HPC Workflow on Shaheen

(Chemistry, Physics & Materials Science)

Zhiyong Zhu

KAUST Supercomputing Core Lab

Outline

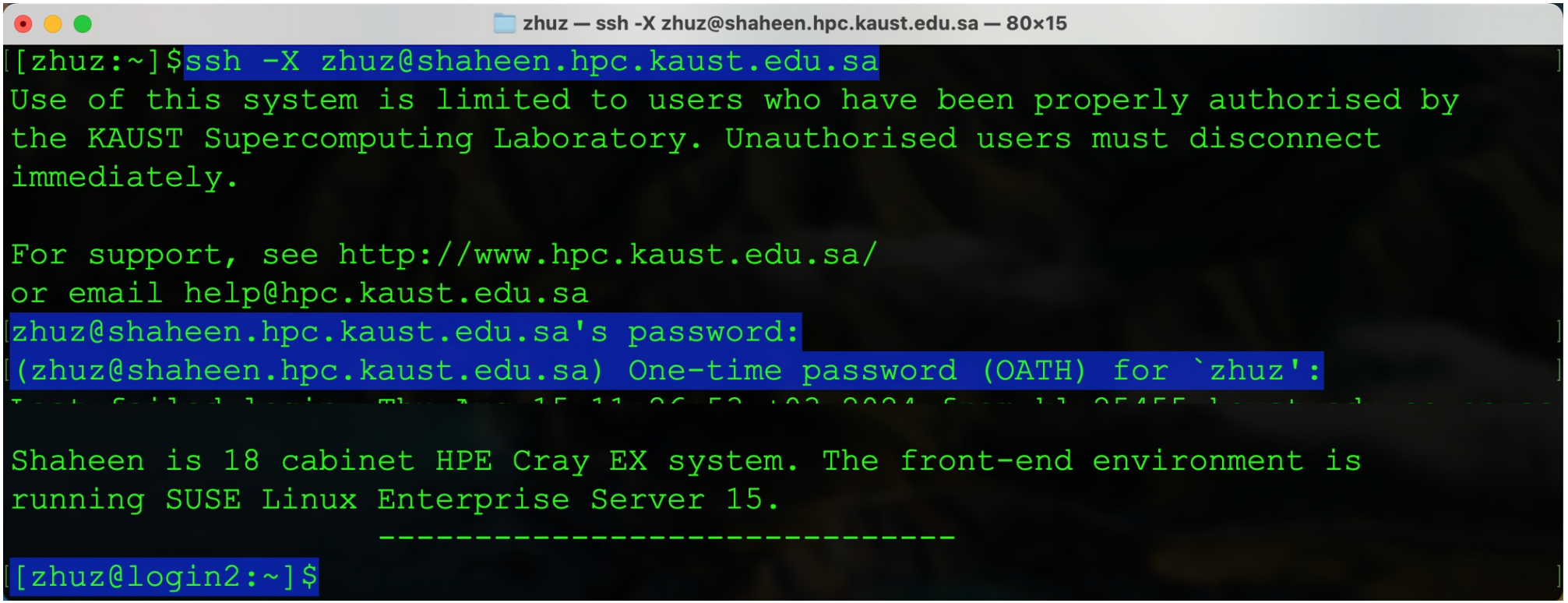
- Workflow
  - Use VASP as an example to show the steps to run
    - The Vienna Ab-initial Simulation Package (VASP) is a computer program for atomic scale materials modeling, e.g. electronic structure calculations and quantum-mechanical molecular dynamics from first-principles (https://www.vasp.at).

Workflow

- Login Shaheen
- Check Code Availability
- Working Directory
- Prepare Input Files for VASP
- Prepare Jobscript for Slurm Job Scheduler
- Job Submission using Slurm Commands
- Check Output Files

