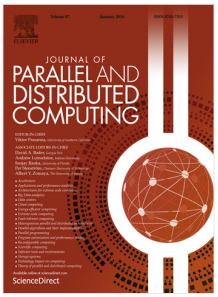
Accepted Manuscript



Making sense of performance in in-memory computing frameworks for scientific data analysis: A case study of the spark system

Xuechen Zhang, Ujjwal Khanal, Xinghui Zhao, Stephen Ficklin

PII: S0743-7315(17)30292-7
DOI: https://doi.org/10.1016/j.jpdc.2017.10.016
Reference: YJPDC 3766

To appear in: *J. Parallel Distrib. Comput.*

Received date : 28 July 2017
Revised date : 11 October 2017
Accepted date : 27 October 2017

Please cite this article as: X. Zhang, U. Khanal, X. Zhao, S. Ficklin, Making sense of performance in in-memory computing frameworks for scientific data analysis: A case study of the spark system, *J. Parallel Distrib. Comput.* (2017), https://doi.org/10.1016/j.jpdc.2017.10.016

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

*Highlights (for review)

Research highlights

- We develop a benchmark, ArrayBench, for benchmarking scientific data analytics that process gene expression matrices using Spark and SciDB.
- We study the correlations between the performance of Spark data analytics and various OS components, e.g., memory, storage, and file systems.
- Our findings shed light on the improvement of Spark and SciDB and the future development of data-intensive data analytics using the in-memory computing frameworks.

ACCEPTED MANUSCRIPT

Download English Version:

https://daneshyari.com/en/article/6874923

Download Persian Version:

https://daneshyari.com/article/6874923

Daneshyari.com