Utilities to Execute Pipelines

Diego Juarez, Jorge Munoz

Department of Physics, The University of Texas at El Paso, El Paso, TX 79968

Abstract

To reduce human error in creating input files for simulations, especially when handling hundreds or even thousands of jobs, we developed Utilities to Execute Pipelines (UTEP), a robust and versatile Python framework to streamline the generation and analysis of large-scale scientific datasets. This framework automates the set up, submission, and execution of scientific codes, generally taking care of the minutiae. Post-processing tasks, such as the development of machine learning models, are easier because of a common data structure. UTEP was deployed on Perlmutter, a National Energy Research Scientific Computing Center (NERSC) supercomputer, but it can be deployed on any Unix system. Here we present instances of UTEP applied to projects in computational thermodynamics of alloys, lattice dynamics of crystals, charge distributions in a semiconductor, and a multi-objective genetic algorithms used to test crystal stability.