Login Shaheen

- Login
  - ssh -X <UserName>@shaheen.hpc.kaust.edu.sa

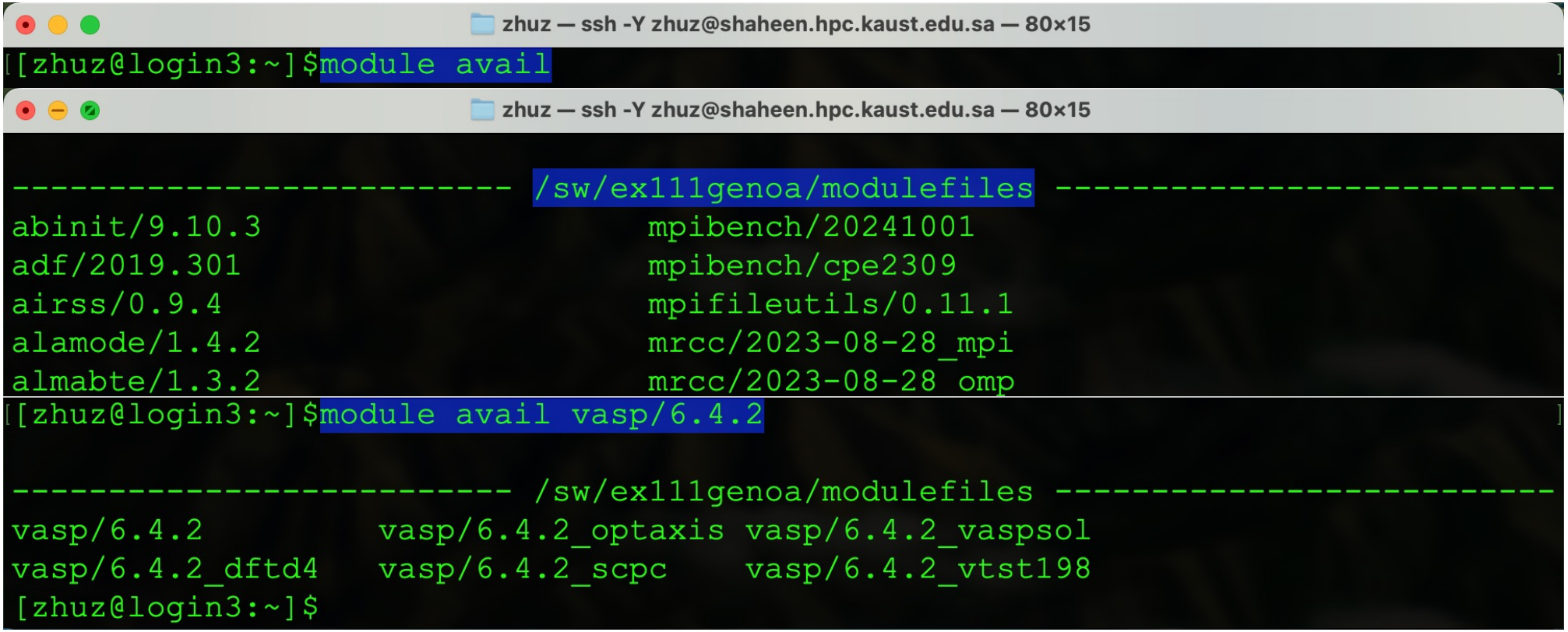


Code Availability

- On Shaheen login node:
  - module avail
- In /sw/ex111genoa software stack:
  - ls /sw/ex111genoa
- From our website:
  - https://docs.hpc.kaust.edu.sa

