



Shay Liu^{1,2}

¹ Indiana University

Mentors: Supreeth Suresh and Cena Miller²

² NCAR CISL

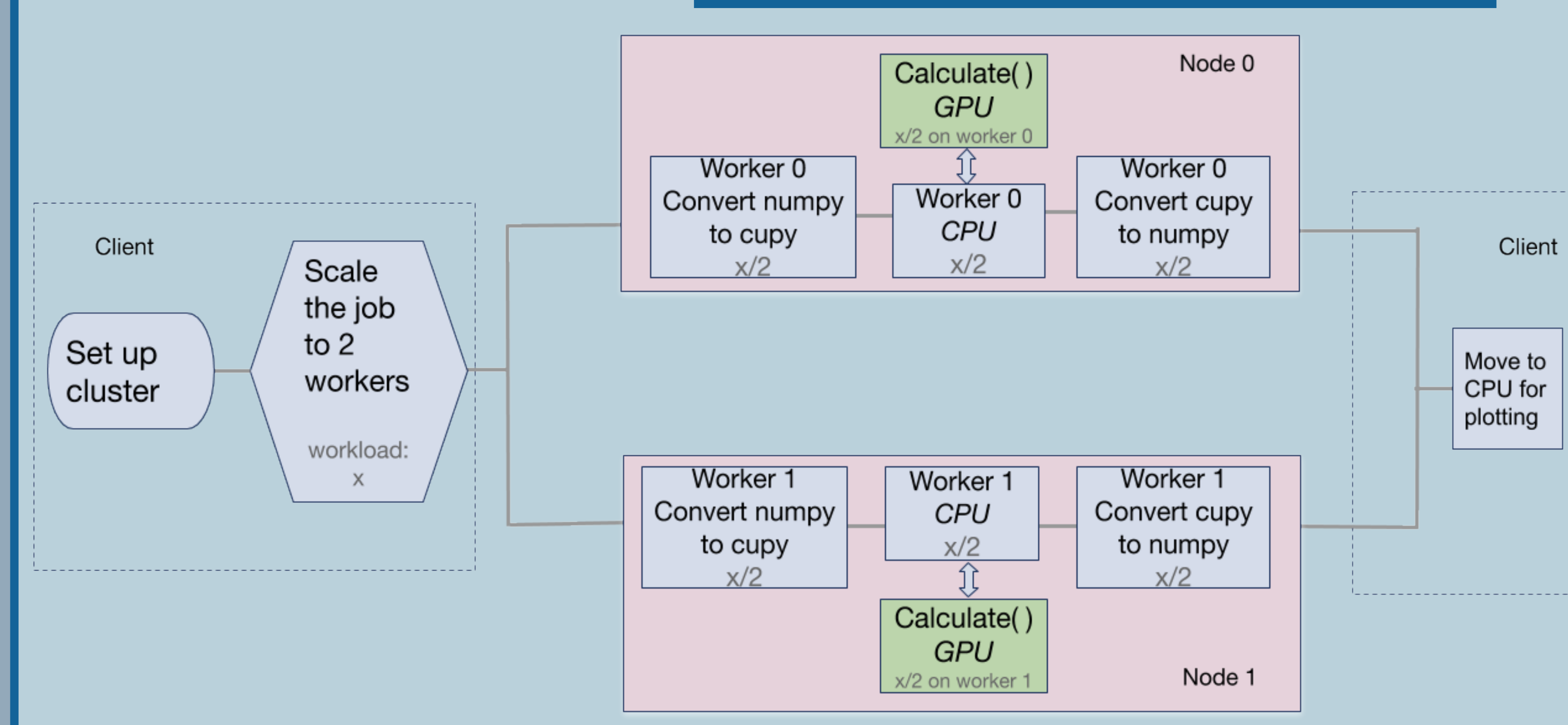
Introduction

- Model analysis is traditionally done on CPUs
- Model analysis is often embarrassingly parallel and compute intensive
- These types of tasks are well suited for GPU acceleration
- *FastEddy*[®] is a GPU-based large eddy simulation (LES) model developed in RAL, which produces large datasets

Project Goals

- Become familiarized with the architecture of GPUs
- Perform *FastEddy*[®] data analysis on GPUs
 - Single GPU execution
 - Multi-GPU execution
- Prototype a simplified GPU acceleration of the data science phase

