SAND2024-11740C

Modeling the Performance and Scaling of Scientific Workflows with Resource-efficient Workflow Mini-apps

Tianle Wang[1], Ozgur Ozan Kilic[1], Matteo Turilli[1,2], Mikhail Titov[1], Andre Merzky[3], Line Pouchard[1], Shantenu Jha[1,2],
Brookhaven National Lab[1], Rutgers University[2], *RADICAL-Computing Inc*[3]

Scientific workflows are becoming pivotal in driving scientific discoveries, necessitating the tailoring of next-generation supercomputers to run these workflows efficiently. However, evaluating the performance of emerging computing systems using production-scale workflows is costly and energy-inefficient, especially at extreme scales. Moreover, application-level mini-apps fail to replicate these workflows' sophistication, heterogeneity, and scale. We propose a new modeling technique as the workflow mini-app to address these challenges. The workflow mini-app is designed to faithfully represent the performance characteristics of real-world workflows without specific constraints of the workflows. Workflow mini-apps[1] are intended for deployment and execution on various systems and architectures, including High-Performance Computing (HPC), enabling performance evaluation and reproducibility analysis at a significantly reduced cost and improved energy efficiency.

In this work, we model the performance and scalability of scientific workflows by targeting various scientific workflows, from inverse problems to AI-steered simulations. We generate workflow mini-apps based on original workflows to model the performance of similar workflows (class of workflow) by enabling adjustable parameters. By running workflow mini-apps across different architectures, we provide insights into the suitability of various software/hardware systems for given scientific workflows. This facilitates informed decisions regarding HPC infrastructure investments, performance optimization, and understanding potential scaling behaviors. The benefits of our approach include but are not limited to balancing the accuracy of performance representation with cost and energy efficiency and overcoming the limitations of application-level mini-apps and production-scale workflows.

Our initial findings demonstrate the effectiveness of workflow mini-apps in estimating execution performance and predicting scalability across different HPC architectures (Figure 1) Notably, these results were achieved while using 5x fewer resources. We highlight the design and implementation of workflow mini-apps, showcasing their ability to offer simple, portable, reproducible, and managed representations of complex workflows that are both cost- and energy-efficient. This work advances the science of workflows, providing valuable tools for performance characterization and optimization in current and future computing systems.

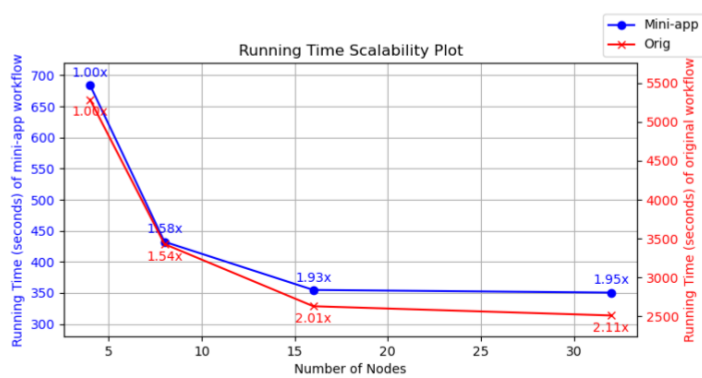

Figure 1: Scalability of Inverse Problem workflow mini-app

[1] Kilic, Ozgur Ozan, et al. "Workflow Mini-Apps: Portable, Scalable, Tunable & Faithful Representations of Scientific Workflows." arXiv preprint arXiv:2403.18073 (2024).