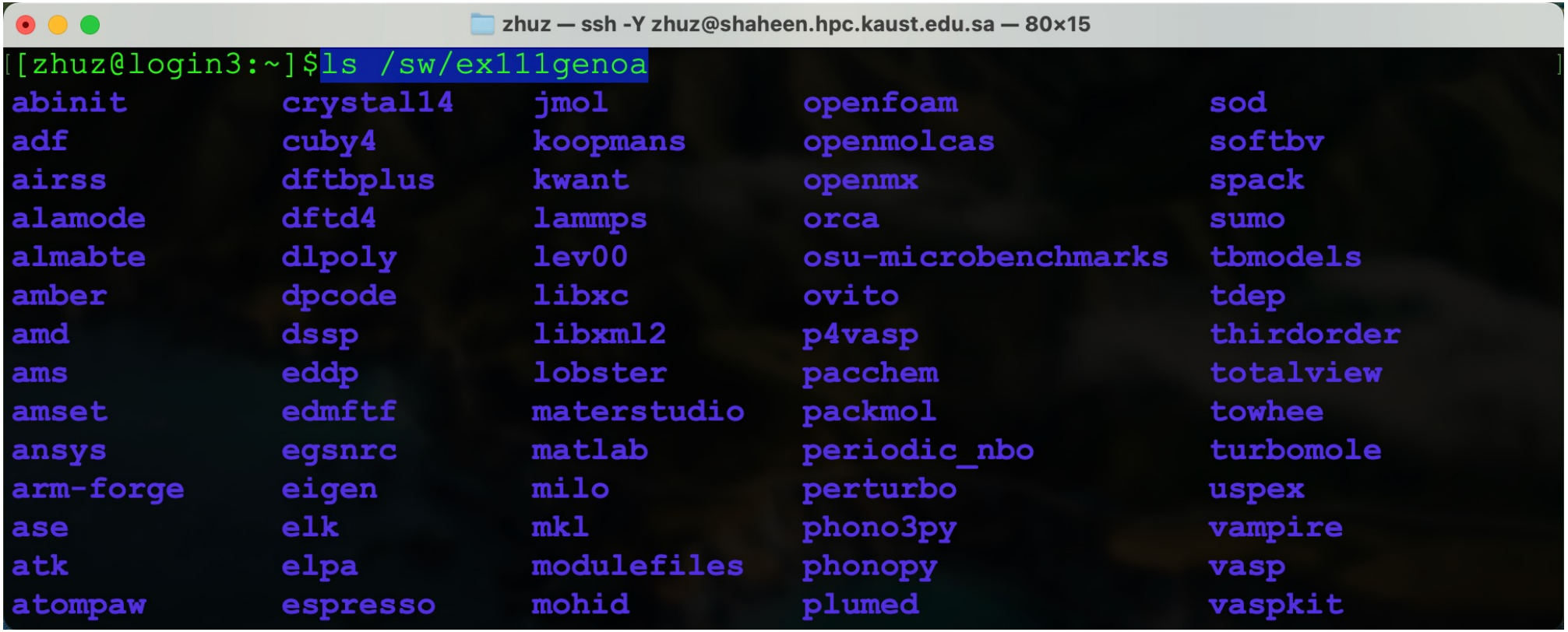
Code Availability

- module avail
  - module avail
  - module avail <code>/<version>



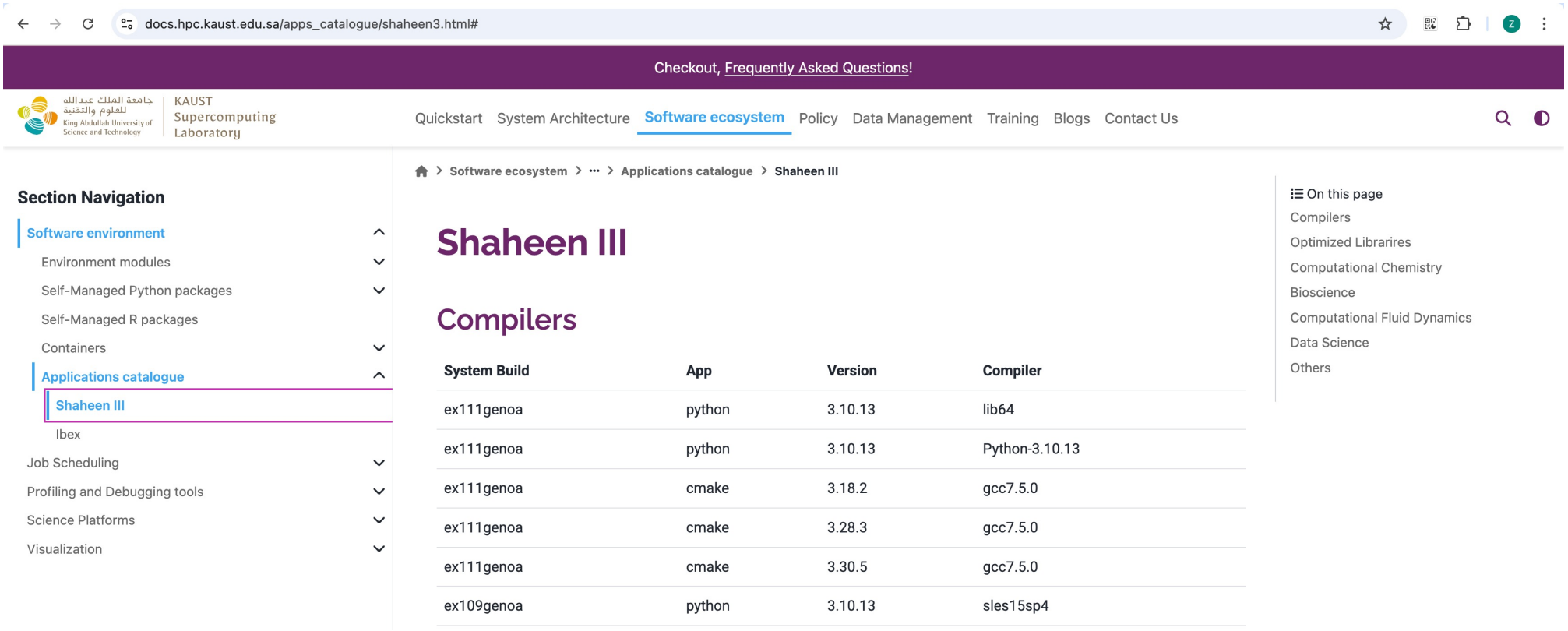
Code Availability

- ls /sw/ex111genoa



Code Availability

- https://docs.hpc.kaust.edu.sa