Dask Execution Workflow



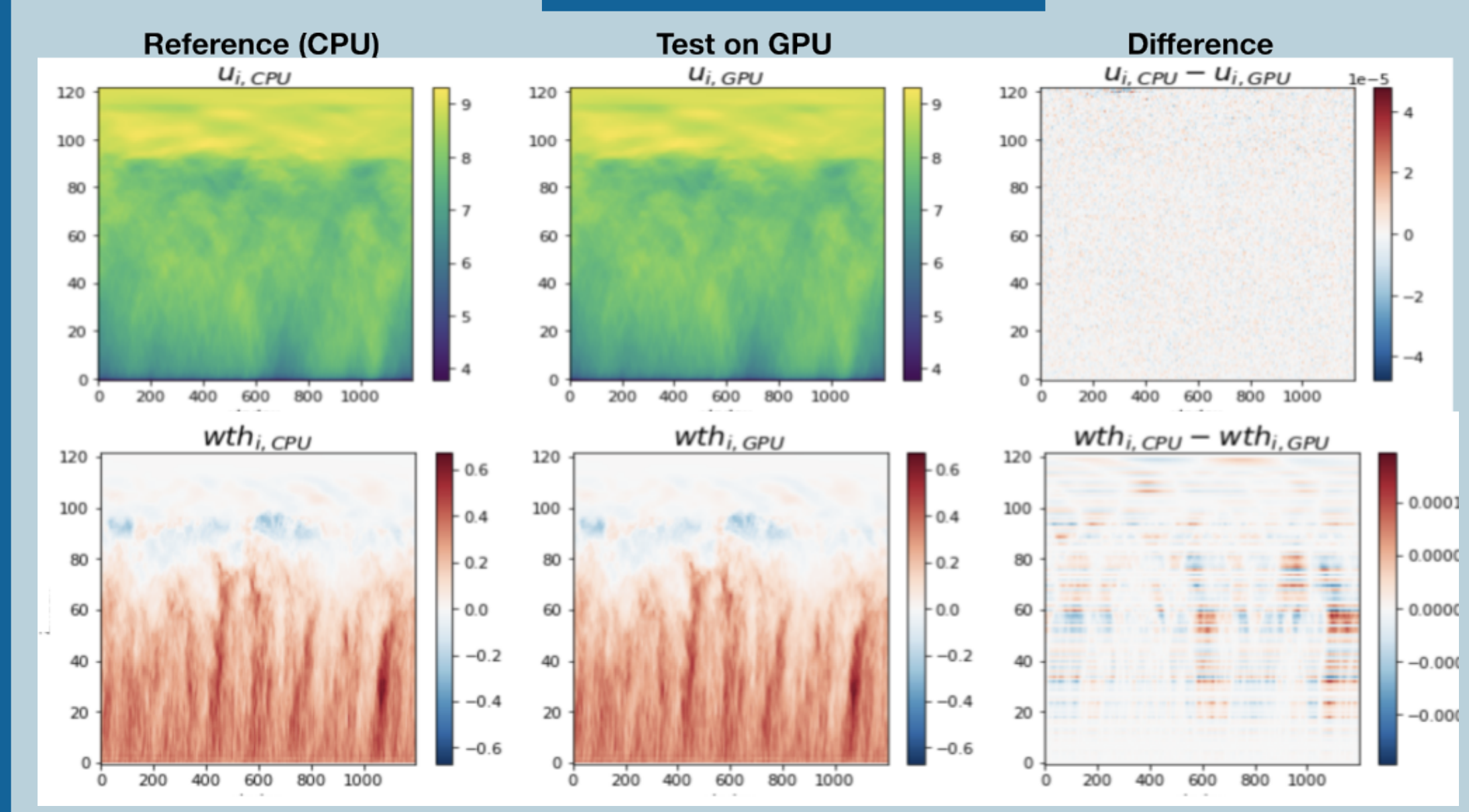
- When two files are analyzed simultaneously, the job is scaled up to two workers. Each worker performs analysis on one file.
- Two delayed functions, read() and calculate(), are executed on each worker. calculate() is executed on GPU.
- The computed data are then brought back to the CPU for plotting.

JupyterLab Dashboard during Multi-node GPU Execution

The screenshot shows the JupyterLab interface during multi-node GPU execution. It includes a code editor with Python code, a Task Stream showing task progress over time, and several resource usage graphs: CPU Utilization, GPU Memory (25.48 GB / 34.14 GB), Dask Progress (total: 4, in-memory: 2), and Dask Graph.

