Hands-on Introduction to UMD's high performance computing resources for the high performance researcher

By: Junaid Merchant, Phillip Alvarez, Dushyanthi Karunathilake, and Daniel Callow

NSF# 1632976

COMBINE
Computation and Mathematics for Biological Networks

Who are we?

Junaid Merchant

Neuroscience and Cognitive Science

Philip Alvarez

Biophysics

Daniel Callow

Neuroscience and Cognitive Science

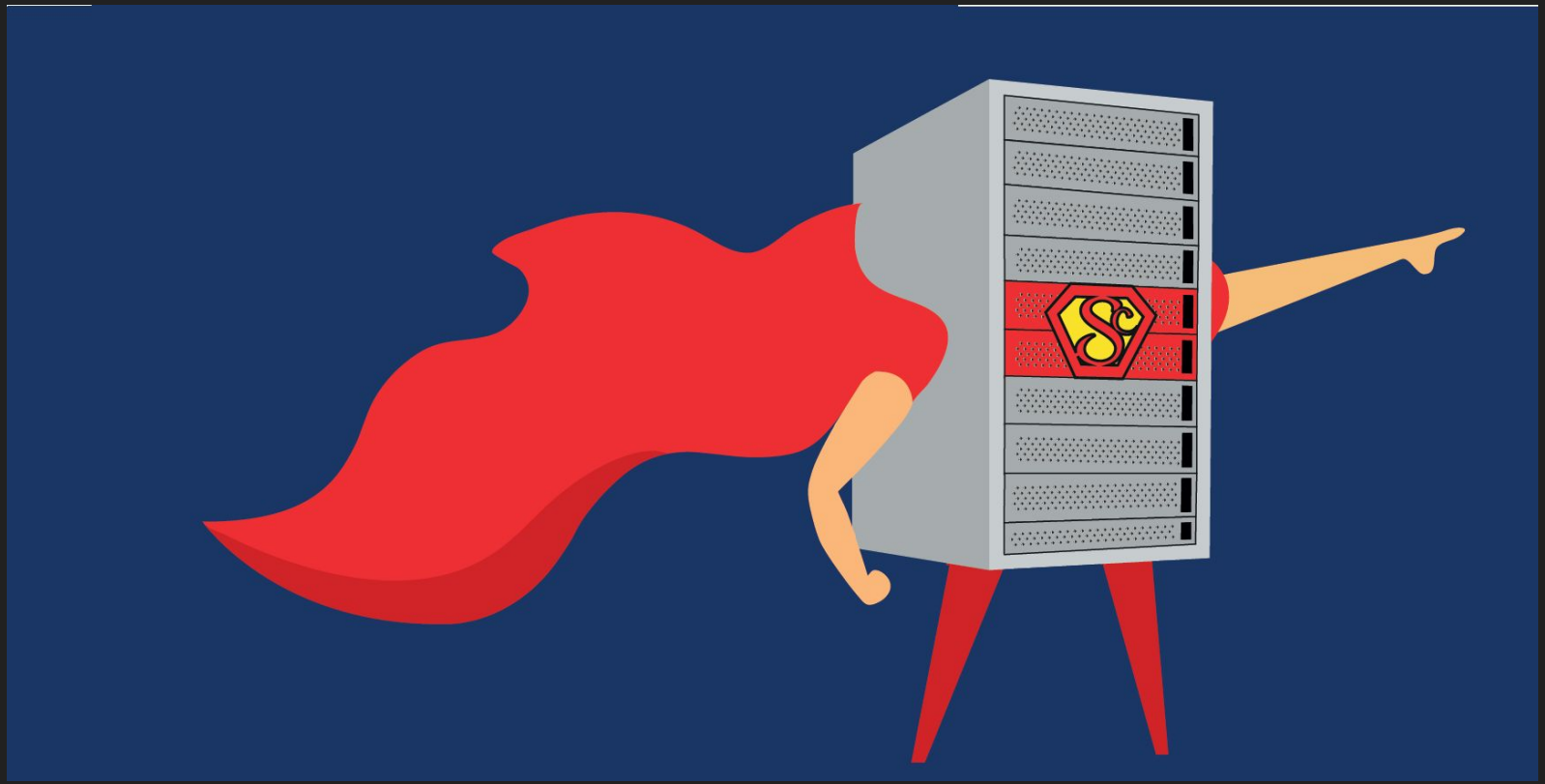
Dushyanthi Karunathilake

Electrical and Computer Engineering

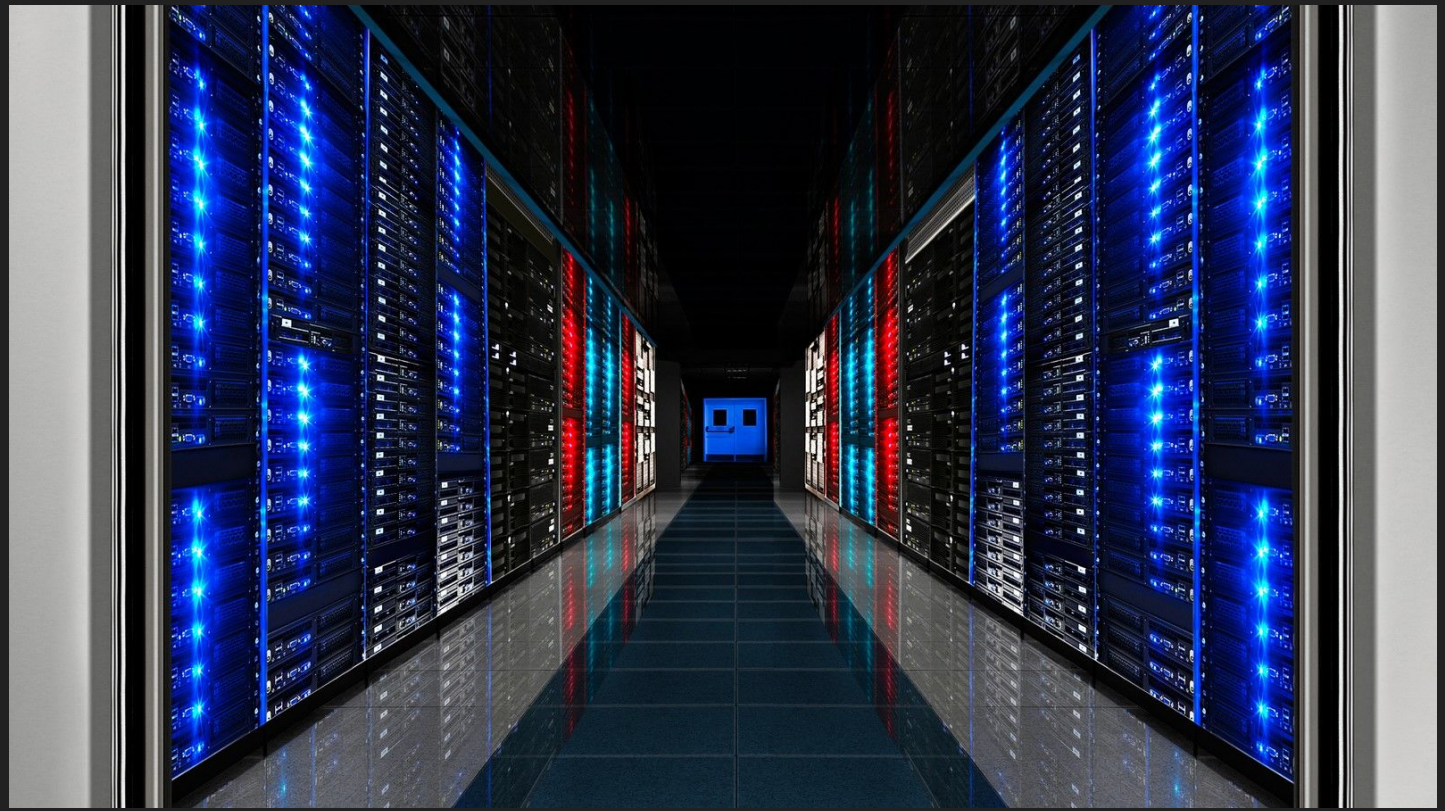
Outline

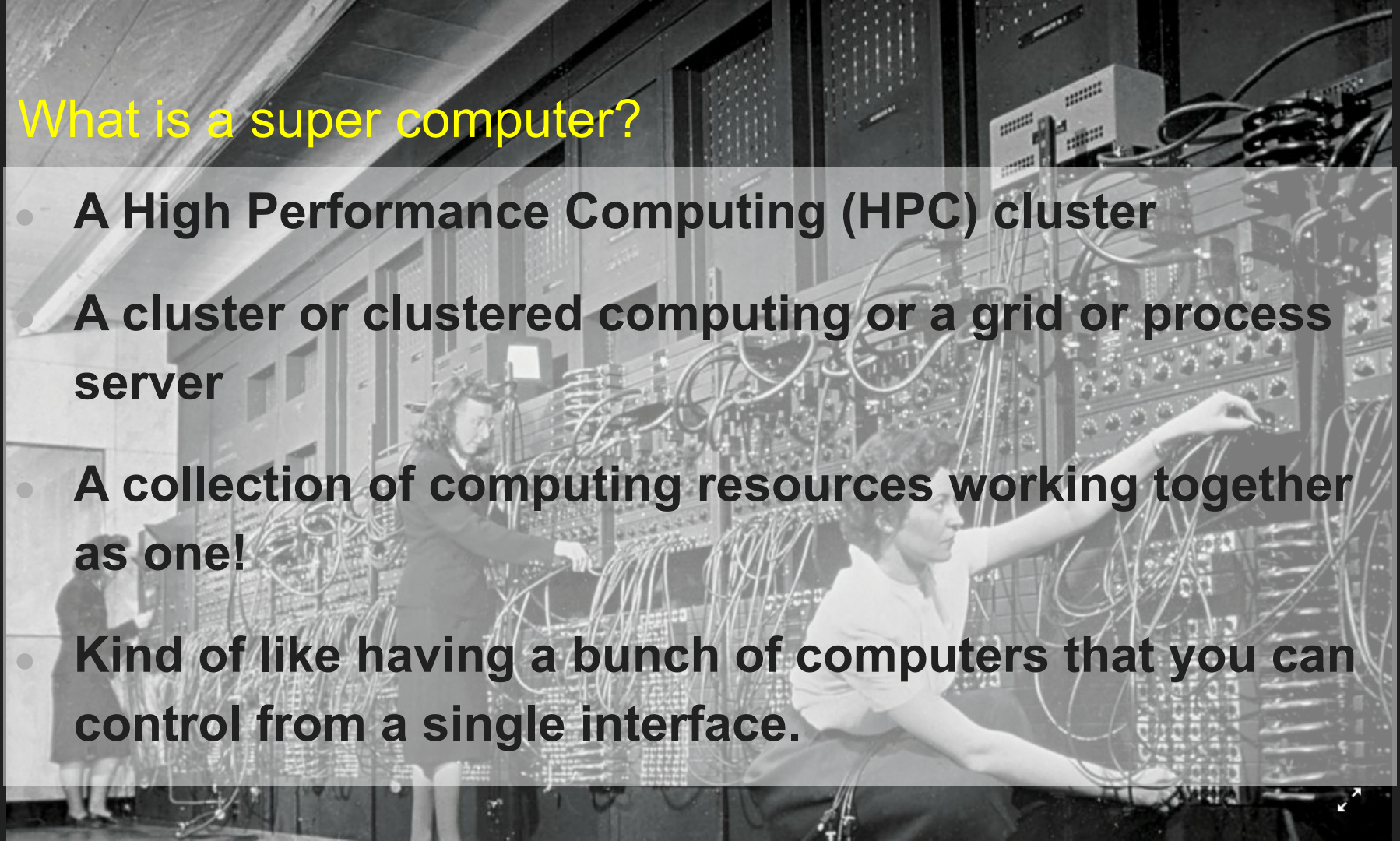
1. What is high performance computing and what is it useful for?
    a. Break out session
2. Overview of the conceptual basis of parallelizing code
    a. Examples
3. What parallel resources are available on campus and how can I get access?
    a. Break out session
4. Job submitting and data retrieval
    a. Discussion
5. Questions

What is a “supercomputer”?



More like...





What is a super computer?

- A High Performance Computing (HPC) cluster
- A cluster or clustered computing or a grid or process server
- A collection of computing resources working together as one!
- Kind of like having a bunch of computers that you can control from a single interface.