  - Software ecosystem -> Software environment -> Applications catalogue -> Shaheen III

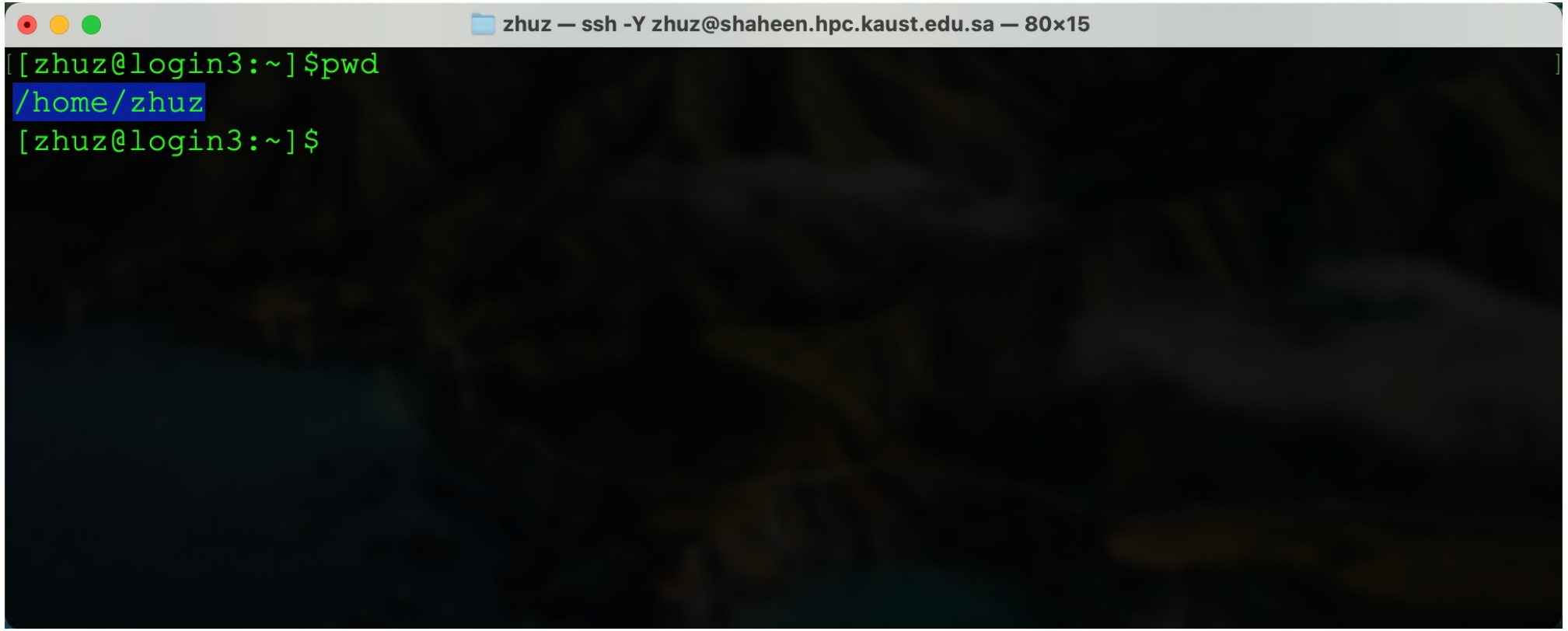


Code Availability

- On Shaheen login node:
  - module avail
- In /sw/ex111genoa software stack:
  - ls -l /sw/ex111genoa
- From our website:
  - https://docs.hpc.kaust.edu.sa
- Not found?
  - help@hpc.kaust.edu.sa

