Accepted Manuscript

Cost optimization for deadline-aware scheduling of big-data processing jobs on clouds

Wei Zheng, Yingsheng Qin, Bugingo Emmanuel, Dongzhan Zhang, Jinjun Chen

PII: S0167-739X(17)31729-6
DOI: https://doi.org/10.1016/j.future.2017.12.004
Reference: FUTURE 3842

To appear in: *Future Generation Computer Systems*

Received date : 31 July 2017
Revised date : 13 November 2017
Accepted date : 3 December 2017

Please cite this article as: W. Zheng, Y. Qin, B. Emmanuel, D. Zhang, J. Chen, Cost optimization for deadline-aware scheduling of big-data processing jobs on clouds, *Future Generation Computer Systems* (2017), https://doi.org/10.1016/j.future.2017.12.004

Cost optimization for deadline-aware scheduling of big-data processing jobs on clouds

Wei Zheng[a], Yingsheng Qin[a], Bugingo Emmanuel[a], Dongzhan Zhang[a,*], Jinjun Chen[b]

[a]*School of Information Science and Engineering, Xiamen University, China*
[b]*Swinburne Data Science Research Institute, Swinburne University of Technology, Australia*

Abstract

Cloud computing has been widely regarded as a capable solution for big data processing. Nowadays cloud service providers usually offer users virtual machines with various combinations of configurations and prices. As this new service scheme emerges, the problem of choosing the cost-minimized combination under a deadline constraint is becoming more complex for users. The complexity of determining the cost-minimized combination may be resulted from different causes: the characteristics of user applications, and providers' setting on the configurations and pricing of virtual machine. In this paper, we proposed a variety of algorithms to help the users to schedule their big data processing workflow applications on clouds so that the cost can be minimized and the deadline constraints can be satisfied. The proposed algorithms were evaluated by extensive simulation experiments with diverse experimental settings.

Keywords: Big-data, Scheduling, Cost-efficient, Cloud Computing

1. Introduction

Big data processing has been receiving more and more attention over recent years, since the increasing needs raised from various disciplines, such as science,

*Corresponding author
Email address: zdz@xmu.edu.cn (Dongzhan Zhang)
A preliminary version has been accepted in The Fifth International Conference on Advanced Cloud and Big Data, August 2017, Shanghai, China [1]