Accepted Manuscript

In-memory big data analytics under space constraints using dynamic programming

Keke Gai, Meikang Qiu, Meiqin Liu, Zenggang Xiong

PII: S0167-739X(17)31601-1
DOI: https://doi.org/10.1016/j.future.2017.12.033
Reference: FUTURE 3871

To appear in: *Future Generation Computer Systems*

Received date : 20 July 2017
Revised date : 12 November 2017
Accepted date : 22 December 2017

Please cite this article as: K. Gai, M. Qiu, M. Liu, Z. Xiong, In-memory big data analytics under space constraints using dynamic programming, *Future Generation Computer Systems* (2018), https://doi.org/10.1016/j.future.2017.12.033

In-Memory Big Data Analytics Under Space Constraints Using Dynamic Programming

Keke Gai[a], Meikang Qiu[b,c,*], Meiqin Liu[d], Zenggang Xiong[c]

[a]*School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China, 100081.*
[b]*College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, China.*
[c]*School of Computer Science and Information Technology, Hubei Engineering University, Hubei, 43200, China.*
[d]*College of Electrical Engineering, Zhejiang University, ZJ 310027, China*

Abstract

The emergence of persistent memories has powered the data processing with the in-memory environment and in-memory data analytics have become an advance of high-performance data processing. Recent explorations of using in-memory technologies address the improvement of the memory performance from re-designing file systems. Most current approaches mitigate data exchanges between buffers and disks by migrating workload to memories. However, this type of solutions will be encountering the restriction of the memory size with the rapid growth of the application volume. This paper focuses on the issue caused by the large amount of data processing within in-memory systems and proposes a novel approach that is designed to dynamically determine whether the data processing should be accomplished in the memory. The proposed approach is called *Smart In-Memory Data Analytics Manager* (SIM-DAM) model, which utilizes a dynamic working manner of the file system, as well as fully uses hardware mappings. The experimental results obtained from our laboratory evaluations represent that the throughputs of SIM-DAM can achieve a high-level performance with different input data sizes without the constraints of the memories' spaces.

Keywords: In-memory data analytics, dynamic programming, heterogeneous computing, big data processing, on-chip memory architecture

*Meikang Qiu (Corresponding Author): School of Computer Science and Information Technology, Hubei Engineering University, Hubei, 43200, China and Department of Computer Science, Pace University, New York City, NY 10038, USA. E-mail: mqiu@pace.edu

Email addresses: kekegai@yahoo.com (Keke Gai), mqiu@pace.edu (Meikang Qiu), liumeiqin@zju.edu.cn (Meiqin Liu), xzg@hbeu.edu.cn (Zenggang Xiong)

[1]This work is supported by China NSFC 61728303 and the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China (ICT170331); NSF of China Grants (61370092) and Youth innovation team project in Hubei Provincial Department of Education (No. T201410).