Teaching Parallel and Distributed Computing Concepts in Simulation with WRENCH

Ryan Tanaka[1] Rafael Ferreira da Silva [1]
Henri Casanova [2]

[1] Information Science Institute / USC, USA
[2] University of Hawai`i at Mānoa, USA

Teaching Challenges

- Teaching Parallel and Distributed Computing (PDC) and High Performance Computing (HPC) concepts in Computer Science curricula should be done more and earlier
- Teaching everything “on the blackboard” is not effective, and students should *learn in a hands-on manner*
- One option: provide students access with some hardware and software platform to learn/apply PDC and HPC concepts
    - e.g., some on-campus cluster
- This comes with challenges!

Real Platforms: Participation Challenge

- An institution may not have an adequate platform
  - Or none readily available for teaching purposes
- There are several solutions:
  - Build a low-cost platform (e.g., raspberry pies, clusters of SoCs)
  - Use virtualization/container technology (e.g., locally, in some cloud)
- But all of these limit what can be done/learned because of their specs and scales

Real Platforms: Pedagogic Challenges

- Real-world stuff gets in the way of learning
  - Possibly intricate platform access mechanisms and policies
  - Platform downtimes (planned or unplanned)
  - Competition for the platform among students and with other users

→ Class and instructor time not spent on learning objectives

- Platform's specifics get in the way of learning
  - "If we had more cores, then this would happen…"
  - "If the network was different, then this wouldn't work as well…"
  - "If we had less RAM, then this would break…"

→ Many learning objectives cannot be achieved hands-on

Simulation as an Alternative

- With simulation: no need for an actual platform, any arbitrary platform configuration, perfect repeatability, quick executions
- Used routinely for teaching in some areas of Computer Science (architecture, network)
- Time-and-again proposed and used for PDC/HPC education since the early 1990s
- Typically used with a “simulate and observe” strategy
  - Simulating the execution of code provided to students and that they cannot modify
  - Simulating the execution of code written by students, allowing them to develop/debug/run all in simulation

GOAL

- Develop a set of pedagogic modules that…

1. Target standard HPC/PDC Student Learning Objectives
2. Can be integrated piecemeal in existing courses starting at freshman levels
3. Rely on simulation to provide students with hands-on, interactive learning opportunities without need for any hardware platform

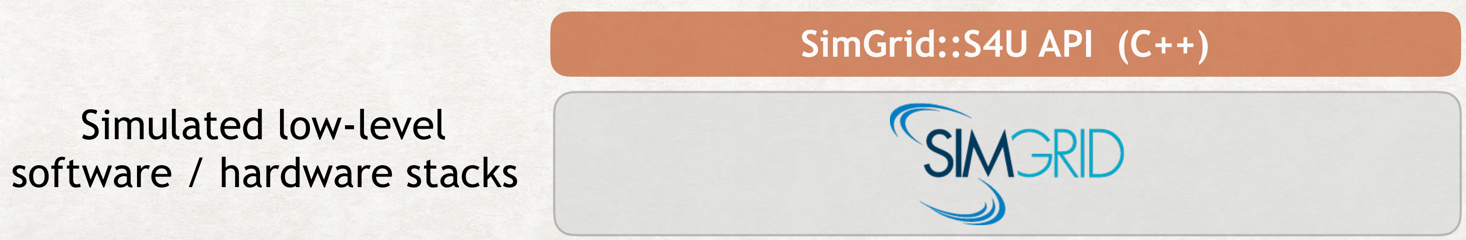
- All developed as part of the WRENCH project…

WRENCH

- To implement our pedagogic modules, we need to develop simulators
- These simulators should be scalable and accurate
- The SimGrid simulation framework has striven to provide both scalability and accuracy for more than a decade, so let's build on it…

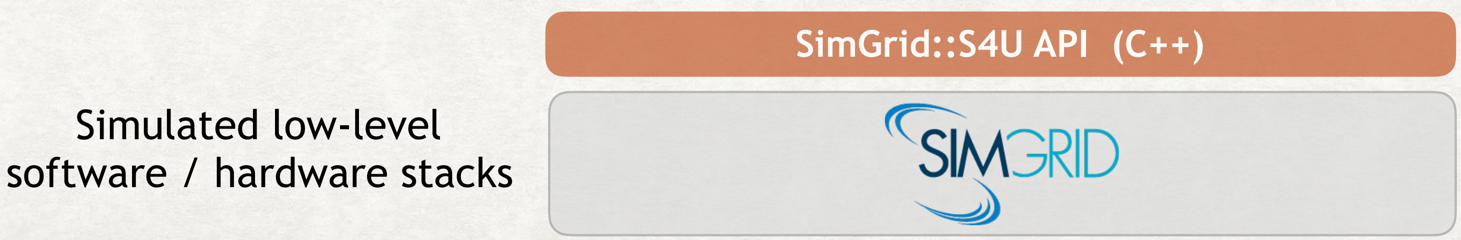
WRENCH

- To implement our pedagogic modules, we need to develop simulators
- These simulators should be scalable and accurate
- The SimGrid simulation framework has striven to provide both scalability and accuracy for more than a decade, so let’s build on it…



