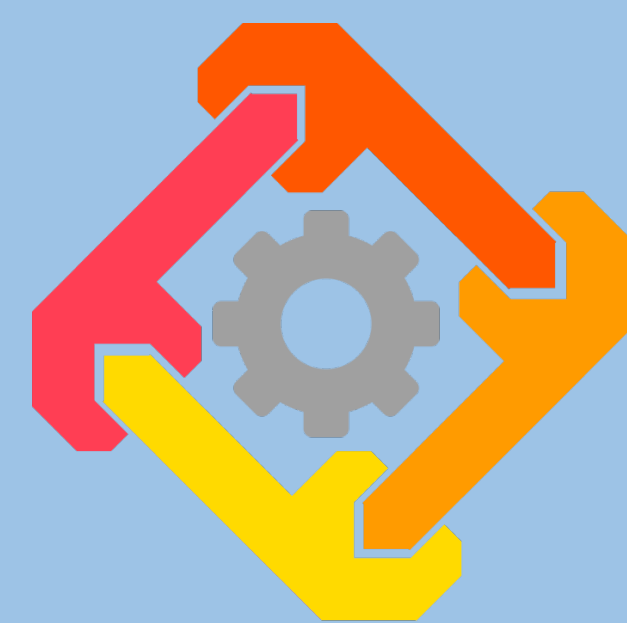


A Pedagogic Module for Energy-Aware Workflow Execution on Distributed Computing Platforms



Derrick Luyen
University of Hawai'i at Mānoa – Information and Computer Sciences
Advisor: Dr. Henri Casanova

What We Proposed

Motivations

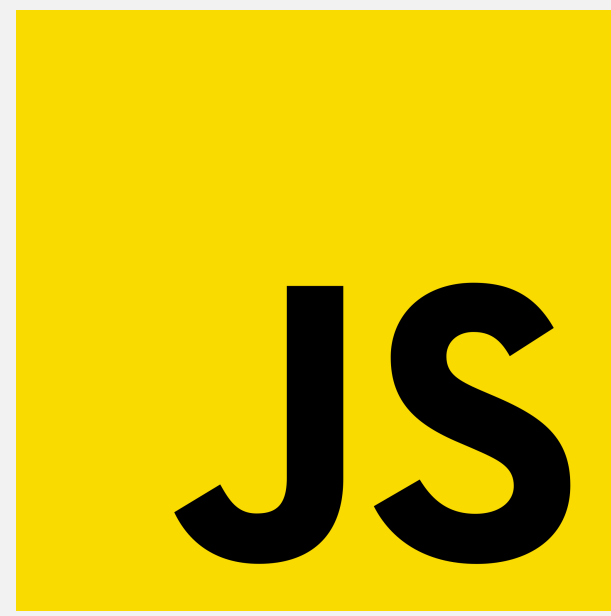
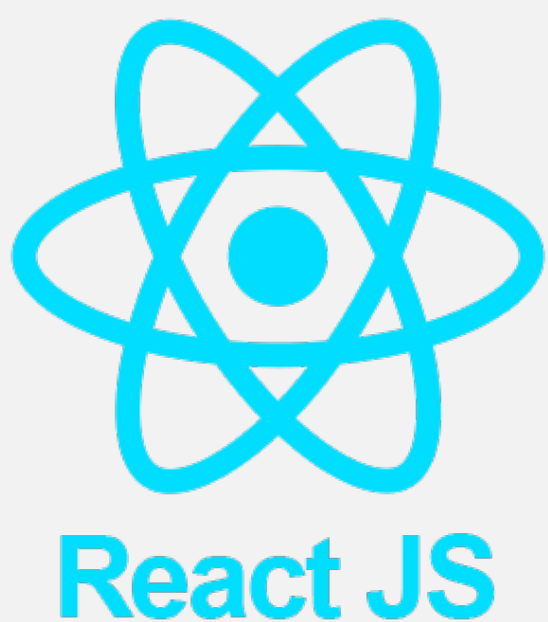
Parallel and Distributed Computing (PDC) topics are crucial as part of Computer Science curricula, they are difficult to teach in a hands-on manner on real systems. An alternative is to teach them hands-on in the browser using interactive *simulations*.

