Accepted Manuscript

A benchmark approach and its toolkit for online scheduling of multiple deadline-constrained workflows in big-data processing systems

Dongzhan Zhang, Wenjing Yan, Emmanuel Bugingo, Wei Zheng, Jinjun Chen

PII: S0167-739X(17)32353-1
DOI: https://doi.org/10.1016/j.future.2018.03.046
Reference: FUTURE 4059

To appear in: *Future Generation Computer Systems*

Received date : 15 October 2017
Revised date : 19 February 2018
Accepted date : 24 March 2018

Please cite this article as: D. Zhang, W. Yan, E. Bugingo, W. Zheng, J. Chen, A benchmark approach and its toolkit for online scheduling of multiple deadline-constrained workflows in big-data processing systems, *Future Generation Computer Systems* (2018), https://doi.org/10.1016/j.future.2018.03.046

A Benchmark Approach and its Toolkit for Online Scheduling of Multiple Deadline-constrained Workflows in Big-data Processing Systems

Dongzhan Zhang[a,*], Wenjing Yan[a], Emmanuel Bugingo[a], Wei Zheng[a], Jinjun Chen[b]

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

Abstract

As distributed systems such as clouds get increasingly popular in the use for big-data processing, there is a need of shifting research attention from minimizing the workflow completion time to satisfying the deadline constraints specified by the users and boosting the benefit of the resource providers. This paper focuses on the online deadline-constrained workflow scheduling problem of how to schedule a set of sequentially submitted workflows with deadline constraints to maximize the resource utilization as well as the success rate of meeting the deadlines. A discrete-event based simulator with a novel benchmark approach is proposed to ease the analysis of the problem. Extensive evaluation has been done to exhibit the effectiveness and significance of the proposed benchmark approach and the developed simulator.

Keywords: Big-data, Workflow, Online Scheduling, Simulator, Benchmark

1. Introduction

Over the recent years, there has been a continuous growth of the already-massive computation and/or storage requirements arising from various activi-

*Corresponding author
Email address: zdz@xmu.edu.cn (Dongzhan Zhang)
A preliminary version has been published in The Third International Conference on Advanced Cloud and Big Data, October 2015, Yangzhou, China [1]