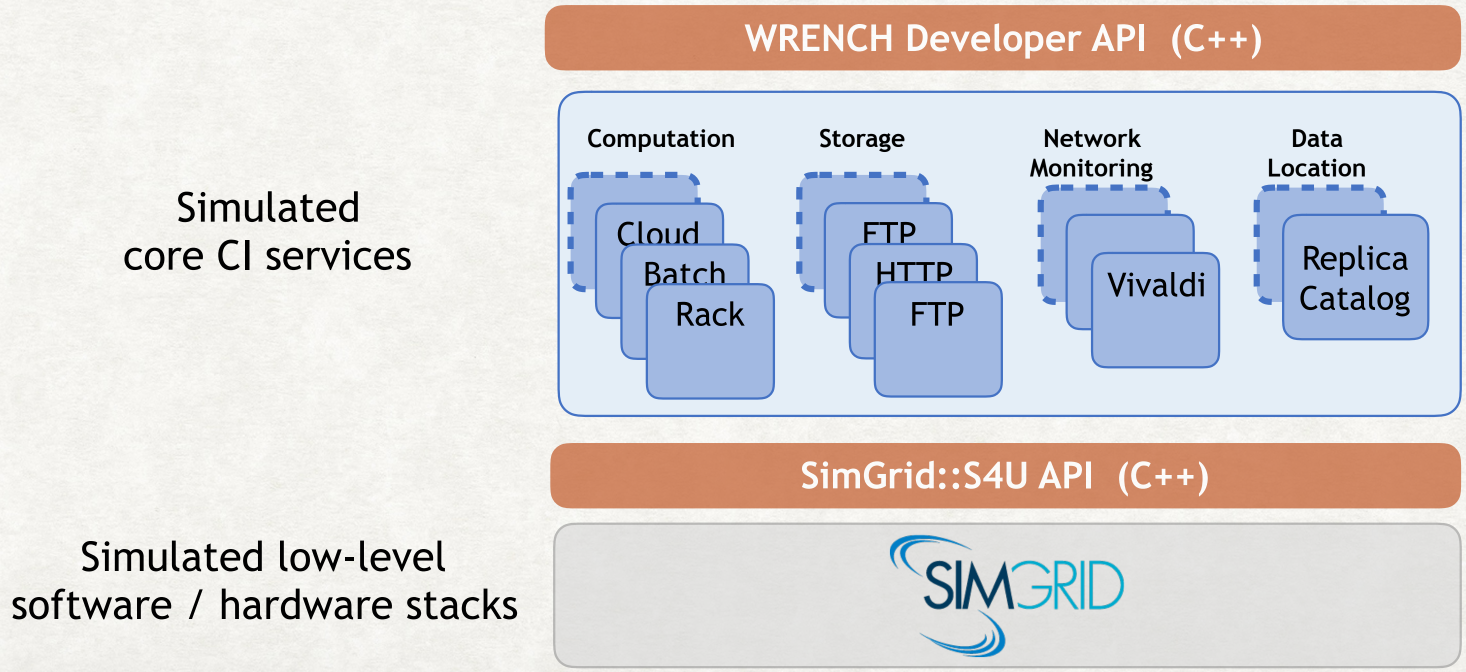
WRENCH

- To implement our pedagogic modules, we need to develop simulators
- These simulators should be scalable and accurate
- The SimGrid simulation framework has striven to provide both scalability and accuracy for more than a decade, so let's build on it…
- But SimGrid provides low-level abstractions, and thus writing simulators can be labor-intensive



WRENCH

- WRENCH builds on top of SimGrid to provide easy, high-level simulation abstractions
- Therefore, we can now have simulators that are accurate, scalable, and easy to develop
- Onward to “WRENCH Pedagogic Modules”



The WRENCH Pedagogic Modules

- Each module has:
  - A set of learning objectives and a narrative
  - One or more simulators that students can execute
  - Guided, practice, and open-ended questions
- The simulators are used by students in various modes:
  - Run-and-observe
  - Run-to-verify-expectations
  - Run-to-discover-answers
- Students only need a browser and Docker

