

Dear Editor-in-Chief,

As a corresponding author, I am sending you our manuscript entitled “Augmenting Amdahl’s Second Law: A Theoretical Model for Cost-Effective Balanced HPC Infrastructure for Data-Driven Science ” by Arghya Kusum Das, Jaeki Hong, Sayan Goswami, Richard Platania, Kisung Lee, Wooseok Chang, Seung-Jong Park et al. We would like to have the manuscript considered for publication in IEEE Access Special Section: Theoretical Foundations for Big Data Applications: Challenges and Opportunities.

A first part of this ongoing study is previously published in conference proceeding (IEEE BigData 2015) with experimental evaluation of a broad range of distributed-cyber-infrastructure. The core part of this paper is substantially different as it proposes a new theoretical model to help the system designers to propose novel distributed-cyber-infrastructure. We estimate percent of new material to be more than 50%. Please see list of modifications below. We attach a copy of the previous conference version in supplementary materials.

With my best regards,

Sincerely yours,

Seung-Jong Jay Park, Ph.D.
Associate Professor
School of Electrical Engineering & Computer Science
Center for Computation & Technology
Louisiana State University

Change list

Original Conference Paper

- Title: Evaluating Different Distributed-Cyber-Infrastructure for Data and Compute Intensive Scientific Application
- Authors: Arghya Kusum Das, Jaeki Hong, Wooseok Chang, and Seung-Jong Park

New Journal Submission

- Title: Augmenting Amdahl's Second Law: A Theoretical Model for Cost-Effective Balanced HPC Infrastructure for Data-Driven Science
- Authors: Arghya Kusum Das, Jaeki Hong, Sayan Goswami, Richard Platania, Kisung Lee, Wooseok Chang, and Seung-Jong Park

Other List of Modifications:

- 1) The core part of this paper proposes a new theoretical model for alternative cluster architecture for data- and compute-intensive applications. The previous work focused only on experimental evaluations.
- 2) In the last paper we observed a tradeoff between performance and cost/performance in scale-out and scale-up clusters. This paper provided enough theoretical insight to resolve the performance and cost conundrum.
- 3) The previous paper used our own genome assembly software as the only benchmark application. Whereas in this paper, along with the assembler we also include other widely used benchmark (such as, Terasort and WordCount) to substantiate the claim of the theoretical model.
- 4) As this paper proposes a theoretical model for new cluster architecture, the architecture of the experimental testbeds that are evaluated here have changed to fit this model from the previous paper.
- 5) The abstract, introduction, related work is modified accordingly to the core part of this paper.