

PERFORMANCE ANALYSIS AND SCHEDULING IN BIG DATA SYSTEMS

Li Zhang, IBM Thomas J. Watson Research Center, New York, USA, zhangli@us.ibm.com

Jian Tan, IBM Thomas J. Watson Research Center, New York, USA, tanji@us.ibm.com

Xiaoqiao Meng, IBM Thomas J. Watson Research Center, New York, USA, xmeng@us.ibm.com

More and more analytic applications are built in Big Data systems providing important insights from analyzing large amount of data. Understanding the performance issues in these large distributed systems is important for providing efficient resource utilization and satisfactory performance guarantees. Due to the large number of jobs sharing limited resources, an efficient job scheduler is critical. We present analytical and experimental studies of existing job schedulers. We also introduce our own job scheduler with superior performance.