Currently Available Modules

Principles of Computing and Distributed Computing

SINGLE-CORE COMPUTING
speed, work, RAM

MULTI-CORE COMPUTING
speedup, efficiency
idle time

I/O
HDD/SSD, data rates,
overlap with
computation

NETWORKING
latencies, bandwidth,
topologies,
contention

Applying Principles To Workflows

SCIENTIFIC WORKFLOWS
Basic concepts

WORKFLOWS AND DATA LOCALITY
network proximity
of data

WORKFLOWS AND PARALLELISM
multi-core,
multi-node

PROVISIONING RESOURCES
meeting performance
goals within budget

CURRENTLY AVAILABLE MODULES

Principles of Computing and Distributed Computing

SINGLE-CORE COMPUTING
speed, work, RAM

MULTI-CORE COMPUTING
speedup, efficiency
idle time

I/O
HDD/SSD, data rates,
overlap with
computation

NETWORKING
latencies, bandwidth,
topologies,
contention

Applying Principles To Workflows

SCIENTIFIC WORKFLOWS
Basic concepts

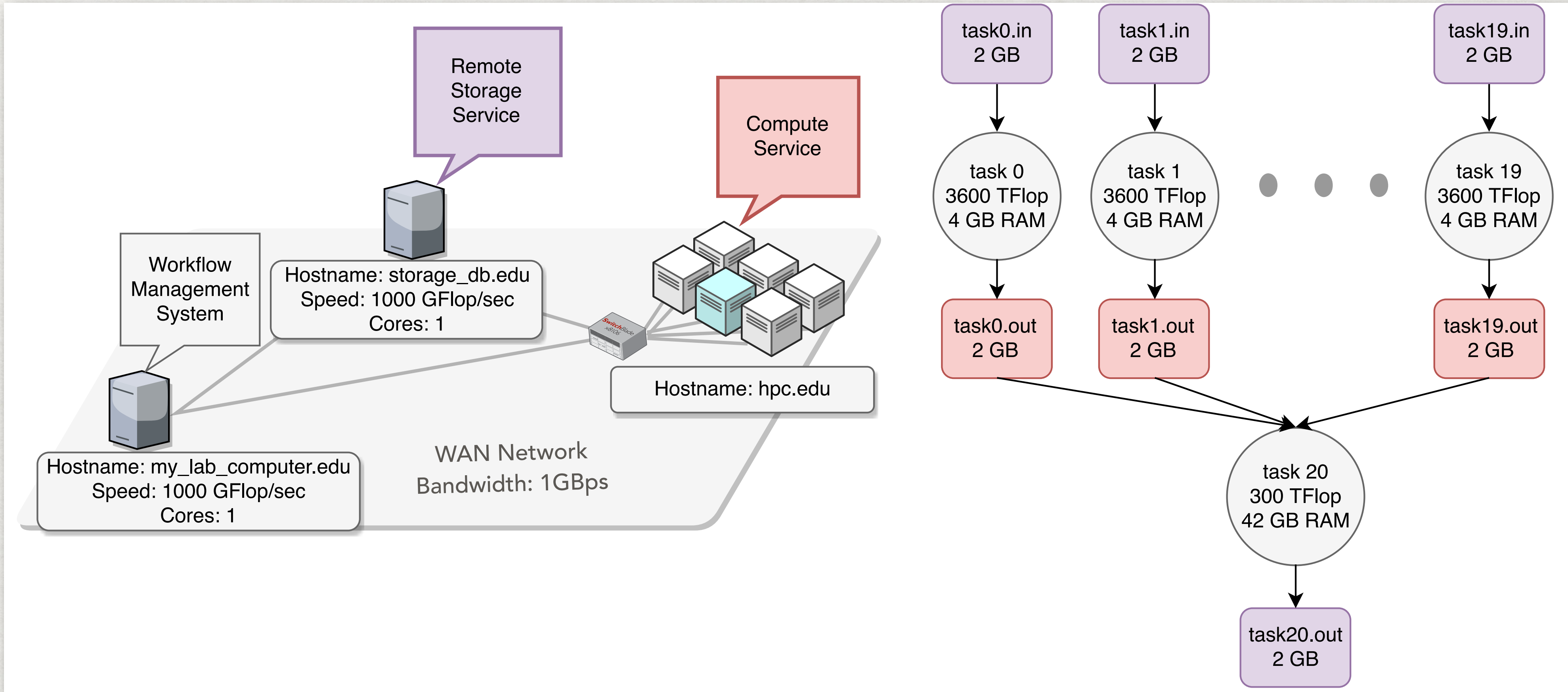
WORKFLOWS AND DATA LOCALITY
network proximity
of data