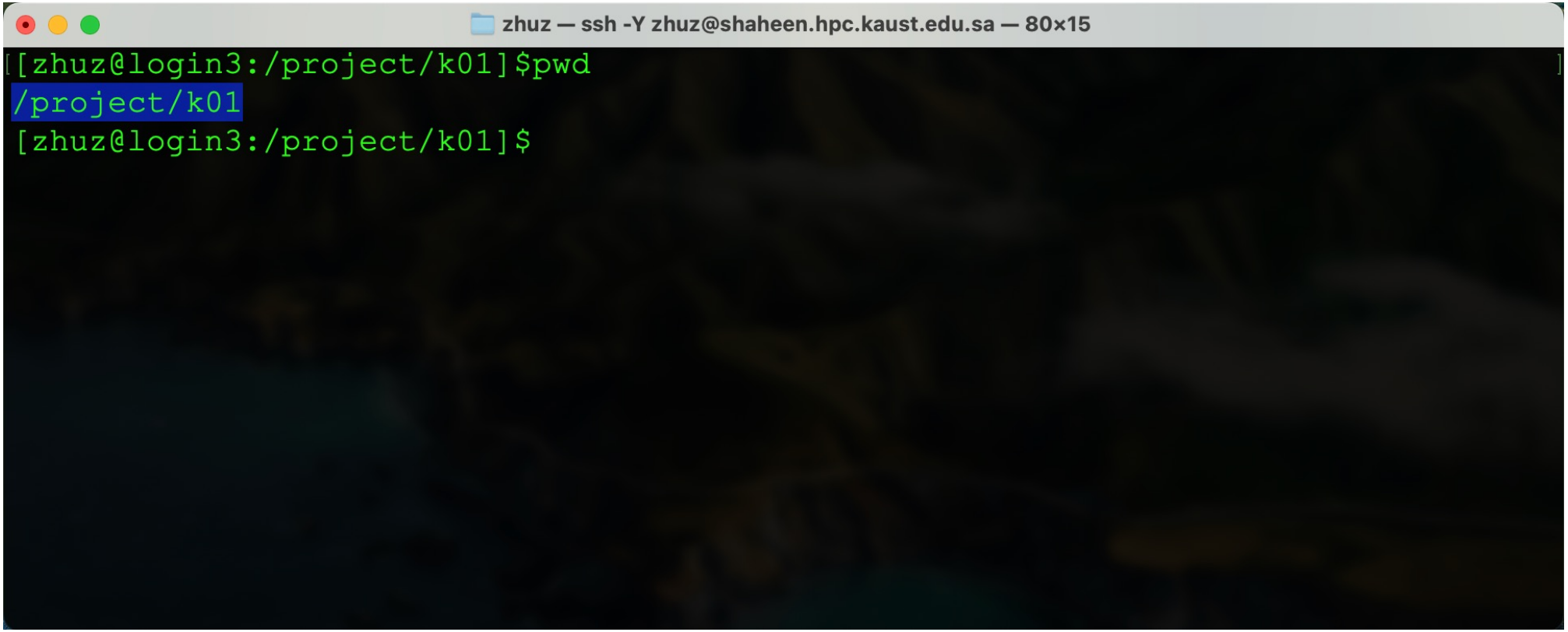
3 Different Working Directories

- /home
  - Very limited space; Not mounted on compute nodes (job submission will fail)



