Contribution ID: **117** Type: **Short Talk**

PandAna: A Python Analysis Framework for Scalable High Performance Computing in High Energy Physics

Wednesday, 19 May 2021 18:45 (13 minutes)

Modern experiments in high energy physics analyze millions of events recorded in particle detectors to select the events of interest and make measurements of physics parameters. These data can often be stored as tabular data in files with detector information and reconstructed quantities. Current techniques for event selection in these files lack the scalability needed for high performance computing environments. We describe our work to develop a high energy physics analysis framework suitable for high performance computing. This new framework utilizes modern tools for reading files and implicit data parallelism. Framework users analyze tabular data using standard, easy-to-use data analysis techniques in Python while the framework handles the file manipulations and parallelism without the user needing advanced experience in parallel programming. In future versions, we hope to provide a framework that can be utilized on a personal computer or a high performance computing cluster with little change to the user code.

Primary authors: GROH, Micah (Fermi National Accelerator Laboratory); BUCHANAN, Norm (Colorado State University); DEREK, Doyle (Colorado State University); KOWALKOWSKI, Jim (Fermilab); PATERNO, Marc (Fermilab); SEHRISH, Saba (Fermilab)

Presenter: GROH, Micah (Fermi National Accelerator Laboratory)

Session Classification: Accelerators

Track Classification: Offline Computing