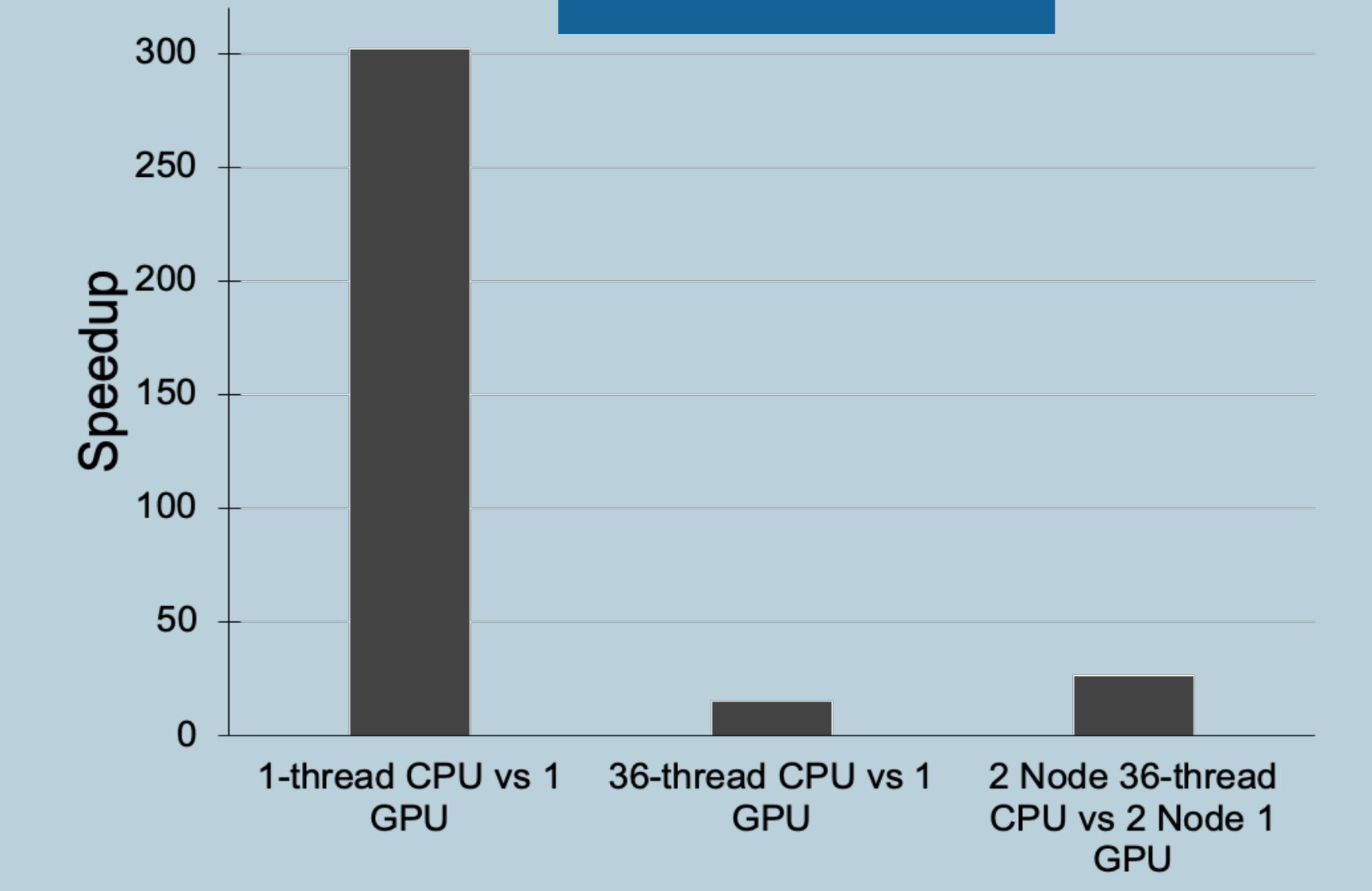
A - GPU memory usage on each worker
 B - Progress of dask execution
 C - Task progress with respect to time

Validation



- Acceptable differences between simple CPU/GPU calculation (e.g., horizontal perturbation wind speed, top panel)
- More sophisticated computations would require more investigation (e.g., turbulent heat flux, bottom panel)

Results



- 300x speedup: achieved comparing the original analysis on single-thread CPU vs single GPU
 - 15x speedup: achieved comparing one node 36-thread CPU vs single GPU
 - 26x speedup: achieved comparing 2 nodes 36-thread CPU vs 2 nodes with single GPUs
- * Casper specs: 384 GB memory per node, 2 18-core skylake CPUs, 1 NVIDIA Quadro GP100 GPU 16GB per node.

Acknowledgements

Mentors: Supreeth Suresh and Cena Miller
SIParCS Admins and CODE Assistants: AJ Lauer, Virginia Do, Jerry Cycone, Jess Hoopengardner
CISL Technical Support: Anderson Banihirwe and Mick Coady
Research Advice: Dr. Jeremy Sauer