Objectives

Design, implement, publish, and evaluate a simulation-driven, self-contained, in-the-browser pedagogic module for students to learn and reason about the trade-offs between power efficiency and performance when executing scientific workflow applications on distributed computing platforms.

What We Accomplished

- Designed a pedagogic narrative, along with interactive pedagogic activities and homework problems that focus on the power efficient execution of a workflow application in two cases:
 - A) On a local cluster with power management via pstates and choosing how many nodes are on/off
 - B) On a local cluster augmented with a remote cloud that has low carbon footprint but reachable via a slow wide-area network



- Implemented a web frontend with Gatsby-React and JS



- Implemented a backend simulator in C++ using the WRENCH API



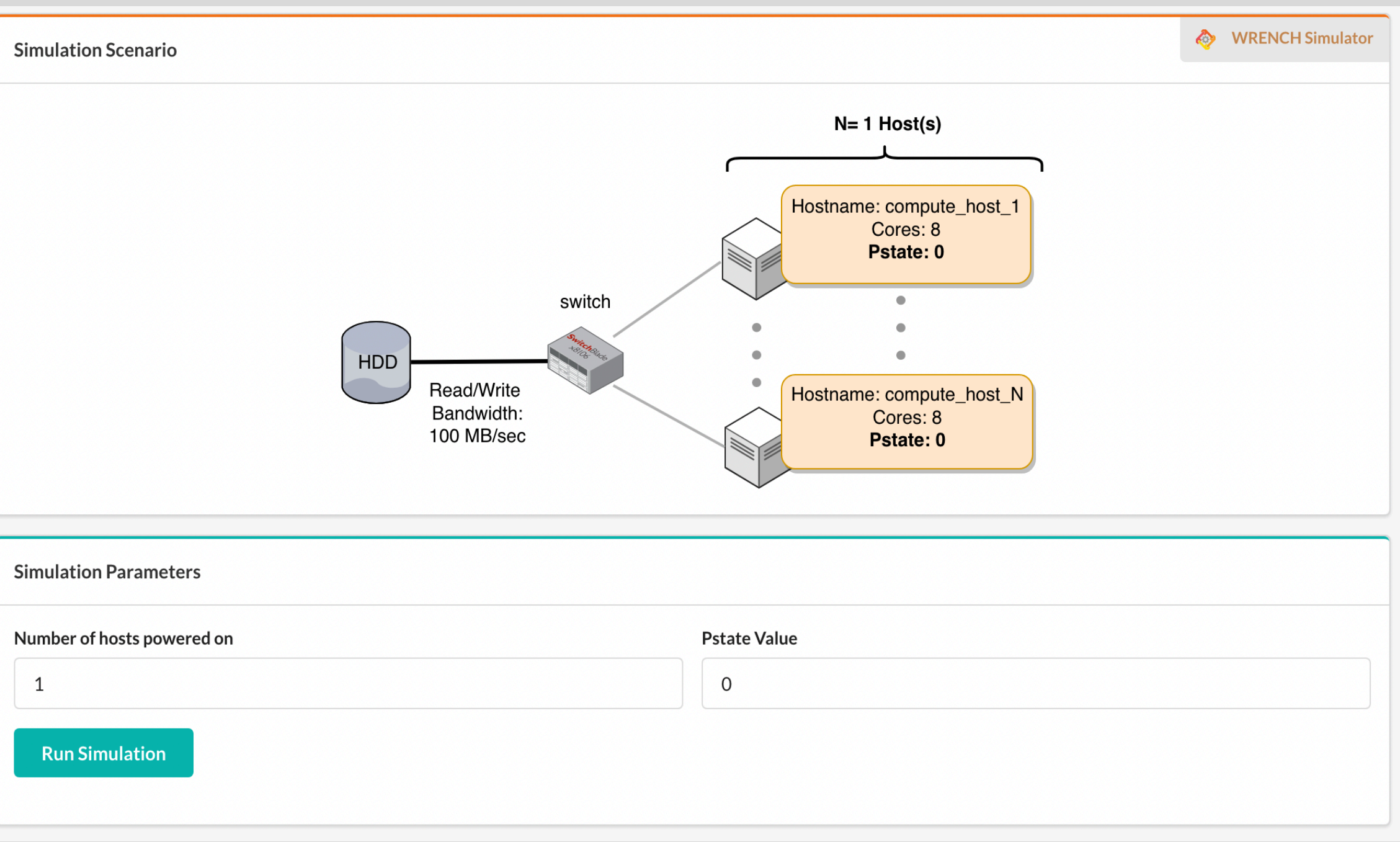
eduWRENCH - Pedagogic Modules
Parallel and Distributed Computing Courseware

