Reference ID: 11100436593_Belotti

Submission Date and Time: 10/29/2019 3:59:10 AM

Author Names & Affiliations

Submitting author: Pietro Belotti - Fair Isaac Corp

Additional authors: None

Contact Email Address (for NSF use only): (hidden)

Research domain(s), discipline(s)/sub-discipline(s)

Optimization/mixed integer nonlinear optimization.

Title of Response

more clusters needed

Abstract

There is no lack of fast, reliable software for solving optimization problems. What is missing is cores where these software packages can be run, both in public (e.g. the Neos optimization servers) and in academic and research institutions.

Question 1 (maximum 400 words) – Data-Intensive Research Question(s) and Challenge(s). Describe current or emerging data-intensive/data-driven S&E research challenge(s), providing context in terms of recent research activities and standing questions in the field. NSF is particularly interested in cross-disciplinary challenges that will drive requirements for cross-disciplinary and disciplinary-agnostic data-related CI.

Large-scale optimization problems with millions of variables and/or constraints, particularly in the context of global optimization of neural network weights with special attention to robustness to adversarial attacks.

Question 2 (maximum 600 words) – Data-Oriented CI Needed to Address the Research Question(s) and Challenge(s). Considering the end-to-end scientific data-to-discovery (workflow) challenges, describe any limitations or absence of existing data-related CI capabilities and services, and/or specific technical and capacity advancements needed in data-related and other CI (e.g., advanced computing, data services, software infrastructure, applications, networking, cybersecurity) that must be addressed to accomplish the research question(s) and challenge(s) identified in Question 1. If possible, please also consider the required end-to-end structural, functional and performance characteristics for such CI services and capabilities. For instance, how can they respond to high levels of data heterogeneity, data integration and interoperability? To what degree can/should they be cross-disciplinary and domain-agnostic? What is required to promote ease of data discovery, publishing and access and delivery?

Clusters are available in most academic and research institutions and are increasing based on the increasing demand, but queuing time is usually in the order of days. Given the speed at which information can be gathered, summarized and visualized with today's software, computation time is the true bottleneck of the research process. An increased availability of computing machinery would make for a faster turnaround in algorithmic changes and data analysis.

Question 3 (maximum 300 words) – Other considerations. Please discuss any other relevant aspects, such as organization, processes, learning and workforce development, access and sustainability, that need to be addressed; or any other issues more generally that NSF should consider.

-- End Submission –