

# Executing Analysis Workflows at Scale with Coffea+Dask+TaskVine

*Tuesday 10 October 2023 15:00 (30 minutes)*

During this talk I will present our experiences executing analysis workflows on thousands of cores. We use TaskVine, a general-purpose task scheduler for large scale data intensive dynamic python applications, to execute the task graph generated by Coffea+Dask. As task data becomes available, TaskVine adapts the cores and memory allocated to maximize throughput and minimize retries. Additionally, TaskVine tries to minimize data movement by temporarily and aggressively caching data at the compute nodes. TaskVine executes these workflows without a previous setup in the compute nodes, as it dynamically delivers the dependencies using conda-based environment file.

**Author:** TOVAR LOPEZ, Benjamin (University of Notre Dame)

**Presenter:** TOVAR LOPEZ, Benjamin (University of Notre Dame)

**Session Classification:** Plenary Session Tuesday