WORKFLOWS AND PARALLELISM
multi-core,
multi-node

PROVISIONING RESOURCES
meeting performance
goals within budget

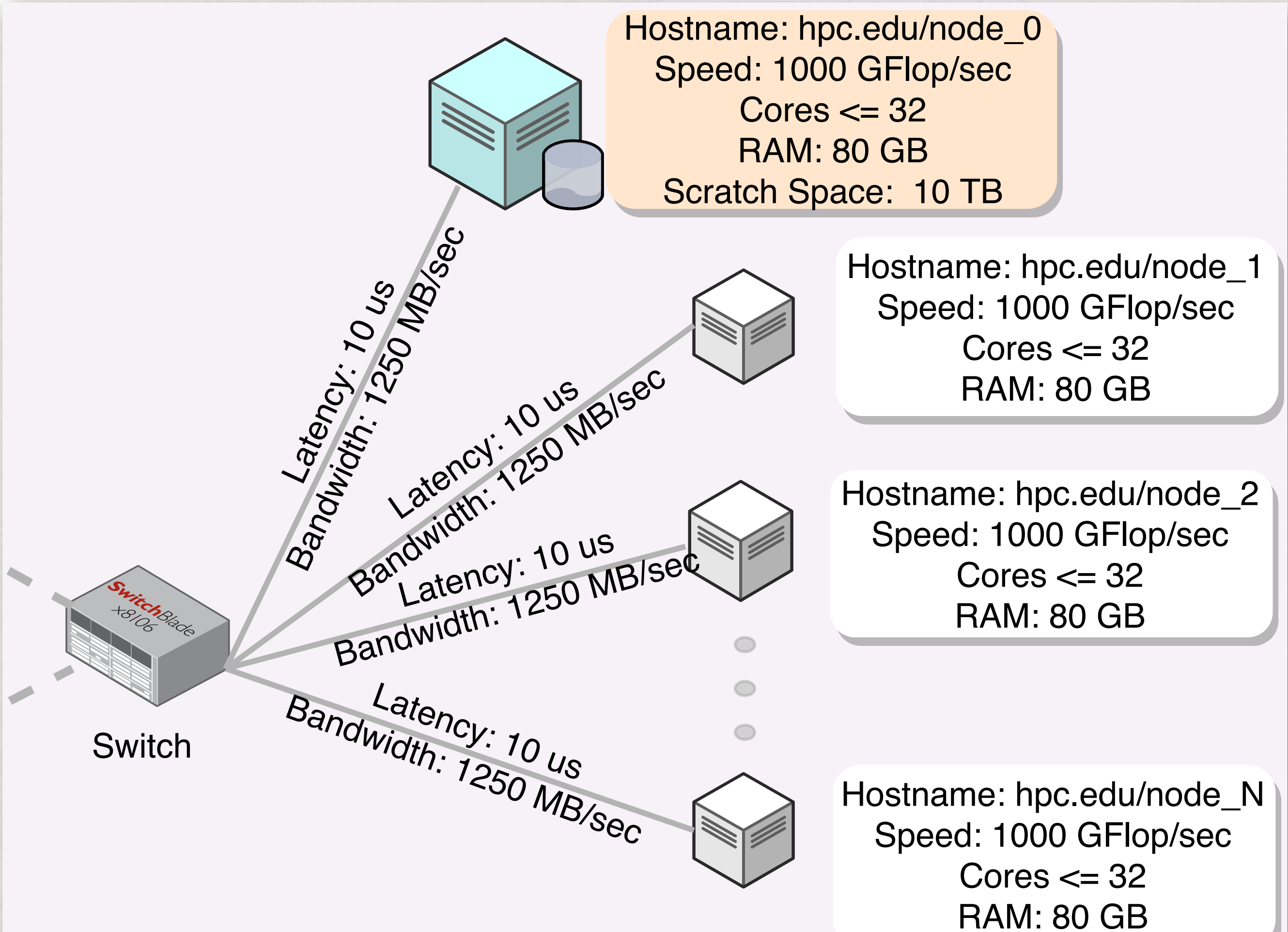