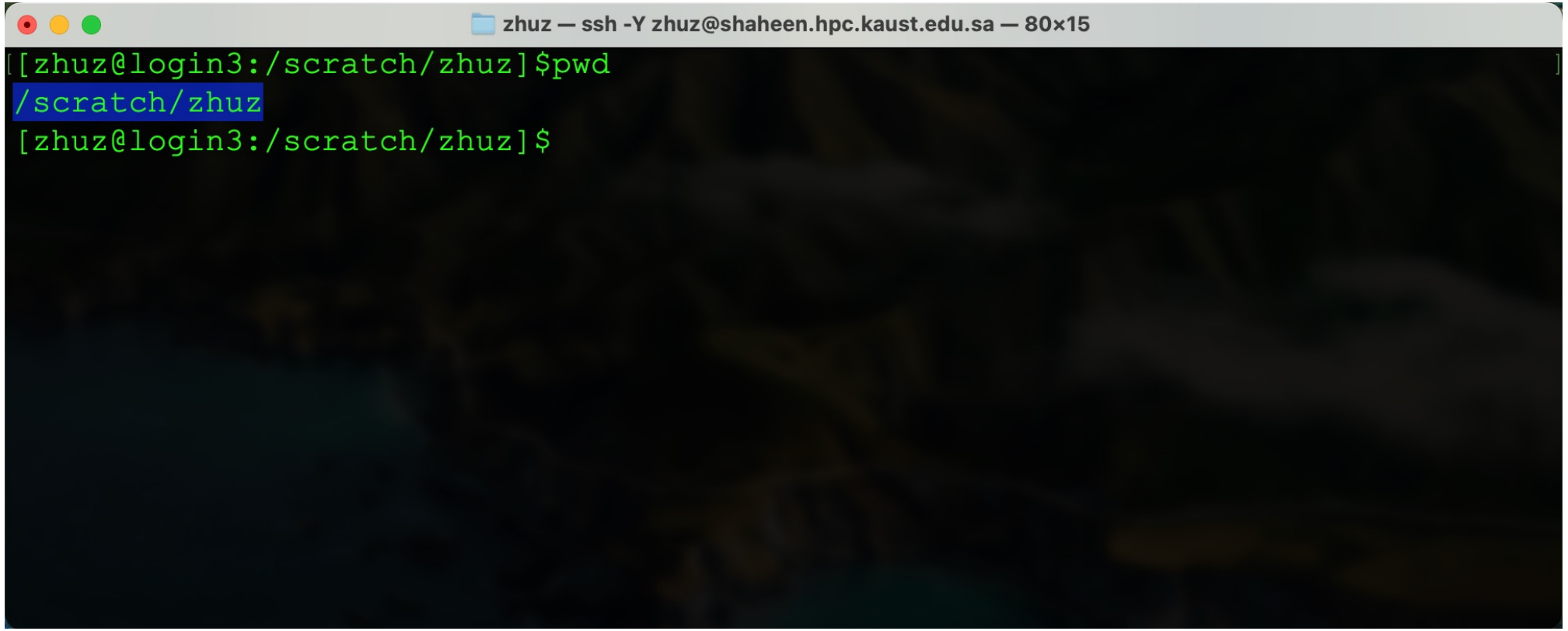
3 Different Working Directories

- /project/<projectname>
  - Read-only for compute nodes (job submission will fail); Used for data backup and data sharing