Important Terminology

Node: a computing unit that is comprised of a CPU, RAM, and maybe GPU

- Like a single computer
- Different nodes might have different specs

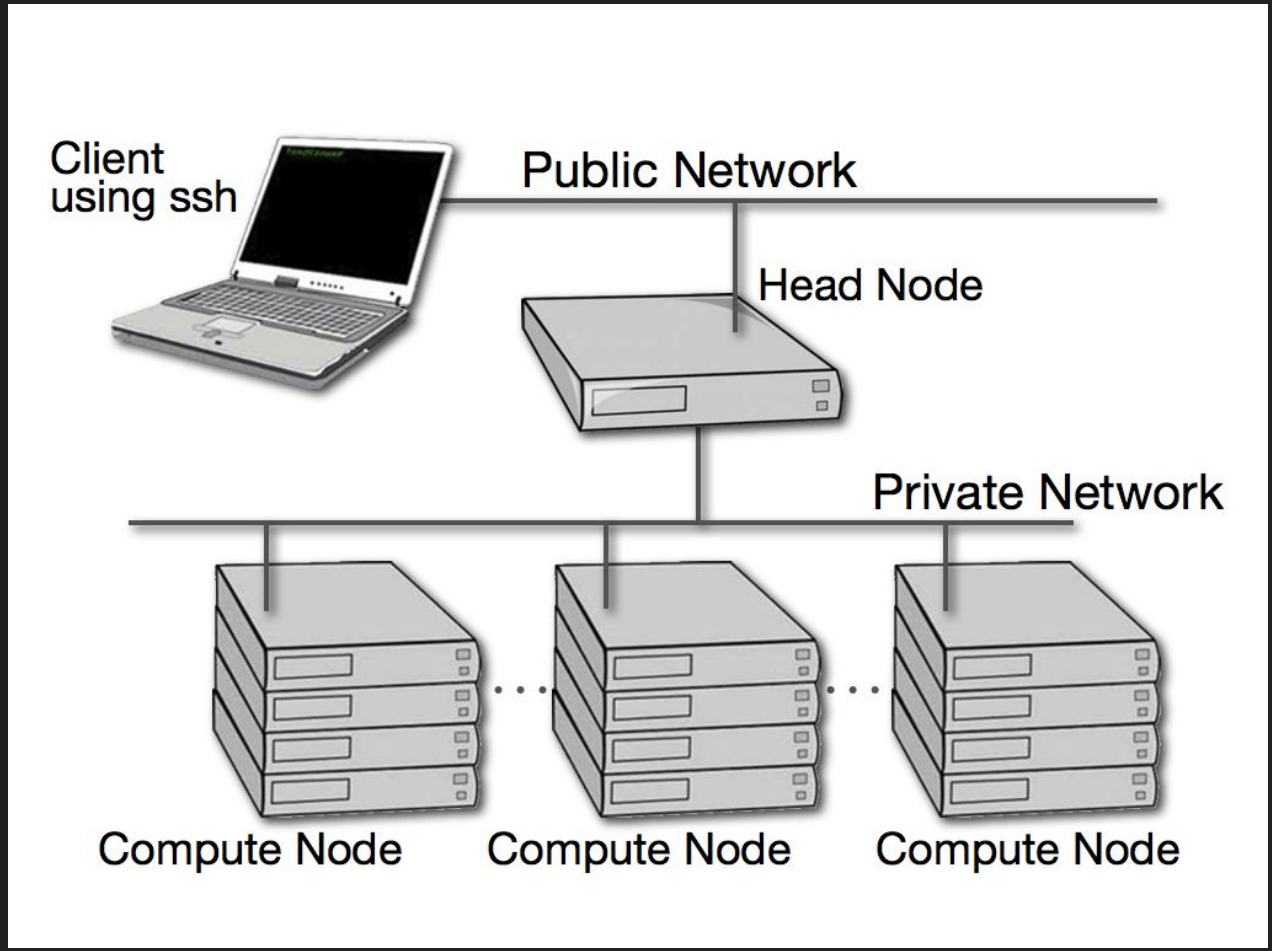
Head node: Where you (and everyone else) login & submit jobs from

- Master node that decides which node jobs are submitted to
- Shared resource, so DON’T run any jobs here

Job: A single process or script or program that you want to run on a node

- Can be comprised of subprocesses
- must run with no interaction (more on this later)

Important Terminology



The first supercomputer was built in 1964

1000x more processing power

iPhone 5s

Today's supercomputers

Often have tens if not hundreds of thousands of processors

Who uses supercomputers? And what for?

- Used for a wide range of computationally intensive tasks in various fields
  - Weather forecasting
  - Finance
  - Climate research
  - Oil and gas exploration
  - Molecular modeling
  - Physical simulations
    - To name a few….

Used extensively by many companies (a skill to learn)

Example - My Research

- Neuroimaging
  - 100+ subjects in a study
  - Brain images take up a lot of space
  - Processing steps for a single subject can take a day
    - Basic computer and code must process 1 at a time
      - Would take Months!
    - Supercomputer processes 30-60+ at a time
      - Done in Days

Why might you use a supercomputer?

- Processing steps with a lot of independent tasks (think lots of subjects)
- Computationally or very time intensive tasks

How about in your own research?

- Placed into break out sessions for 5 minutes
  - Discuss with your group how you might be able to employ high performance computing in your own research
    - Do you have any computationally or time intensive tasks?
    - What are the units in your processing pipeline that could be submitted?

- Designate one person from each group to share an example from their break out session to the entire group when we come back