Symposium of the Center for Network and Storage Enabled Collaborative Computational Science



Contribution ID: 18

Type: Presentation

Highly scalable metadata management with signac

Friday 19 May 2017 11:00 (30 minutes)

Continually increasing computational resources and improved efficiency of parallelized software for data generation and manipulation in the field of scientific computation have led to the requirement of more systematic approaches for data management. We present a data management framework designed to work on both desktop computers and in high-performance computing environments with special emphasis on low entry barriers for both new and experienced users. The *signac* framework assists in the decentralized storage of data and metadata on the file system by providing all basic components needed for building simple to complex data pipelines largely agnostic of data source and format. These managed data spaces are immediately searchable through a homogeneous interface and in this way more accessible to data owners, but also collaborators. Sharing of data across different endpoints is simplified through the generation of metadata indices that contain information about data provenance and current location. The framework's data model is designed not to require absolute commitment to the presented implementation. This reduces barriers for the integration into existing workflows and increases the accessibility to archived data sets. The presented approach simplifies the production of scientific results and collaboration on shared data sets.

Author: ADORF, Carl S. (University of Michigan)

Co-authors: DODD, Paul M. (University of Michigan); Prof. GLOTZER, Sharon C. Presenter: ADORF, Carl S. (University of Michigan)

Session Classification: Complementary Technology Solutions

Track Classification: Complementary Technology Solutions