Let's look at one module…

Sample Module: Workflow and Parallelism



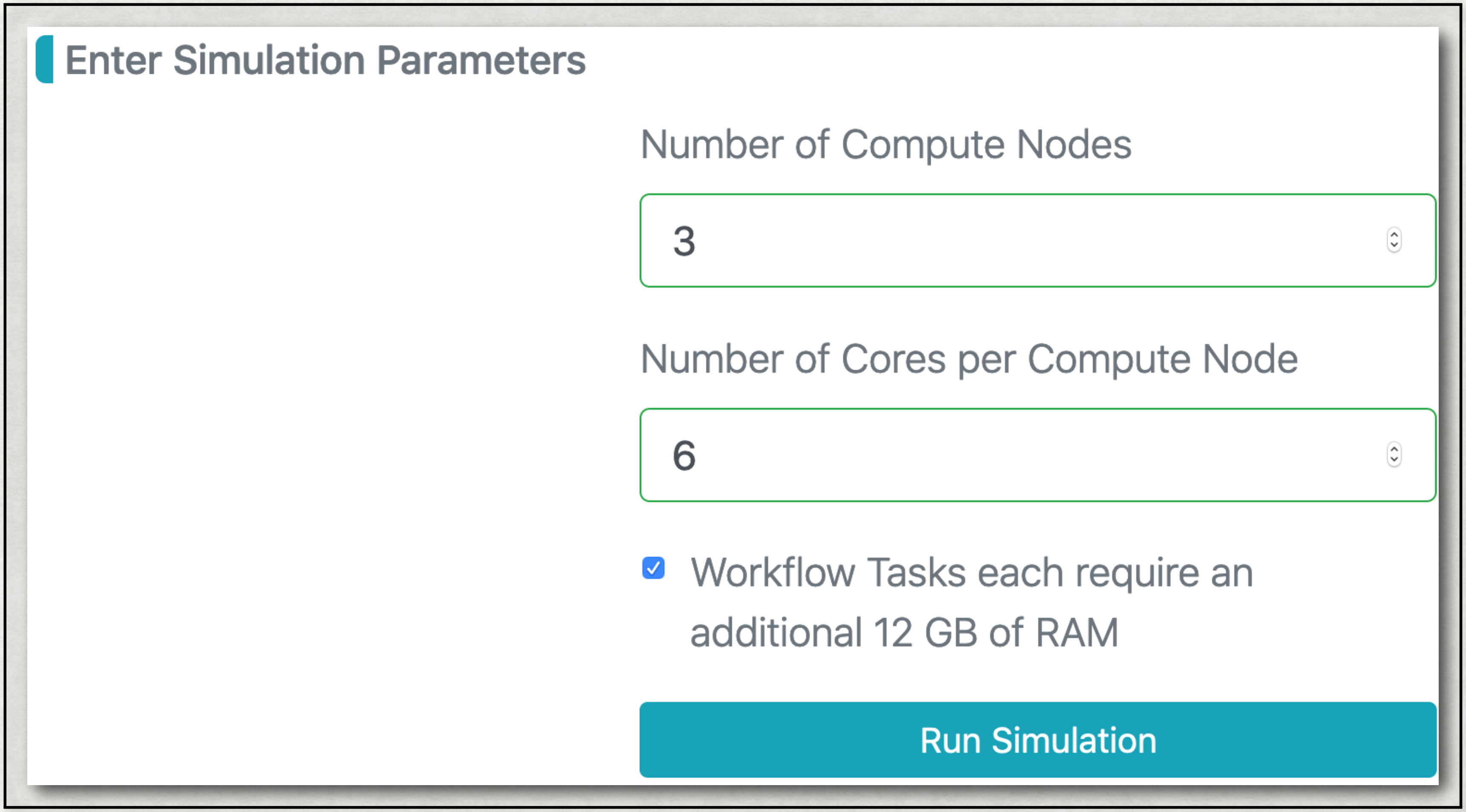
Students are shown a platform and an application at a high level

