MapReduce scheduling algorithms: a review

Ibrahim Abaker Targio Hashem[a,b], Nor Badrul Anuar[b], Mohsen Marjani[a], Ejaz Ahmed[c], Haruna Chiroma[d], Ahmad Firdaus[e], Muhamad Taufik Abdullah[f], Faiz Alotaibi[f], Waleed Kamaleldin Mahmoud Ali[b], Ibrar Yaqoob[b] & Abdullah Gani[a]

[a] School of Computing and Information Technology, Taylor's University, Subang Jaya, 47500, Malaysia
[b] Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia
[c] Centre for Mobile Cloud Computing Research, University of Malaya, Kuala Lumpur, Malaysia
[d] Department of Computer Science, Federal College of Education (Technical), Gombe, Nigeria
[e] Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Gambang, Kuantan, Pahang 26300, Malaysia
[f] Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Selangor 43400, Malaysia

ABSTRACT

Recent trends in big data have shown that the amount of data continues to increase at an exponential rate. This trend has inspired many researchers over the past few years to explore new research direction of studies related to multiple areas of big data. The widespread popularity of big data processing platforms using MapReduce framework is the growing demand to further optimize their performance for various purposes. In particular, enhancing resources and jobs scheduling are becoming critical since they fundamentally determine whether the applications can achieve the performance goals in different use cases. Scheduling plays an important role in big data, mainly in reducing the execution time and cost of processing. This paper aims to survey the research undertaken in the field of scheduling in big data platforms. Moreover, this paper analyzed scheduling in MapReduce on two aspects: taxonomy and performance evaluation. The research progress in MapReduce scheduling algorithms is also discussed. The limitations of existing MapReduce scheduling algorithms and exploit future research opportunities are pointed out in the paper for easy identification by researchers. Our study can serve as the benchmark to expert researchers for proposing a novel MapReduce scheduling algorithm. However, for novice researchers, the study can be used as a starting point.