- Deployed the module on the <https://eduwrench.org> pedagogic site
- Used the module at UHM in ICS 632 in Fall 2021 & collected student feedback and self-assessments of knowledge acquisition

Tab 1: Local Cluster Simulation

pstate	Clock rate (GHz)	Core speed (Gflops)	Power consumption (watts)
0	1.8	22.43	120
1	2.1	26.17	130
2	2.4	29.91	140
3	2.7	33.65	150
4	3.0	37.39	160
5	3.3	41.13	170
6	3.5	43.00	190

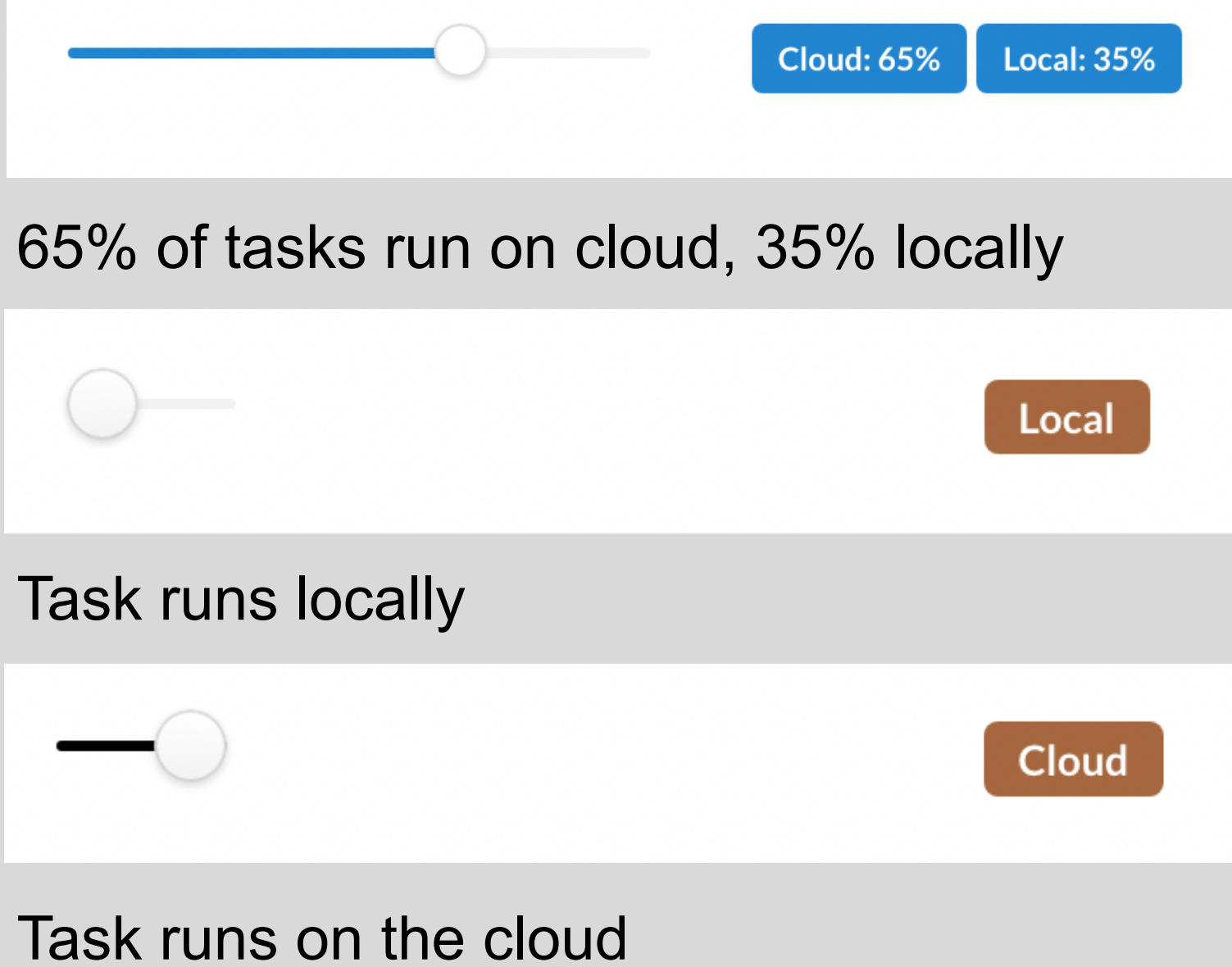
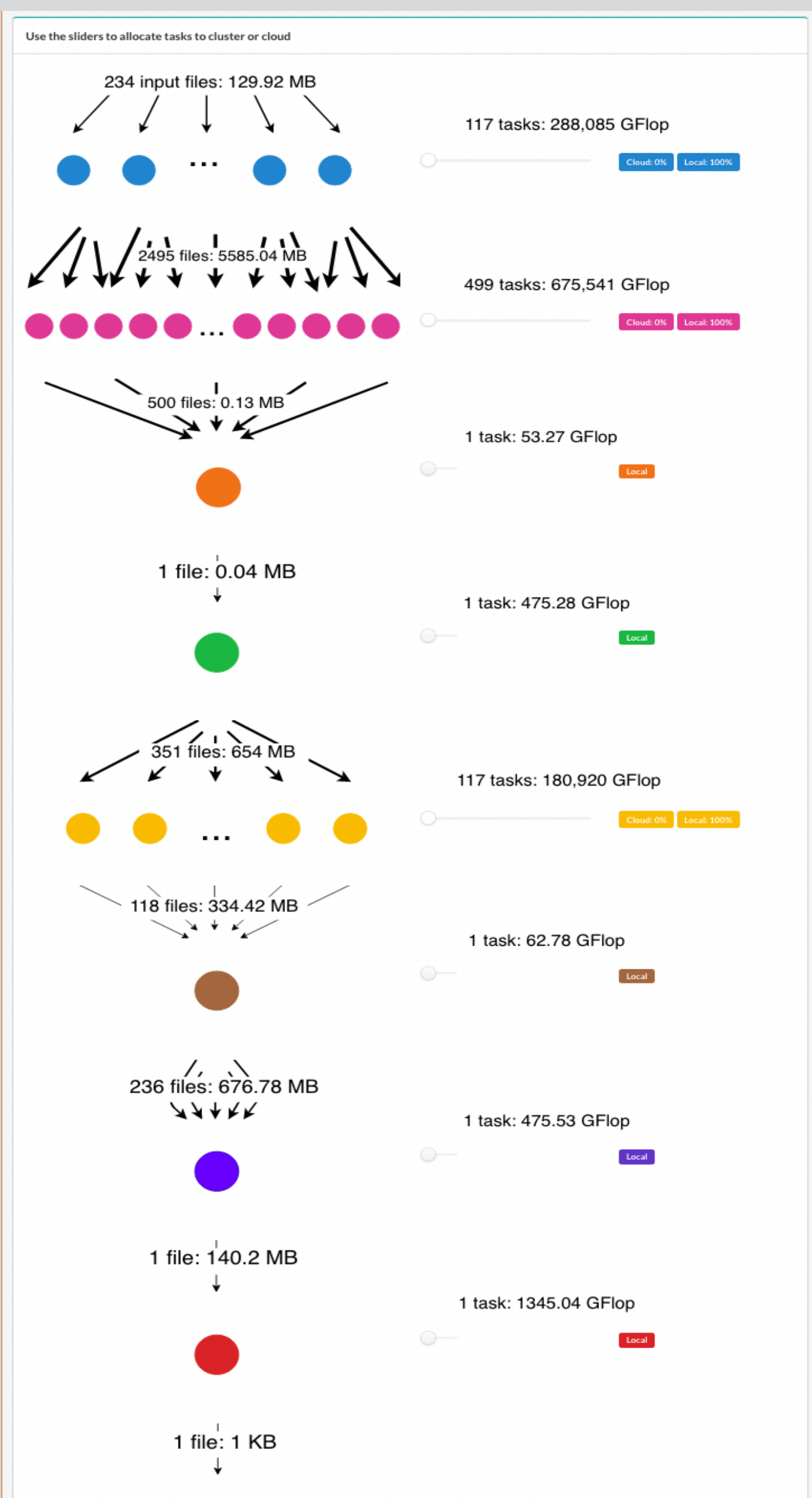
P-state values table