Sample Module: Workflow and Parallelism



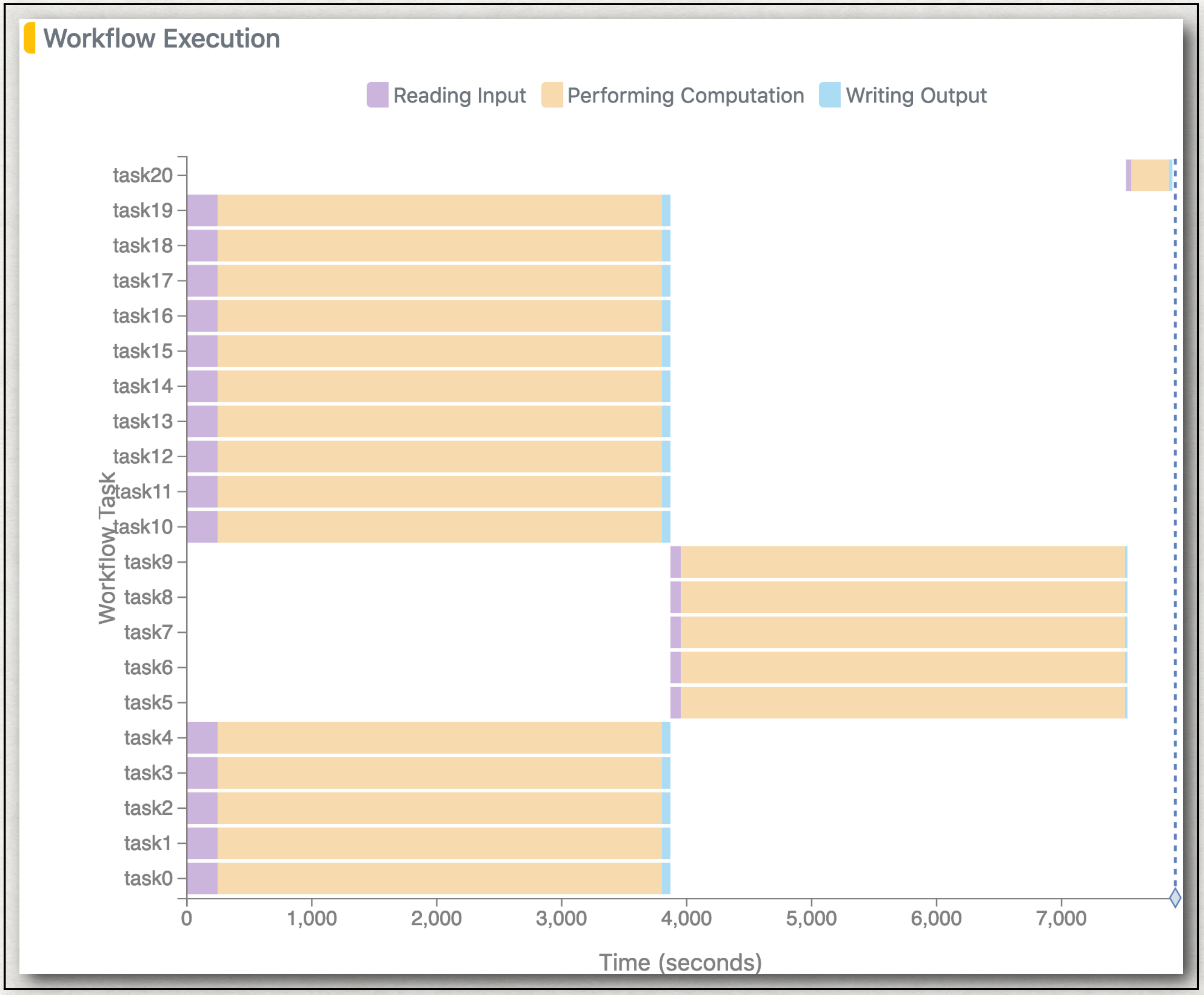
Students then learn about specifics

Sample Module: Workflow and Parallelism



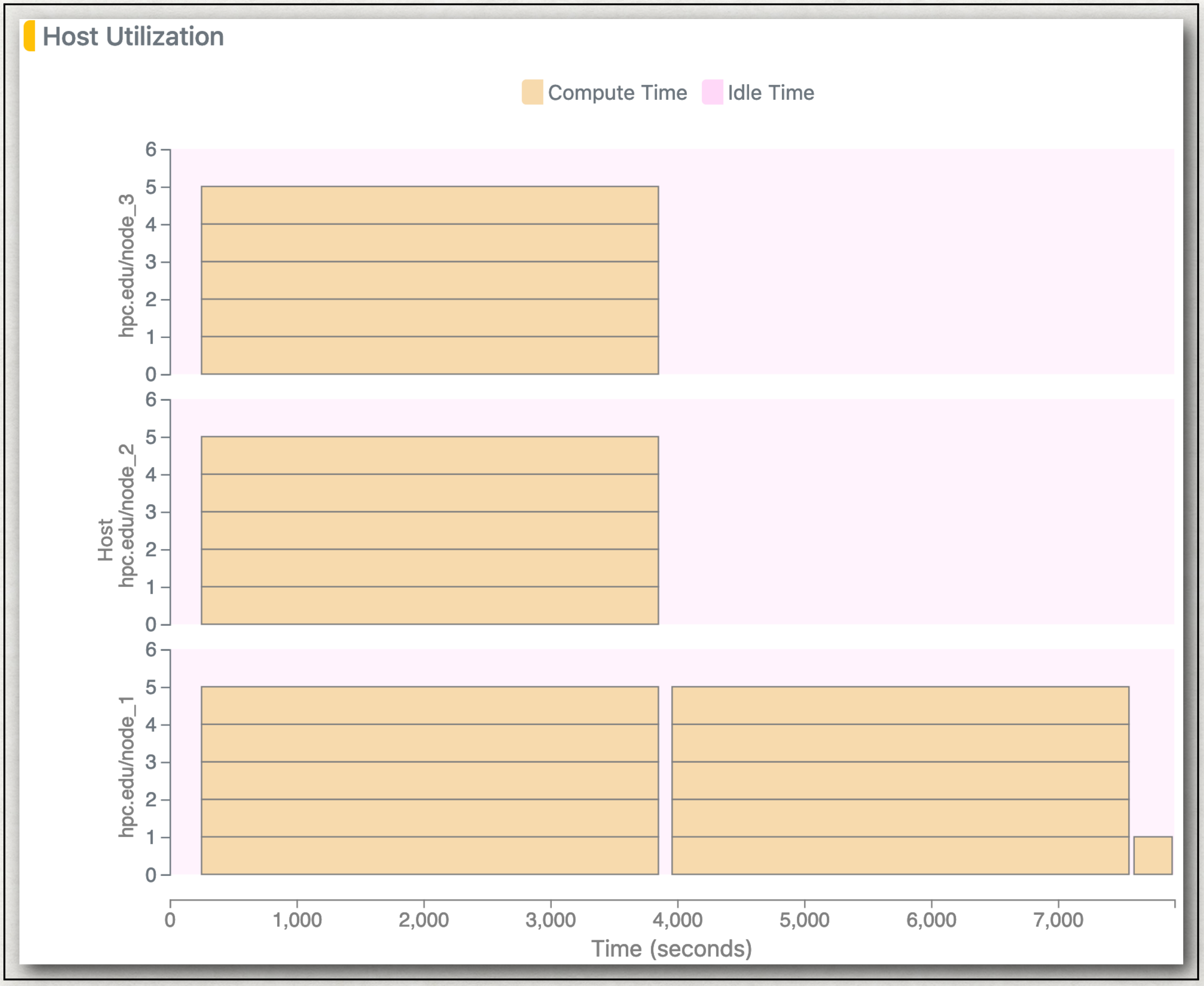
Students are able to try different specs and simulate application execution

Sample Module: Workflow and Parallelism



Execution Gantt chart for all tasks

SAMPLE MODULE: WORKFLOW AND PARALLELISM



Core utilization

Sample Module: Workflow and Parallelism

- Sample Questions #1: Assuming the cluster has 4 8-core compute nodes, what can we expect the execution time of the workflow to be? Write a simple formula. Run the simulation and check your results against the simulator.

- Sample Question #2: Assuming that you can add an arbitrary number of 5-core nodes, with the same per-core compute speed, is it possible to decrease the workflow execution time? Why or why not?

In-Class Evaluation (1)

