agreement (SLA), which is established through negotiation and accessible as a composable service via Web Service technologies such as SOAP and REST (Schute & Groenwold, 2005). The key enablers of cloud computing include virtualization and multi-tenancy (Tasgetiren, Sevkli, Liang, & Gencyilmaz, 2004). It becomes complex to quantify the performance of scheduling and allocation policy on cloud infrastructures for different applications under varying work load and system size. Some of the classical cloud-based applications include Social Networking, Web Hosting, Content Delivery, and Real-Time Instrumented data processing. These application types differ in composition, configuration and deployment requirements. This paper focuses on a hybrid meta-scheduler that is built over Particle Swarm Optimization (PSO) and Backfilling Strategy.

The developed scheduler fulfills the deadline requirements of the resource consumers as well as energy conservation requirement of the resource provider. The use of real test-bed limits the experiments and makes reproduction of results a difficult task. The alternative solution is the use of simulation tools, which enable the possibility of evaluating the hypothesis prior to software development and deployment. The simulation based approach significantly benefits the researchers to test their proposed algorithms and protocols in a scalable and controlled environment, and to find solution to the performance bottlenecks before deploying in the real cloud. CloudSim is a widely used cloud simulation tool which is opted for implementation and evaluation of the work done.

2. PROBLEM DEFINITION

In cloud IaaS model, service consumer specify hardware and software configuration of Virtual Machines(VM) to be created and maps a set of jobs to each VM. On a broader classification, the nature of the submitted jobs are parallelizable, non-pre-emptive, independent and time-
Word-Of-Mouth, Trust, and Perceived Risk in Online Shopping: An Extension of the Technology Acceptance Model
[www.igi-global.com/article/word-of-mouth-trust-and-perceived-risk-in-online-shopping/161769?camid=4v1a](www.igi-global.com/article/word-of-mouth-trust-and-perceived-risk-in-online-shopping/161769?camid=4v1a)

Computational Intelligence for Information Technology Project Management
Robert J. Hammell, Julie Hoksbergen, James Wood and Mark Christensen (2010). *Intelligent Systems in Operations: Methods, Models and Applications in the Supply Chain* (pp. 80-104).
[www.igi-global.com/chapter/computational-intelligence-information-technology-project/42656?camid=4v1a](www.igi-global.com/chapter/computational-intelligence-information-technology-project/42656?camid=4v1a)