Montage Workflow

Sample question:
According to your boss, it is only necessary to run the workflow in under 3 minutes (e.g., running in 2 minutes brings no extra benefits when compared to running in 3 minutes).
Configure the cluster so as to have the lowest CO2 emission provided the workflow executes in less than 3 minutes.

Tab 2: Local Cluster + Remote Cloud Simulation



Sample Question: The two "extreme" options are to either run all tasks on the local cluster or run all tasks on the remote cloud. Note that the input data to the workflow initially resides on the storage system of the local cluster. Use the simulation to evaluate both these options in terms of workflow execution time and answer:

- *Considering total execution time and carbon footprint, is one option strictly better than the other, and if so which one?*
- *If the wide-area link were of infinite bandwidth, how much faster would the cloud-only execution be?*

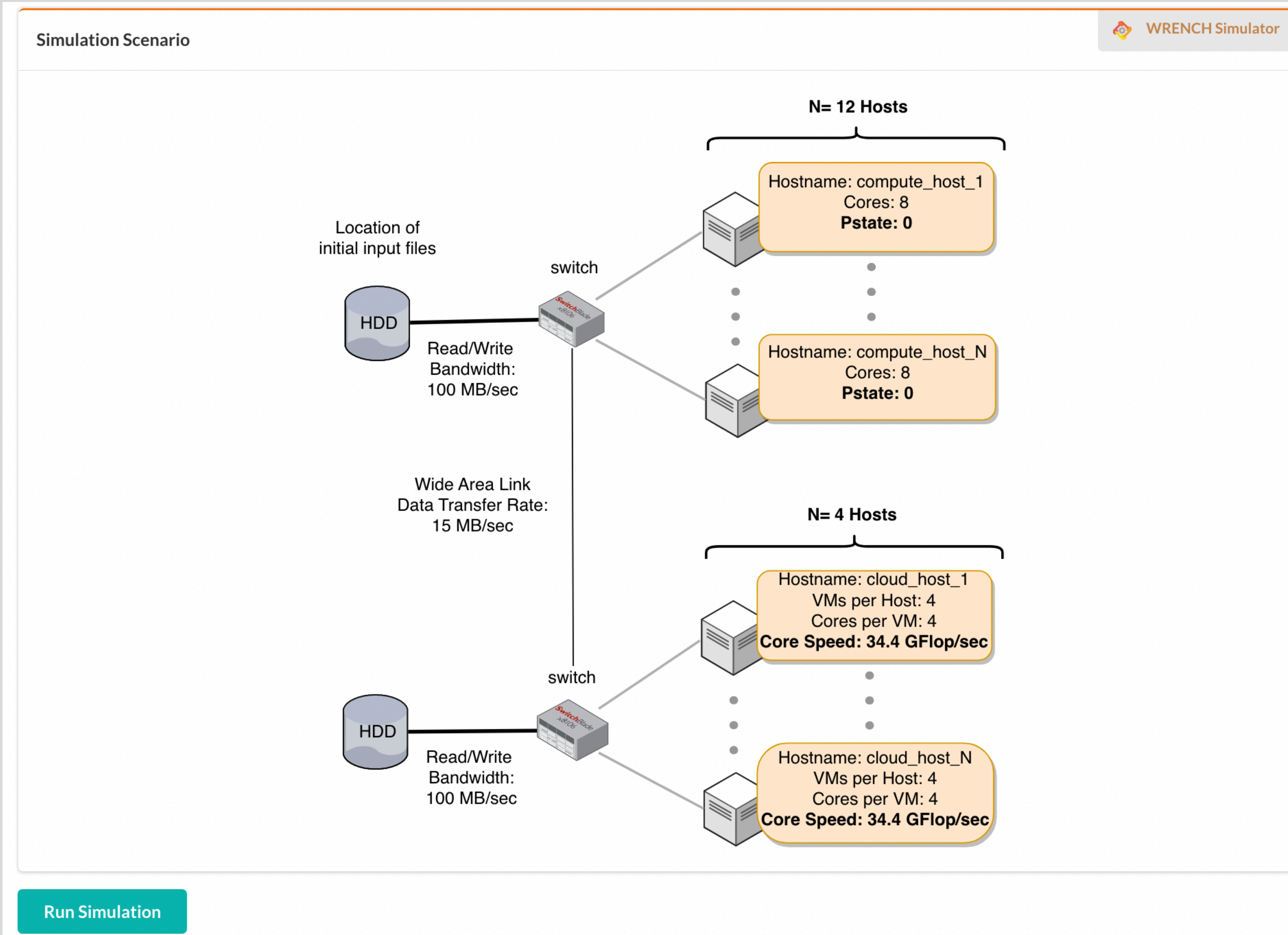
Module Evaluation Results

Student Feedback

TABLE I
STUDENT FEEDBACK ($n = 11$).

Question	Choices	#Answers
How easy / difficult is the assignment?	very easy	1
	somewhat easy	6
	neither easy nor difficult	4
	somewhat difficult	-
	very difficult	-
How useful is the assignment?	very useful	5
	useful	3
	somewhat useful	3
	of little use	-
	not useful	-
To what extent did the assignment help you learn new things?	to a great extent	5
	to a moderate extent	4
	to some extent	2
	to a small extent	-
	not at all	-
Are you interested in learning more about this topic?	yes	10
	no	1
How useful is simulation in this assignment?	very useful	6
	useful	3
	somewhat useful	3
	of little use	-
	not useful	-
How valuable is the overall learning experience in the module?	very much	7
	quite a bit	3
	somewhat	1
	a little	-
	not at all	-

Tab 2: Local Cluster + Remote Cloud (Cont.)



- Local cluster and remote cloud platform graph
- Simulation input is provided via sliders to decide what fractions of tasks run locally / remote
- Number of compute hosts is preset for local cluster and remote cloud
- When running remote tasks, data must be transferred back and forth between the local cluster and the remote cloud