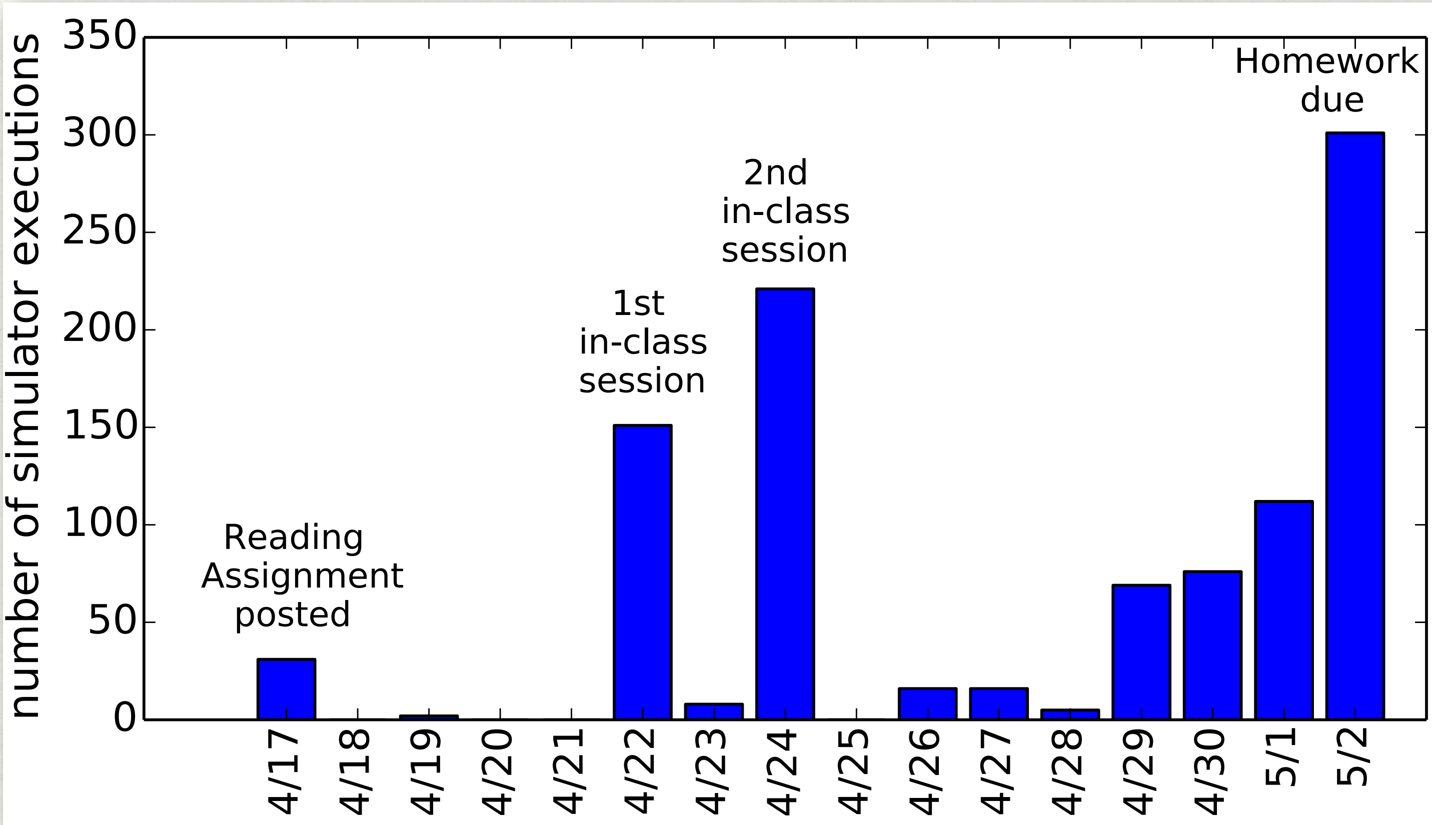
- These modules were used in the ICS332 course at UH Manoa in Spring 2019
  - And will be used next week again!
- Students were given:
  - A 30-minute lecture on PDC
  - A reading assignment in which students did foundational modules on their own
  - Two 75-minute in-class interactive sessions, going through modules with instructor scaffolding
  - A homework that consisted in completing the 2nd half of one of the workflow modules
  - Three final exam questions on these topics (10% of the exam grade)

In-Class Evaluation (2)

- In the evaluation we gathered:
  - Anonymous post questionnaire about the modules and about perceived learning
  - Anonymous pre and post knowledge tests
  - Non-anonymous grades for homework and exam questions
  - Non-anonymous time-stamps of simulation activities
- What we don’t have: a control group that does not use simulation
  - Unclear how that would be feasible/fair

WHAT WE LEARNED (1)

- Students are using the simulation



- 45 out of 55 students ran simulations (22 times on average)
- 40% of simulations were for input settings not suggested to them

WHAT WE LEARNED (2)

- Students are learning the material (thanks to simulation?)
- Students who never ran a simulation did poorly on the exam (but perhaps they were just unengaged)
- Pre to post knowledge tests: ~20% success rate to ~80% rate
- Interesting correlation between grades and number of simulation runs:

CORRELATION BETWEEN NUMBER OF SIMULATIONS EXECUTED AND AVERAGE GRADE ON PDC-FOCUSED FINAL EXAM QUESTIONS.

# of simulations	# of students	grade average
0	10	67.6
1-10	14	88.8
11-20	13	99.8
21-30	6	81.0
31+	12	75.5

WHAT WE LEARNED (3)

- Students had a positive experience
- Students appeared engaged during in-class sessions
- Perceived difficult level:
  - 60% “just right”, 23% “too difficult but useful”, 10% “too hard to be useful”, 7% “to easy to be useful”
- Written-in comments in course evaluation were very positive
- Two students since then have joined the WRENCH project as undergraduate researchers
- One technical issue: Docker on Windows 10 Home

Conclusion

- The modules are publicly available
- Many more are being developed as part of an NSF Cybertraining award
- Please contact us if you want to use these modules, or have feedback, or want to contribute

http://wrench-project.org/wrench-pedagogic-modules/