

Dask Tutorial

Friday 16 September 2022 14:00 (1 hour)

Dask provides a foundation to natively scale Python libraries and applications. Dask collection libraries like `dask.array` and `dask.dataframe` mimic the ubiquitous APIs of NumPy and Pandas to parallelize and/or distribute NumPy-like and Pandas-like workflows. The `dask.delayed` collection supports parallelization of custom algorithms. In this tutorial we will introduce the core Dask collections, the concepts behind them (partitioned objects represented by task graphs), and Dask's distributed execution engine that is compatible with common HEP batch compute systems. Finally, we will introduce recently developed Dask collections that support partitioned and distributed representations of awkward arrays and boost-histogram objects.

Author: DAVIS, Doug

Presenter: DAVIS, Doug

Session Classification: Plenary Session Friday