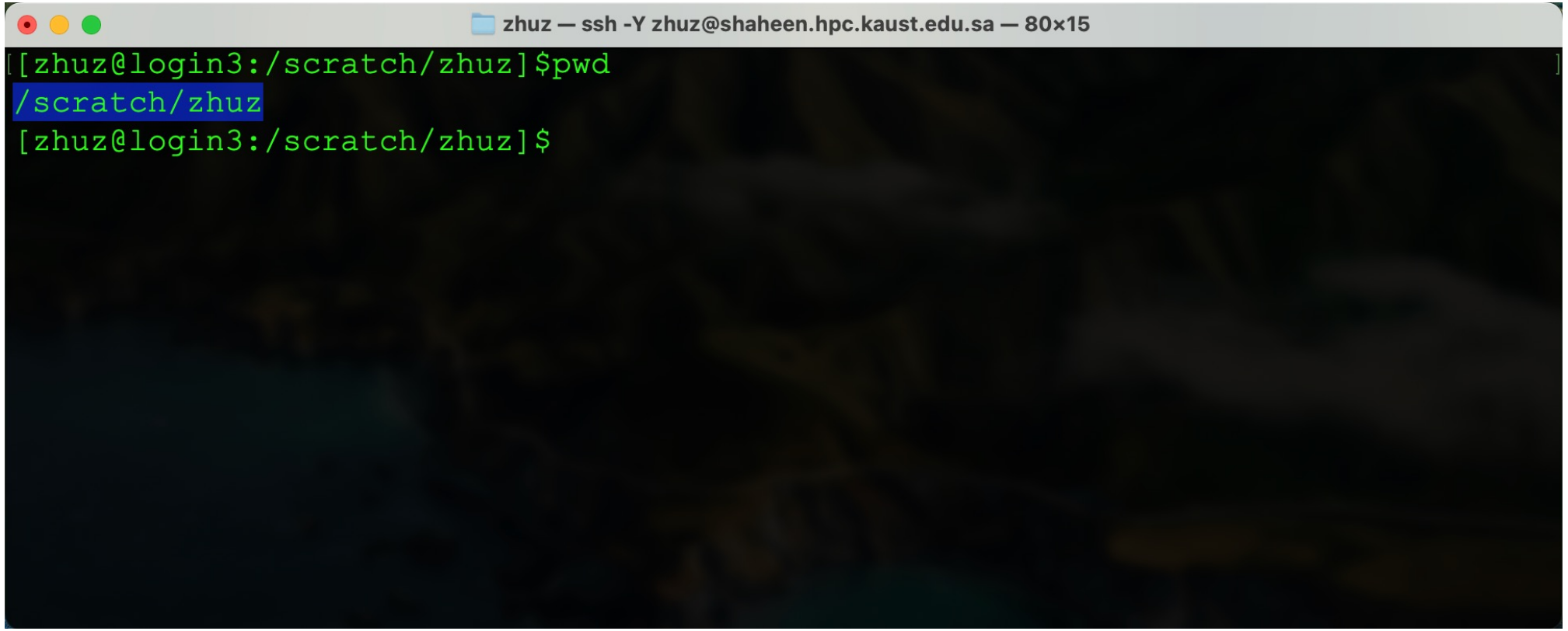
3 Different Working Directories

- /scratch/<username>
  - Almost unlimited space



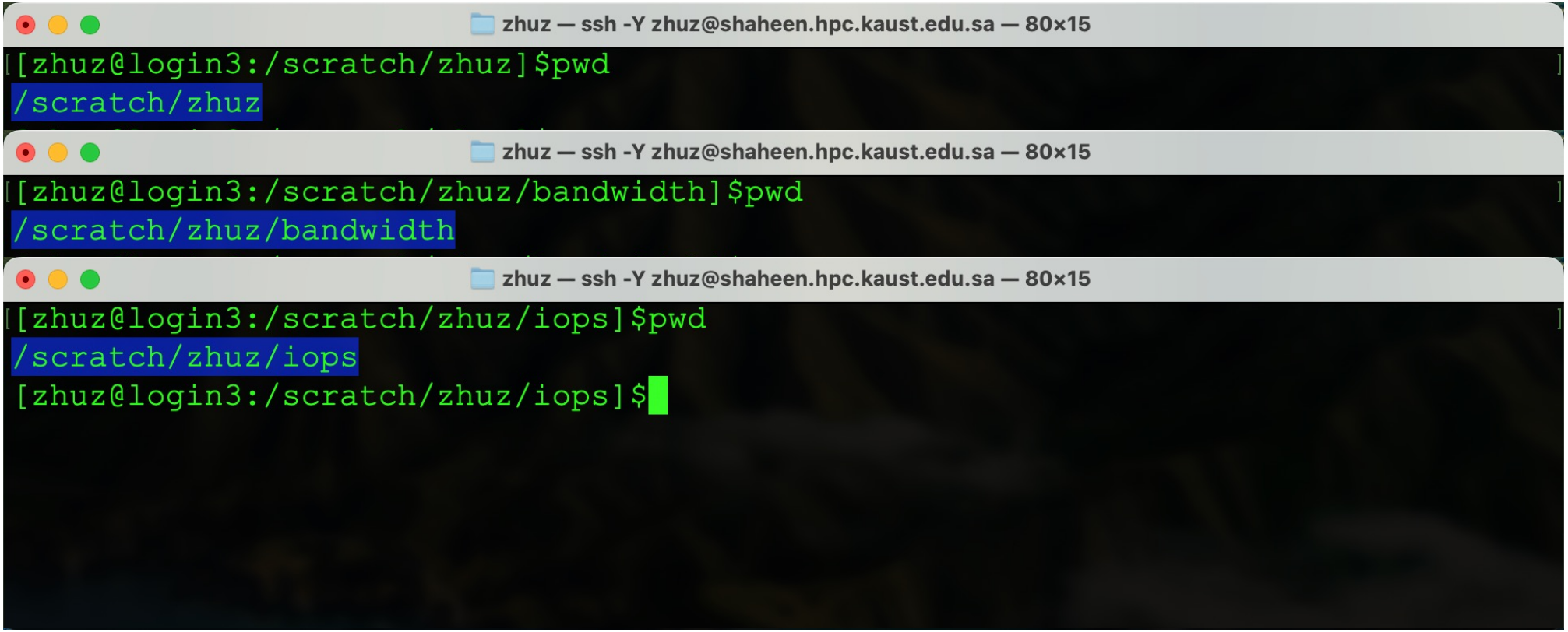
3 Different Working Directories

- Where to run? /scratch!
  - The only place to run
  - Remember to backup important data to /project



