Strategies for deploying stochastic HPC applications on large-scale systems

Ana Gainaru[*1], Guillaume Aupy[2], Hongyang Sun[3], and Padma Raghavan

[1]Vanderbilt University [Nashville] – United States
[2]Department of Electrical Engineering Computer Science [Nashville] – Vanderbilt University VU Station B 351824 2301 Vanderbilt Place Nashville, TN 37235-1824 USA, United States
[3]Laboratoire de l'Informatique du Parallélisme (LIP) – École Normale Supérieure (ENS) - Lyon, INRIA, CNRS : UMR5668, PRES Université de Lyon, Université Claude Bernard - Lyon I (UCBL) – 46 Allée d'Italie 69364 LYON CEDEX 07, France

Abstract

The last few years have seen an increase in new emerging fields that require the use of large-scale systems. These fields use high-performance computing applications with heterogeneous, dynamic and data-intensive requirements that put a high emphasis on productivity and are not tuned to run efficiently on today's large-scale systems. This results in a loss of efficiency and innovation when deploying them on current HPC systems as well as a decrease in effective system utilization for platform providers due to ad-hoc solutions and unsupervised resource management strategies implemented by the user. This talk presents a brief characterization of these workflows and proposes a couple of solutions to adapt the runtime system to the computational needs of stochastic applications. We focus on improving the overall system utilization as well as application response time, without disrupting either the current HPC platform model nor the application development process.

Keywords: stochastic applications, hpc, runtime system, resource manager

---

[*]Speaker