Accepted Manuscript

Design and implementation of reconfigurable acceleration for in-memory distributed big data computing

Junjie Hou, Yongxin Zhu, Sen Du, Shijin Song

PII: S0167-739X(18)31210-X
DOI: https://doi.org/10.1016/j.future.2018.09.049
Reference: FUTURE 4484

To appear in: *Future Generation Computer Systems*

Received date : 18 May 2018
Revised date : 17 August 2018
Accepted date : 19 September 2018

Please cite this article as:, Design and implementation of reconfigurable acceleration for in-memory distributed big data computing, *Future Generation Computer Systems* (2018), https://doi.org/10.1016/j.future.2018.09.049

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Design and Implementation of Reconfigurable Acceleration for In-memory Distributed Big Data Computing

Junjie Hou[a], Yongxin Zhu[b,c,a,*], Sen Du[a], Shijin Song[a],

[a]*School of Microelectronics, Shanghai Jiao Tong University, Shanghai 200240, P. R. China*
[b]*Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, P. R. China*
[c]*University of Chinese Academy of Sciences, Beijing 100049, P. R. China*

Abstract

Apache Spark is an efficient distributed computing framework for big data processing. It supports in-memory computation of RDDs (Resilient Distributed Dataset) and provides a provision of reusability, fault tolerance, and real-time stream processing. However, the tasks in Spark framework are only performed on CPU. The low degree of parallelism and power inefficiency of CPU may restrict the performance and scalability of the cluster. In order to improve the performance and power dissipation of the data center, heterogeneous accelerators such as FPGA, GPU, MIC (Many Integrated Core) exhibit more efficient performance than general-purpose processors in big data processing. In this work, we propose a framework to integrate FPGA accelerators into a Spark cluster, which achieves performance improvement and power dissipation reduction for distributed applications. We propose a method for connecting Spark with OpenCL application which is a standard for cross-platform, parallel programming of diverse processors and widely used in heterogeneous computing, and use FPGA to accelerate the Spark tasks developed with Python. We illustrate the performance and the energy efficiency of FPGA based Spark framework with a case study of K-means algorithm acceleration. The results show that FPGA based Spark implemen-

*Corresponding author
Email address: zhuyongxin@sari.ac.cn (Yongxin Zhu)

Download English Version:

https://daneshyari.com/en/article/11012473

Download Persian Version:

https://daneshyari.com/article/11012473

Daneshyari.com