A Smart Speculative Execution Strategy based on Node Classification for Heterogeneous Hadoop Systems

Qi Liu*, Weidong Cai*, Jian Shen*, Zhangjie Fu**, Nigel Linge***

* College of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China

** Jiangsu Engineering Centre of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China

*** The University of Salford, Salford, Greater Manchester, UK

qi.liu@nuist.edu.cn, caiweidongsuzhou@163.com, s_shenjian@126.com, wwwfzj@126.com, n.linge@salford.ac.uk

***Abstract*—MapReduce (MR) has been widely used to process distributed large data sets. Meanwhile, speculative execution is known as an approach for dealing with same problems by backing up those tasks running on a low performance machine to a higher one. In this paper, we have modified some pitfalls and taken heterogeneous environment into consideration. We also have implemented it in Hadoop-2.6 based on node classification, this strategy is called Speculation-NC and optimized Hadoop is called Hadoop-NC. Experiment results show that our method can correctly backup a task, improve the performance of MRV2 and decrease the execution time and resource consumption compared with traditional strategy.**

***Keyword*—MapReduce; speculative execution; time prediction; Node classification**

Qi Liu (M'11) received his BSc degree in Computer Science and Technology from Zhuzhou Institute of Technology, China in 2003, and his MSc and PhD in Data Telecommunications and Networks from the University of Salford, UK in 2006 and 2010. His research interests include context awareness, data communication in MANET and WSN, and smart grid. His recent research work focuses on intelligent agriculture and meteorological observation systems based on WSN.

Weidong Cai received his bachelor's degree in Software Engineering from Nanjing University of Information Science and Technology in 2014, and he is pursuing a master's degree in software engineering at the Nanjing University of Information Science and Technology. His research interests include Cloud Computing, Distributed Computing and Data Mining.

Jian Shen received his bachelor's degree in Electronic Science and Technology Specialty from Nanjing University of Information, Science and Technology in 2007, and he received his masters and PhD in Information and communication from CHOSUN University, South Korean in 2009 and 2012. His research interests includes Computer network security, information security, mobile computing and network, wireless ad hoc network.

Zhangjie Fu received his BS in education technology from Xinyang Normal University, China, in 2006; received his MS in education technology from the College of Physics and Microelectronics Science, Hunan University, China, in 2008; obtained his PhD in computer science from the College of Computer, Hunan University, China, in 2012. Currently, he works as an assistant professor in College of Computer and Software, Nanjing University of Information Science and Technology, China. His research interests include cloud computing, digital forensics, network and information security.

Nigel Linge received his BSc degree in Electronics from the University of Salford, UK in 1983, and his PhD in Computer Networks from the University of Salford, UK, in 1987. He was promoted to Professor of Telecommunications at the University of Salford, UK in 1997. His research interests include location based and context aware information systems, protocols, mobile systems and applications of networking technology in areas such as energy and building monitoring.