

An Adaptive Tuning Strategy on Spark Based on In-memory Computation Characteristics

Yao Zhao*, Fei Hu**, Haopeng Chen*

* *School of Electronic Information and Electrical Engineering,
Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, China*
GuiLong@sjtu.edu.cn, hu.fei@sjtu.edu.cn, chen-hp@sjtu.edu.cn

Abstract—We present an adaptive tuning method to improve Spark performance, especially for its in-memory computation. This manner serves one purpose: making a better use of memory reasonably through adaptively adopting suitable category based on Spark application runtime statistics on different working sets. This solution works in two steps. Firstly, it collects run-time statistics dynamically and stores them in round-robin structures to save memory. Secondly, it can change system storage category based on these statistics. Additionally we focus on serialization strategy optimization. For this purpose we test Spark integrated serialization algorithms: Java and Kryo serialization algorithms, and make a comparison of their performance. In order to gain flexibility we change Spark serialization mechanism by setting the default serialization unit from one RDD to one block. In this way, for the case which RDD has huge amount of blocks our solution can use different serialization algorithms to serialize different blocks in one RDD. We show that our solution is expressive enough to obtain 2x speedup than original Spark when there is inadequate memory resource.

Keyword—Spark, RDD, adaptive tuning, in-memory computation, serialization.



Yao Zhao. He got his Bachelor degree in computer science and technology, Xi'an Jiao Tong University in 2012 Xi'an, Shanxi Province, China. And now he is working for Master degree in software engineering, Shanghai Jiao Tong University, Shanghai, China. He is interested in machine learning algorithm, natural language identification and big data processing. Mr. Zhao is working for his Master degree in Shanghai Jiao Tong University.



Fei Hu. He received his Bachelor degree from Department of computer software, Northwest University, Xi'an, Shanxi Province, China in 1990 and received his Master degree of computer science and engineering and Ph.D of Precision Guidance and Control both from Northwest Polytechnical University, Xi'an, Shanxi Province, China in 1993 and 1998. He has worked in Department of Computer Science and Engineering, Northwestern Polytechnical University lecturer, from 1993 to 2006. From 2006/ 9 to now he has worked in School of Software, Shanghai Jiao Tong University. His Publications are as follows: Zhiyang Zhang, Fei Hu and Jian Li, "Autonomous Flight Control System Designed for Small-Scale Helicopter Based on Approximate Dynamic Inversion," The 3rd IEEE International Conference on Advanced Computer Control (ICACC 2011), 18th to 20th January 2011, Harbin, China.



Haopeng Chen. He received his Ph.D degree from Department of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an, Shanxi Province, China in 2001. He has worked in School of Software, Shanghai Jiao Tong University since 2004 after he finished his two-year postdoctoral research job in Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. He got the position of Associate Professor in 2008. In 2010, he studied and researched in Georgia Institute of Technology as a visiting scholar. His research group focuses on Distributed Computing and Software Engineering. They have kept researching on Web Services, Web 2.0, Java EE, .NET, and SOA for several years. Recently, they are also interested in cloud computing and researching on the relevant areas, such as cloud federation, resource management, dynamic scaling up and down, and so on.