
On Understanding the Energy Impact of Speculative Execution in Hadoop

Tien-Dat Phan^{*†1}

¹KerData (INRIA - IRISA) – Universite de Rennes 1, INRIA, CNRS : UMR6074 – Campus de Beaulieu
35042 Rennes cedex, France

Abstract

Hadoop emerged as an important system for large-scale data analysis. Speculative execution is a key feature in Hadoop that is extensively leveraged in clouds: it is used to mask slow tasks (i.e., stragglers) — resulted from resource contention and heterogeneity in clouds — by launching speculative task copies on other machines. However, speculative execution is not cost-free and may result in performance degradation and extra resource and energy consumption. While prior literature has been dedicated to improving stragglers detection to cope with the inevitable heterogeneity in clouds, little work is focusing on understanding the implications of speculative execution on the performance and energy consumption in Hadoop cluster. In this paper, we have designed a set of experiments to evaluate the impact of speculative execution on the performance and energy consumption of Hadoop in homo- and heterogeneous environments. Our studies reveal that speculative execution may sometimes reduce, sometimes increase the energy consumption of Hadoop clusters. This strongly depends on the reduction in the execution time of MapReduce applications and on the extra power consumption introduced by speculative execution. Moreover, we show that the extra power consumption varies in-between applications and is contributed to by three main factors: the duration of speculative tasks, the idle time, and the allocation of speculative tasks. To the best of our knowledge, our work provides the first deep look into the energy efficiency of speculative execution in Hadoop.

*Speaker

†Corresponding author: