Submitting Batch Jobs to the SCC

1/22/25

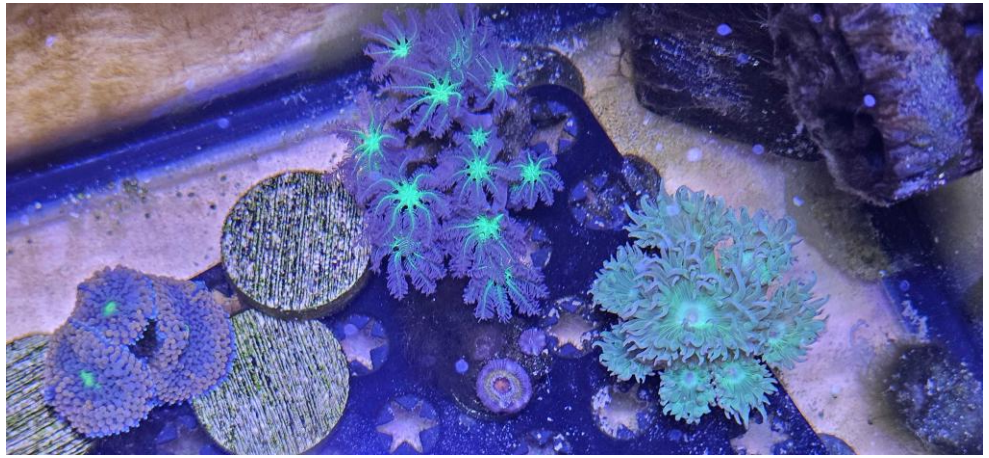
Dennis Milechin,
Research Computing Services

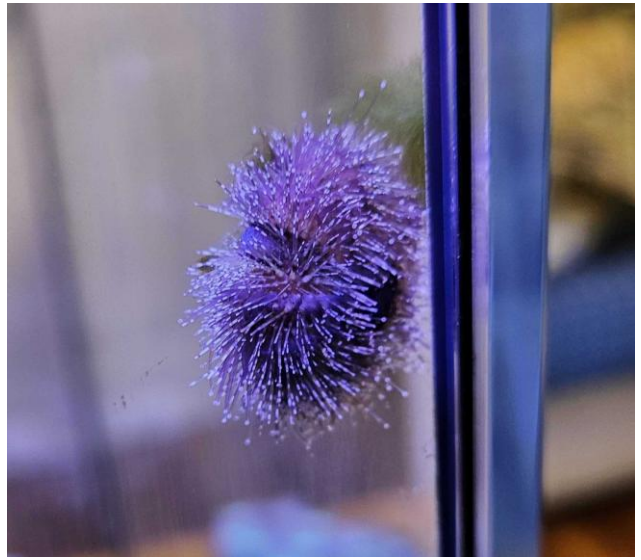
About Me

About Me

- B.S. and M.S. in Environmental Engineering
- 7 years in Environmental Consulting
- 2018 - joined Research Computing Services
- Self-taught computer programmer.

My Hobby





Getting Help

- Research Computing Support (RCS)
- http://rcs.bu.edu

- My Office - CAS 334G
- milechin@bu.edu
- help@scc.bu.edu

- Things to include in your request
  1. Job Number
  2. Screenshot of the error message
  3. Path on the SCC to the script you are running

Outline

- Session 1 (Today)
  - Overview of SCC
  - SCC Terminology and Architecture
  - Interactive Versus Batch Jobs
  - Create a qsub script and submit a batch job
  - Review SGE Environment Variables and Directives
- Session 2 (Next week)
  - Answer questions
  - Selecting resources for your job
  - qsub scripts for specific tasks (e.g. downloading files, array jobs)

Shared Compute Cluster (SCC)

The Shared Compute Cluster (SCC)

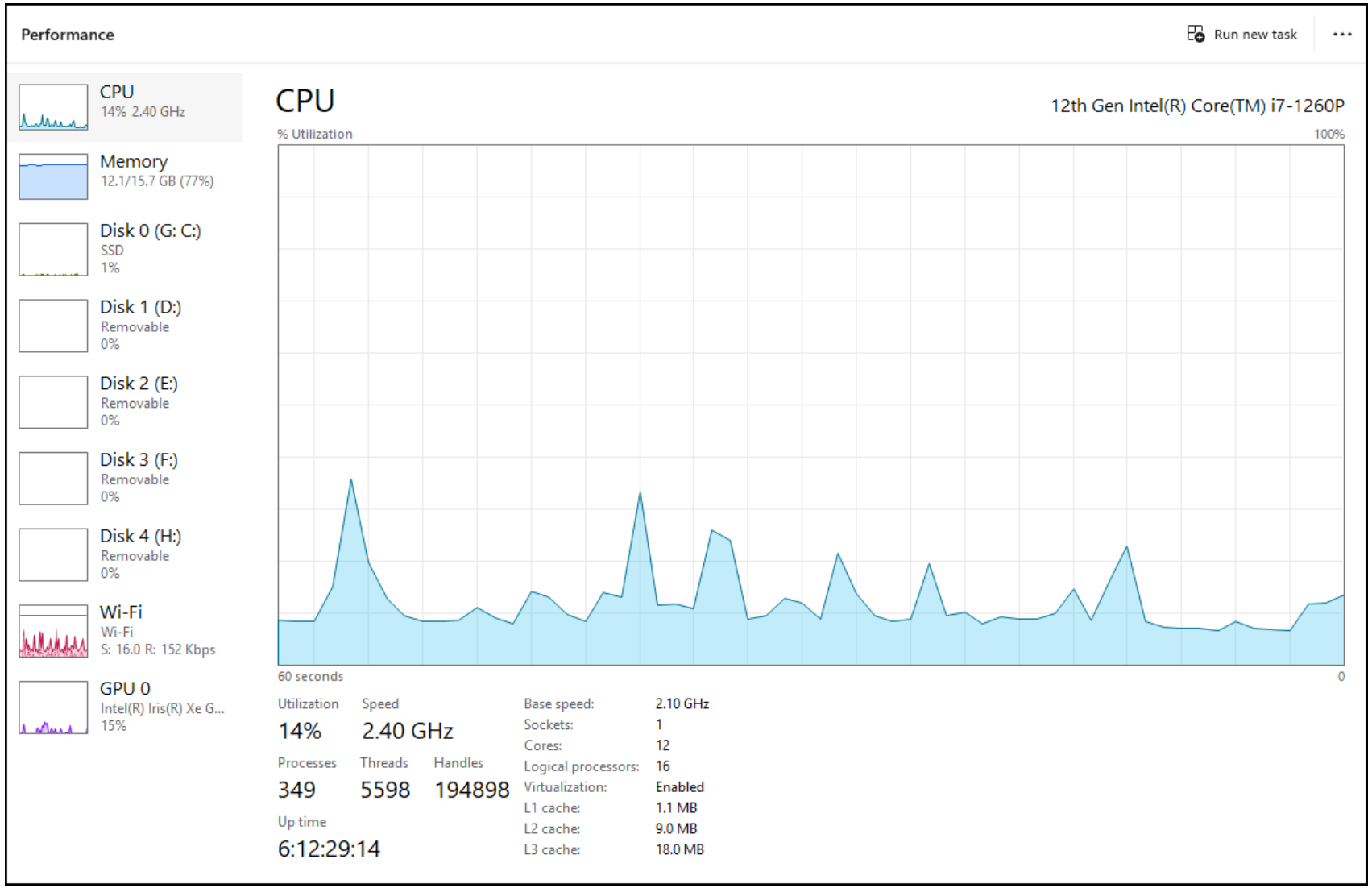
- Owned and Maintained by Boston University.
- Located in Massachusetts Green High Performance Computing Center (MGHPCC) in Holyoke, MA
- Went into production in June 2013
- Free to use by BU Research Community

For Info: http://www.bu.edu/tech/support/research/rcs/mghpcc/


MOHPCC

Why Use the SCC?

My Laptop Specs



SCC Technical Specs

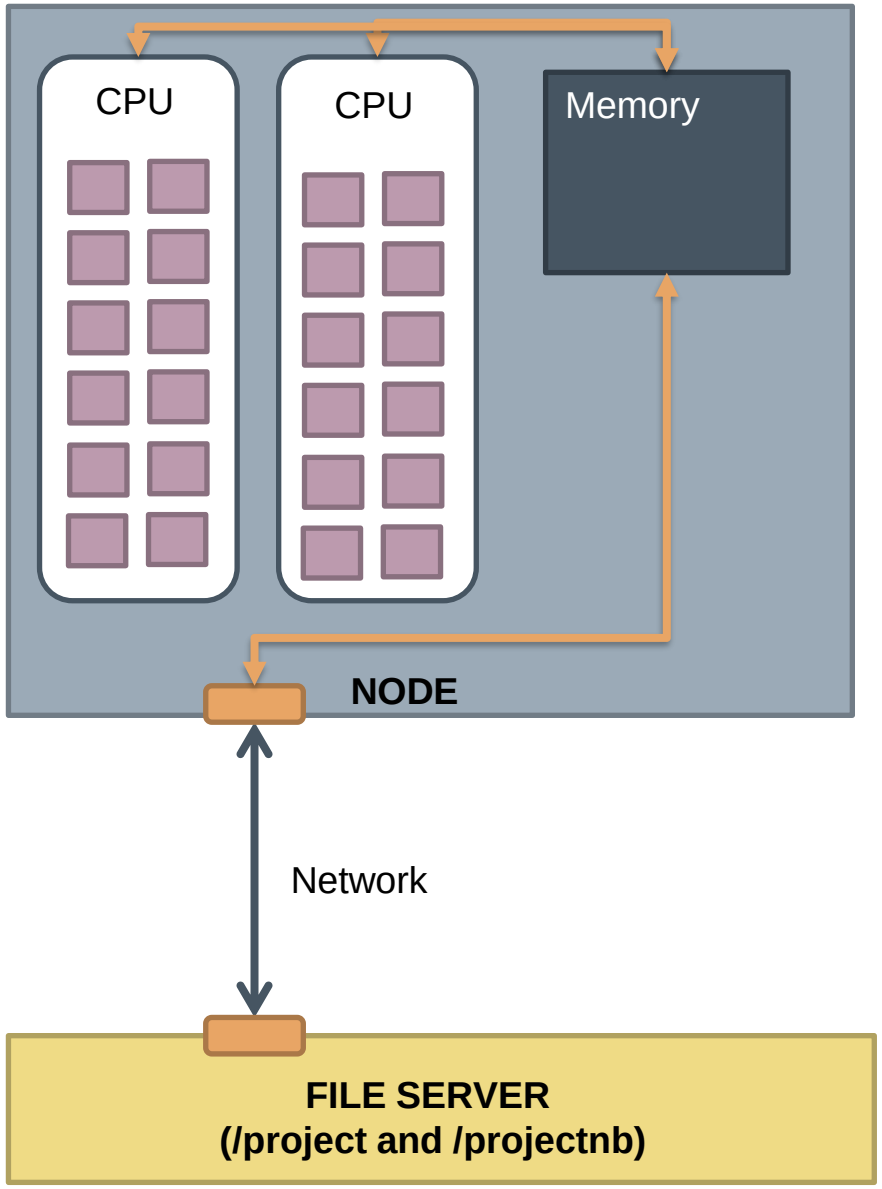
- 28,000 CPU cores (nodes with 16 up to 32 cores)
- 400 GPUs
- 14 Petabytes of data storage (An SCC project gets 1 TB for free)
- Memory 192 GB up to 1 TB

- https://www.bu.edu/tech/support/research/rcs/publications-and-grants/facilities/

SCC Terminology

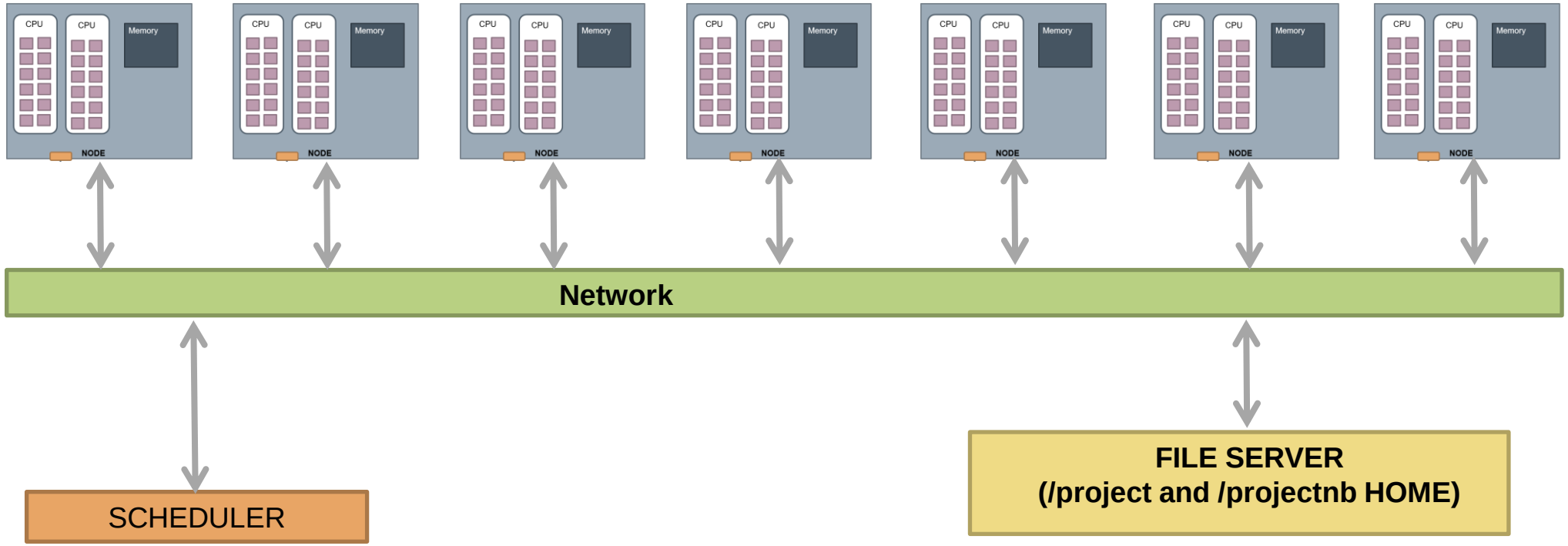
SCC Terminology

- NODE – One computer, may contain multiple CPUs
- CPU – Central Processing Unit, the brain of the computer
- Cores **(Slots)** – Workers available within the CPU.
- Memory – Also known as RAM. Feeds the data to the CPU and receives computed data back.
- **FILE SERVER** - Where all the project data is stored.



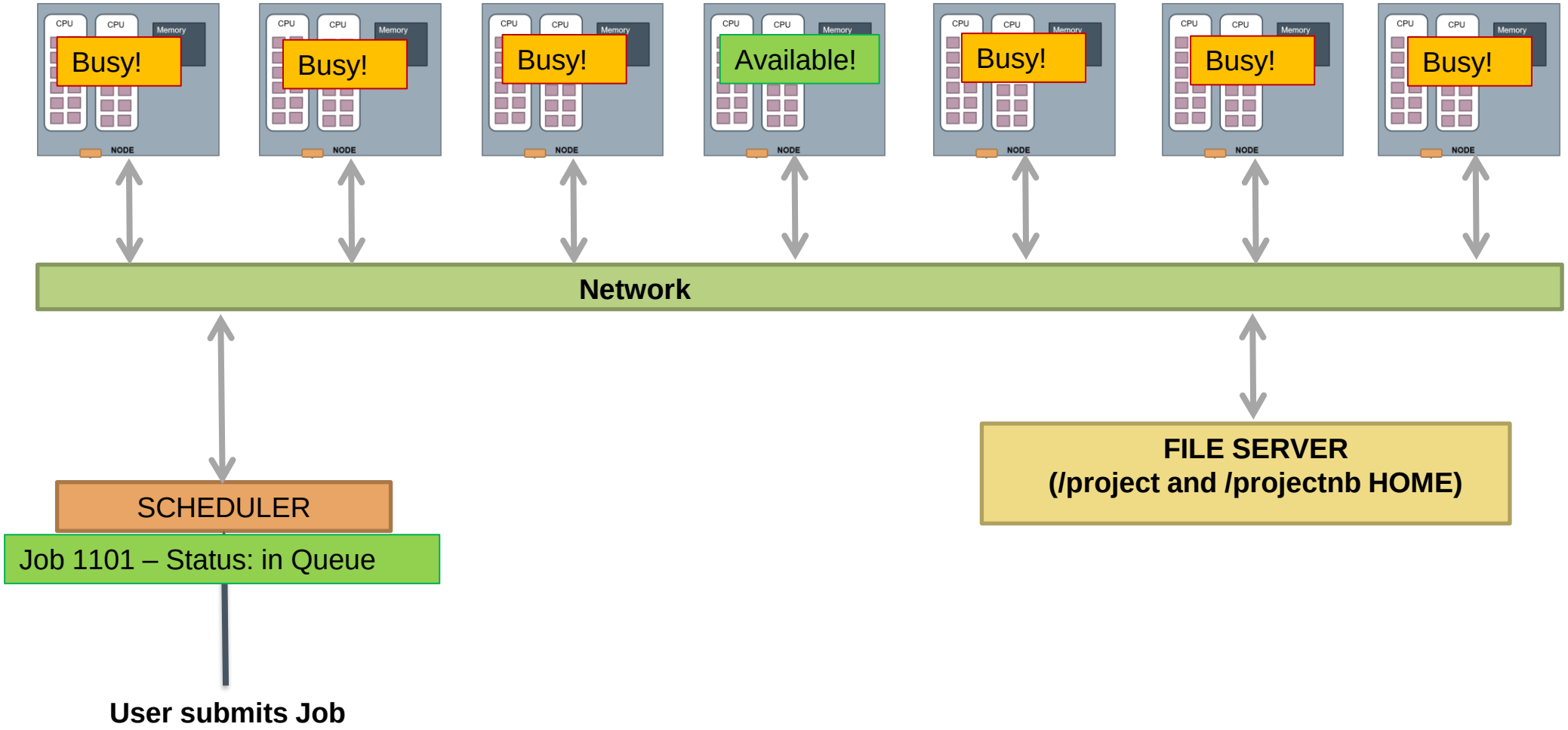
SCC Terminology

CLUSTER – Collection of computers (or compute nodes)



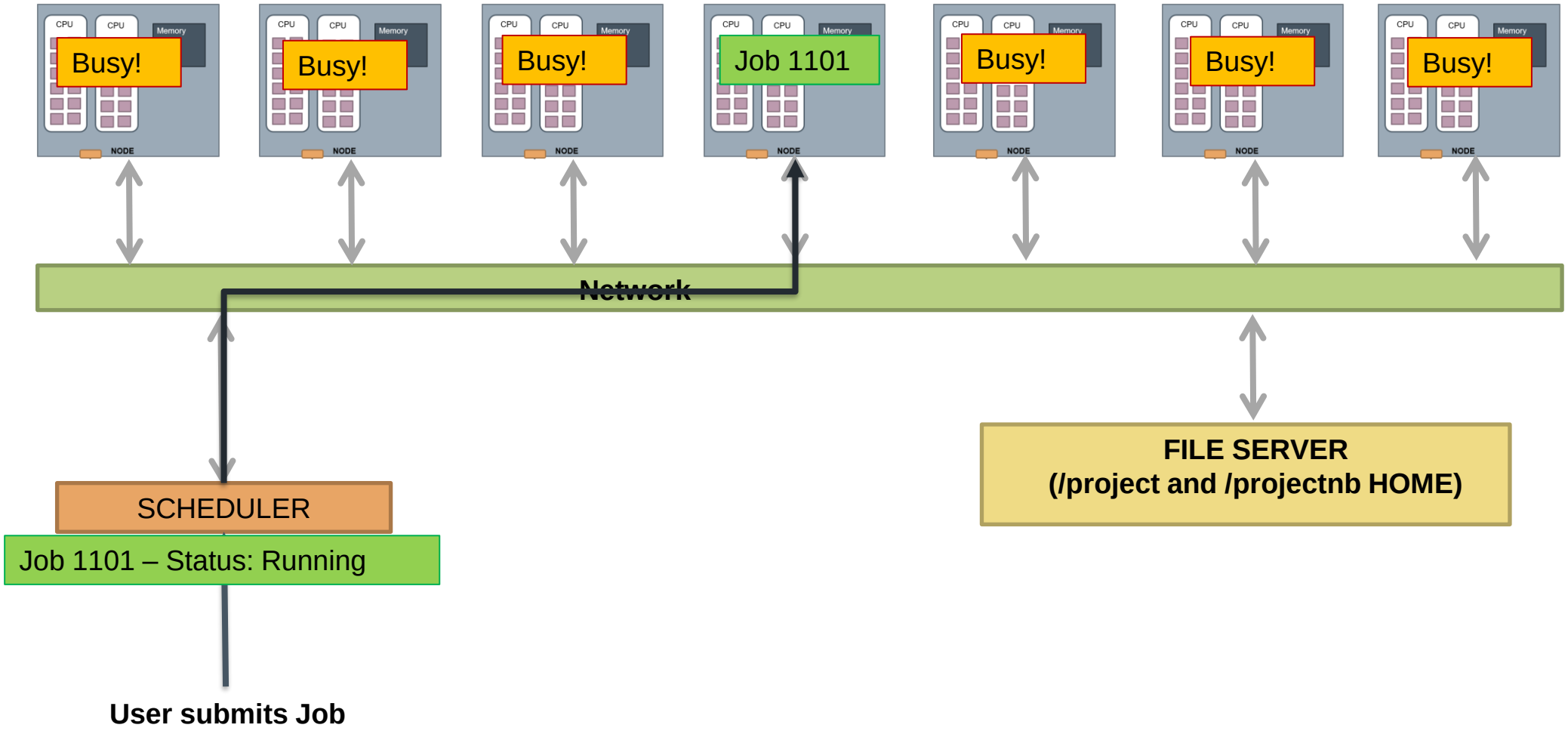
SCC Terminology

CLUSTER – Collection of computers (or compute nodes)



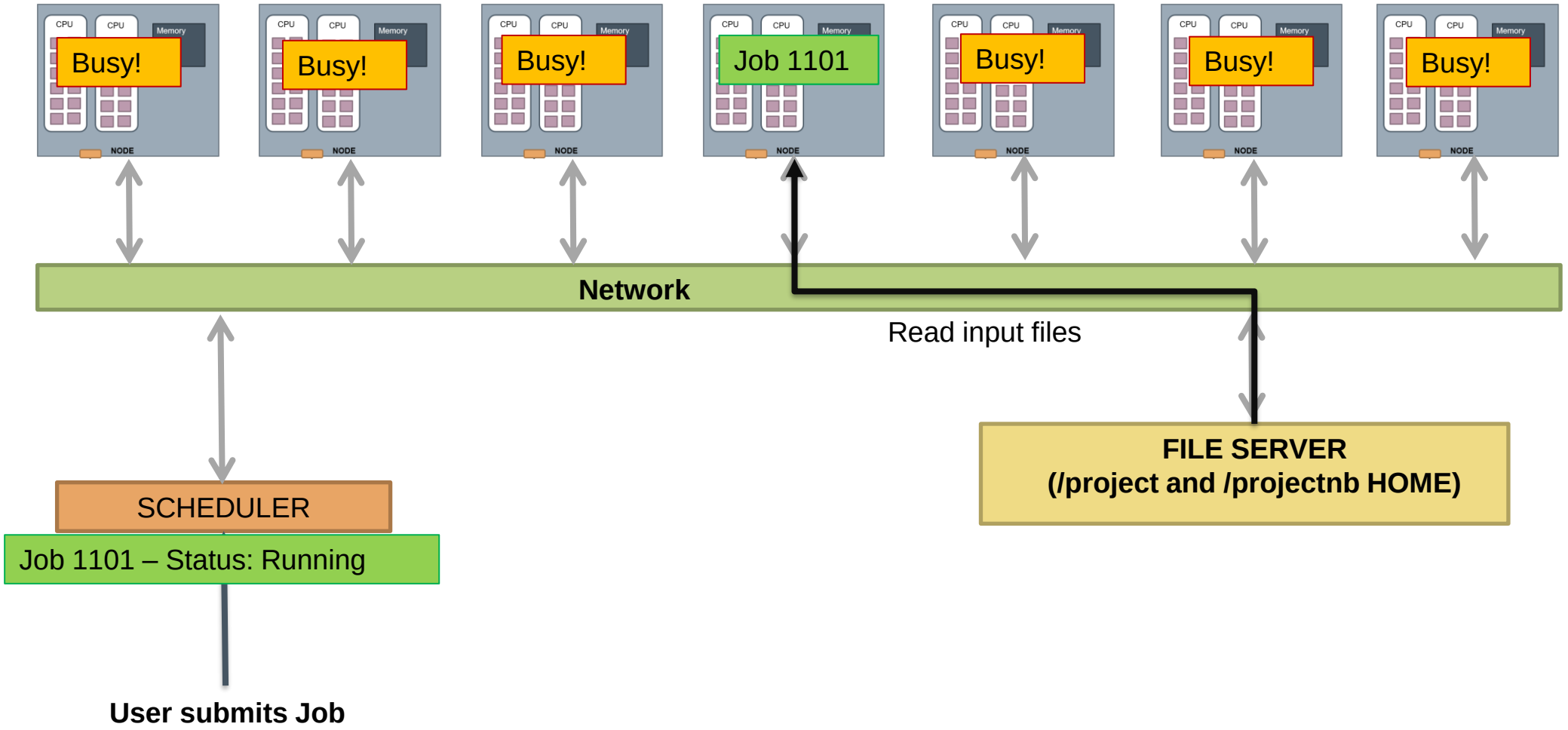
SCC Terminology

CLUSTER – Collection of computers (or compute nodes)



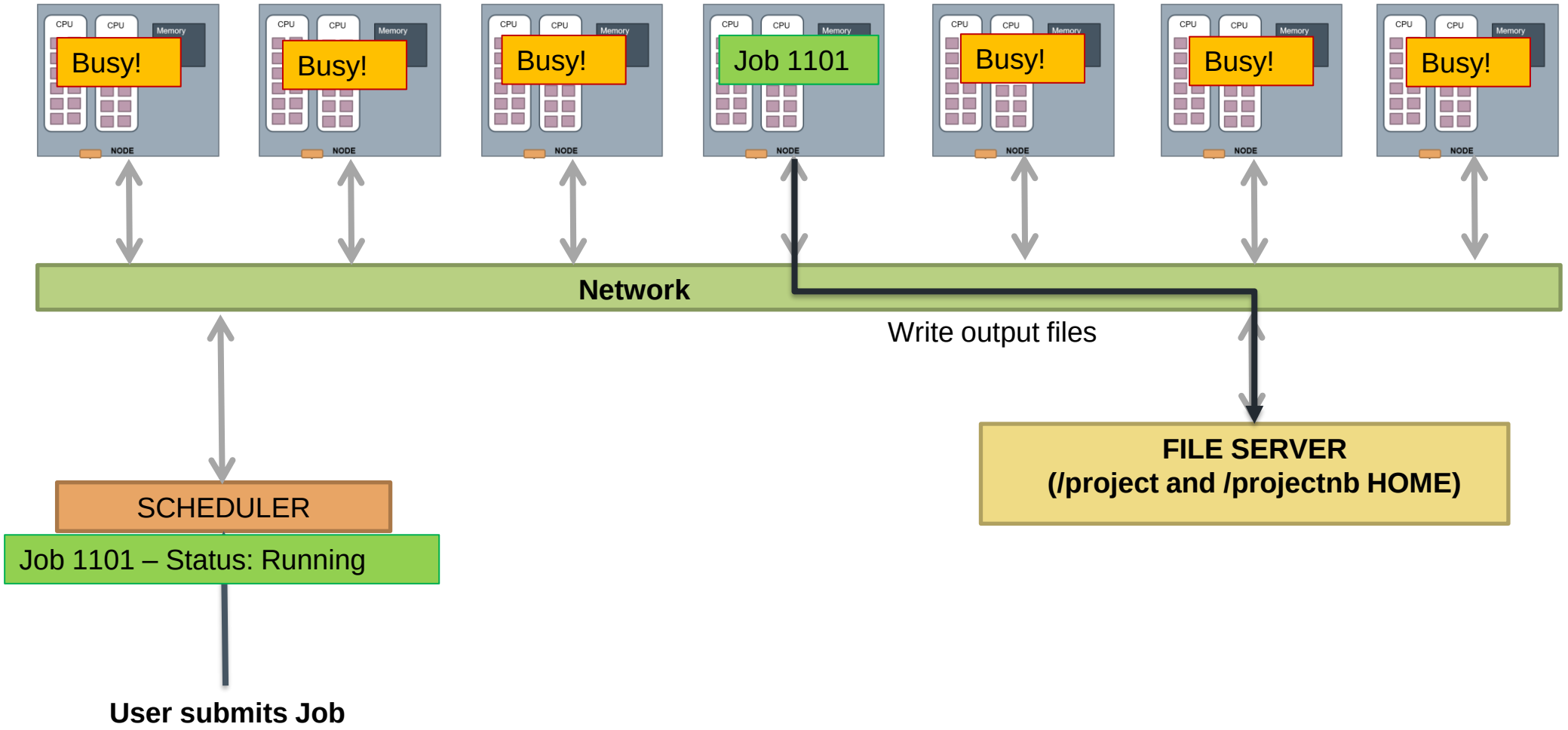
SCC Terminology

CLUSTER – Collection of computers (or compute nodes)



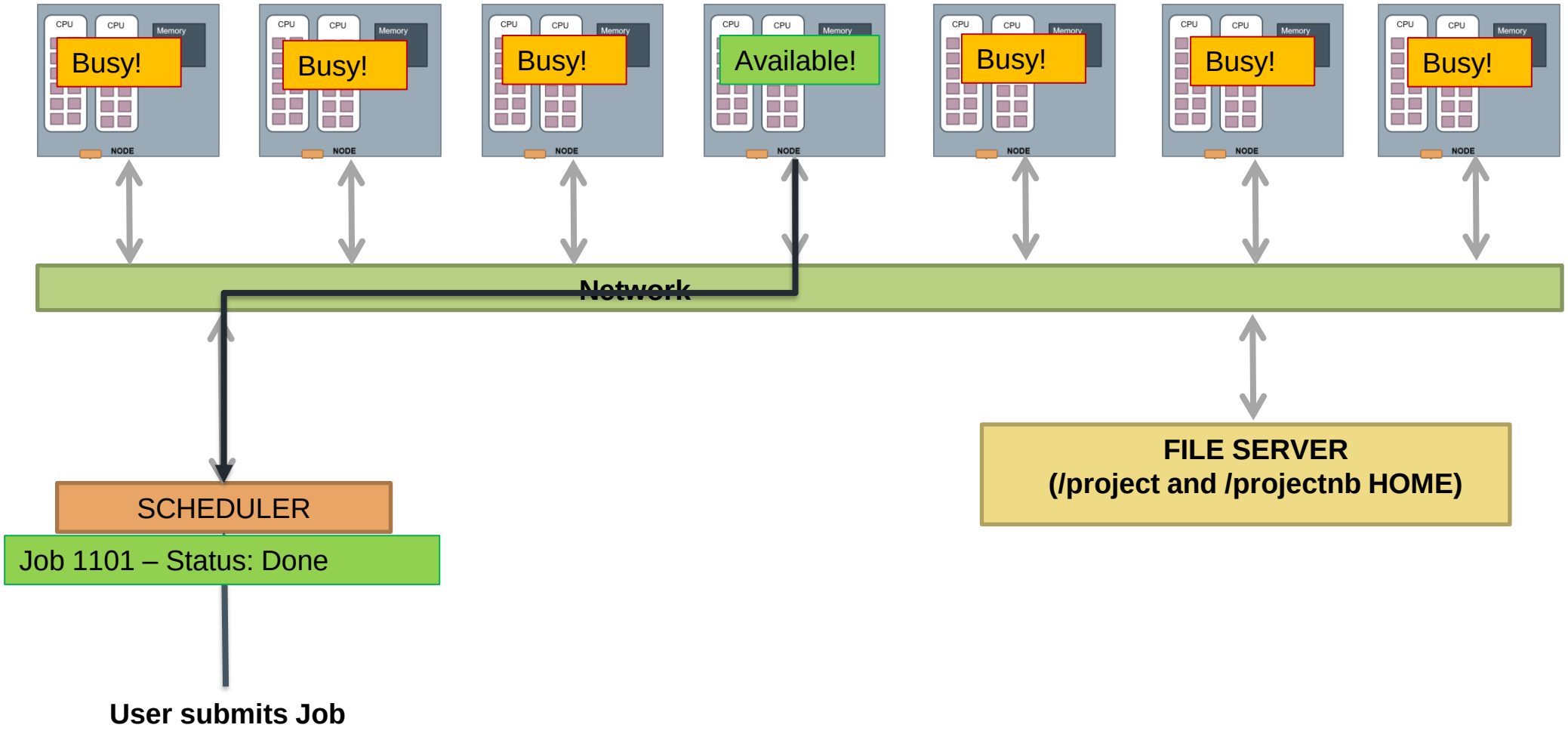
SCC Terminology

CLUSTER – Collection of computers (or compute nodes)



SCC Terminology

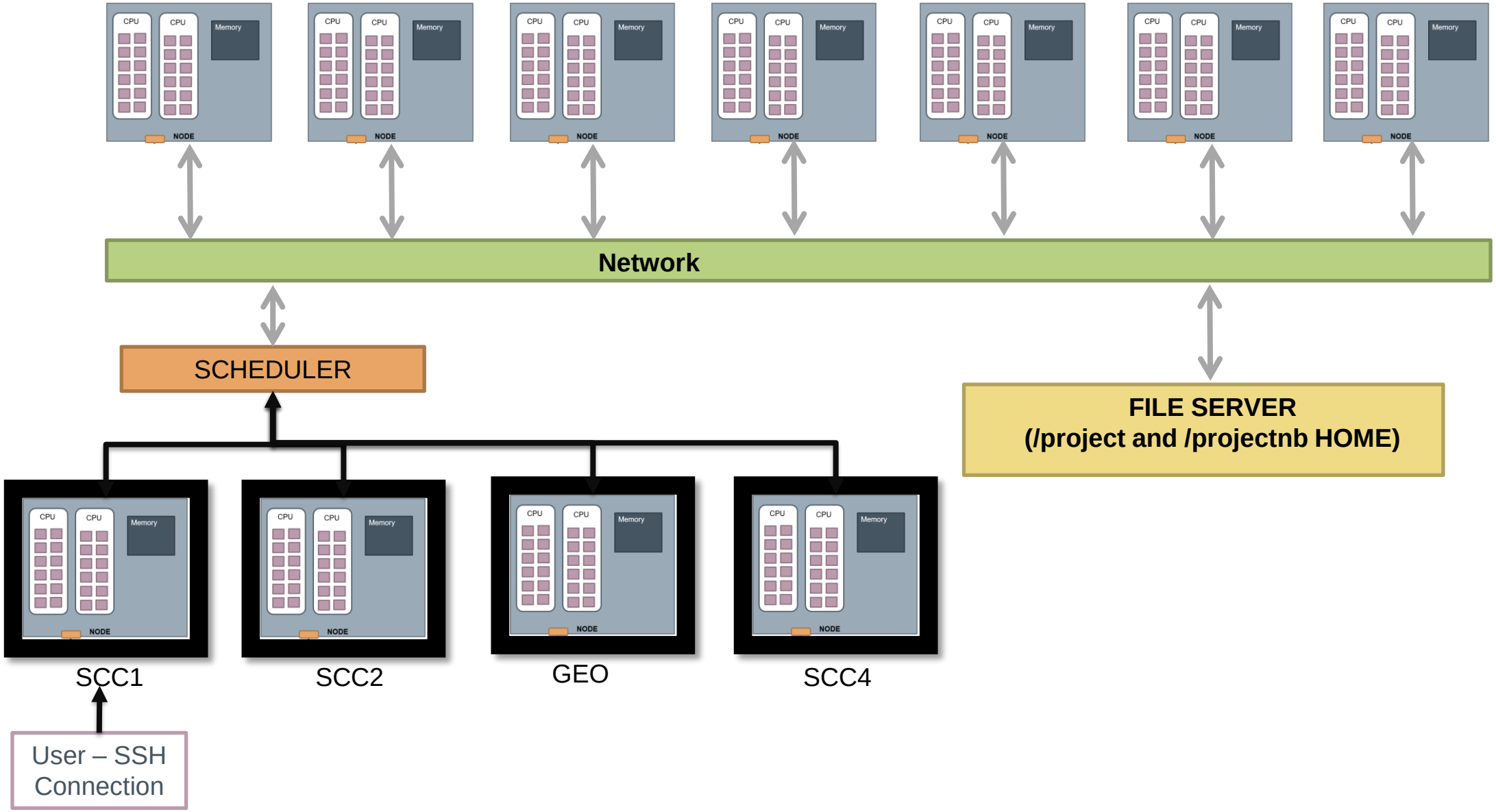
CLUSTER – Collection of computers (or compute nodes)



How do login nodes (SCC1, SCC2, GEO, and SCC4) fit in this?

Connecting to the SCC

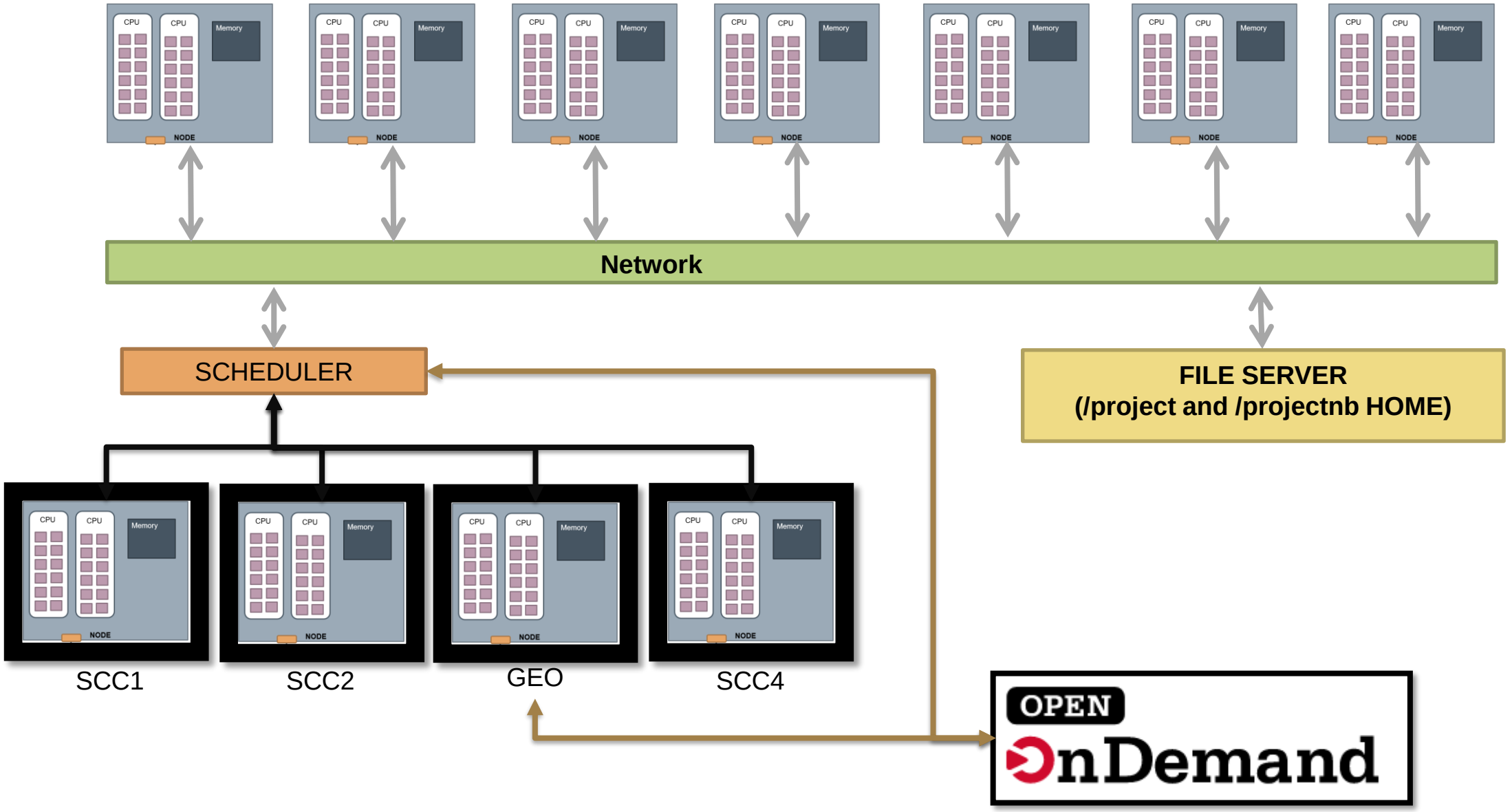
CLUSTER – Collection of computers (or compute nodes)



How about OnDemand?

Connecting to the SCC

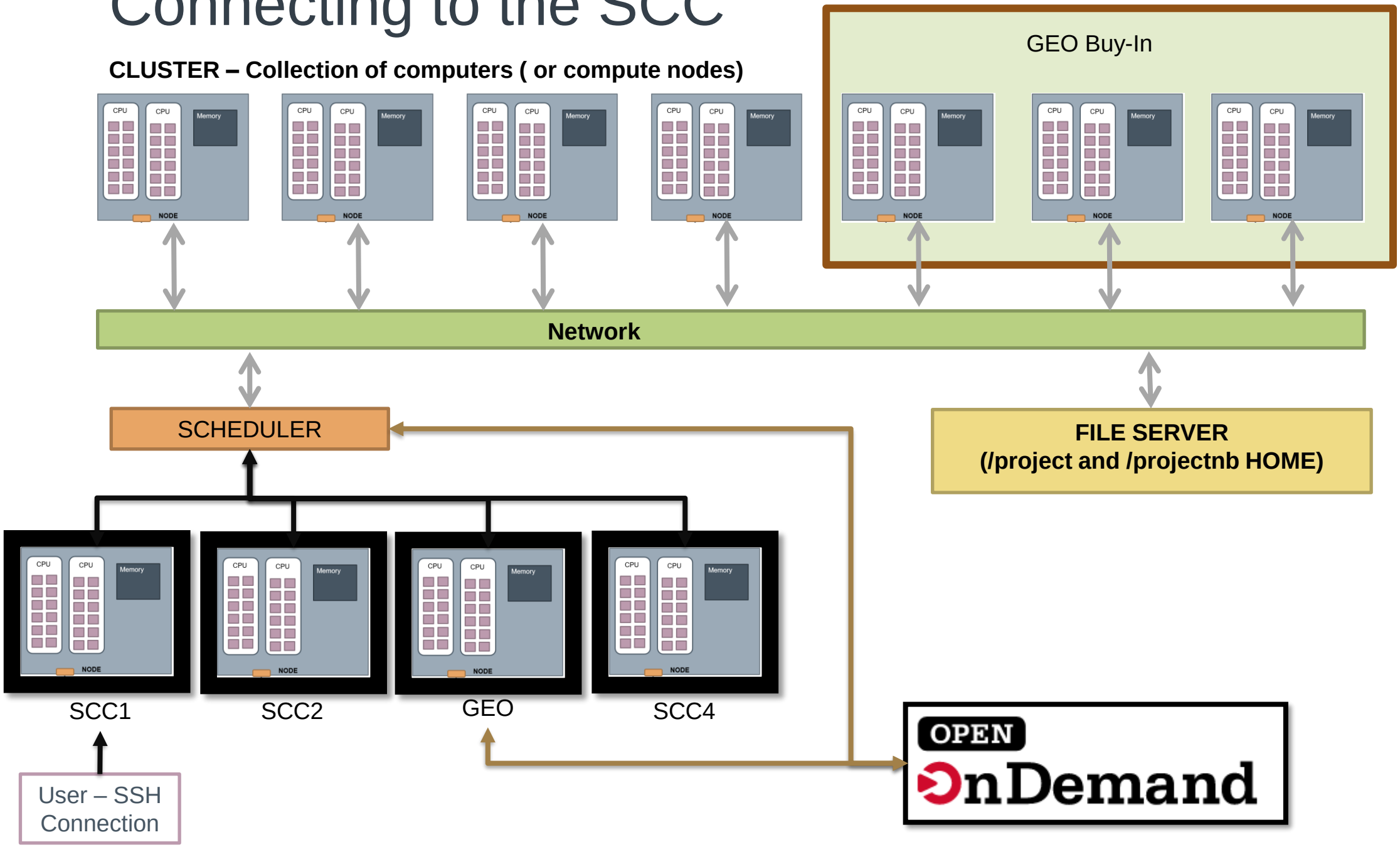
CLUSTER – Collection of computers (or compute nodes)



What is the GEO queue?

Connecting to the SCC

CLUSTER – Collection of computers (or compute nodes)



Interactive Versus Batch Jobs

- Batch Job
  - Write script that needs no input from user
  - Submit script as a job to the SCC
  - Scheduler finds node and executes the script
  - Wait for job to complete
- Interactive Job
  - Computation task needs input from user
  - Submit request for interactive job
  - Get direct connection to a compute node

Questions?

Let’s create a batch submission script.