3 Different Working Directories

- /scratch/<username> - 3 tiers
  - capacity, bandwidth, iops



3 Different Working Directories

- /scratch/<username>
  - Tier “capacity”
  - Large capacity (10T per user by default)
  - Low performance in terms of bandwidth and iops
  - Used for calculations that are not sensitive to IO performance
  - Data purged after 60 days of no access

3 Different Working Directories

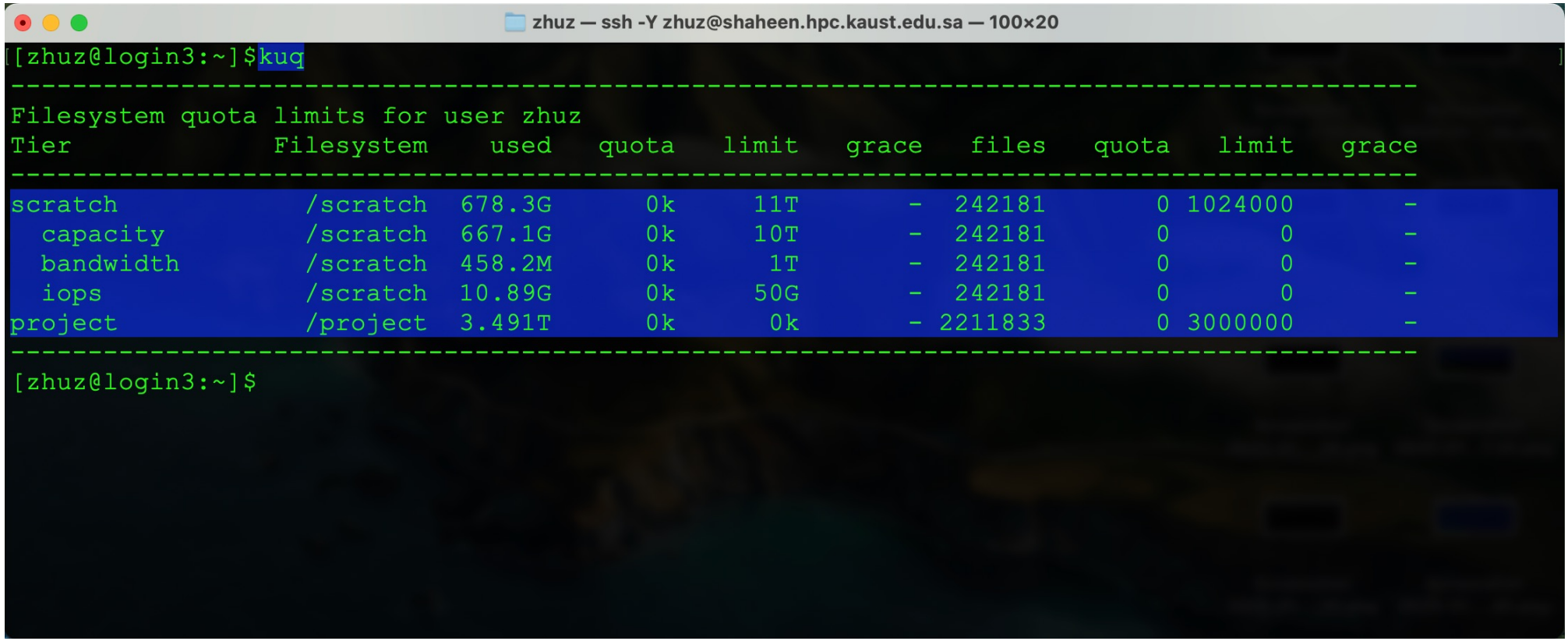
- /scratch/<username>/bandwidth
  - tier “bandwidth”
  - Low capacity (1T per user by default)
  - High performance in terms of IO bandwidth
  - Used for calculations that read/write a large amount of data
  - Data purged after 60 days of no access

3 Different Working Directories

- /scratch/<username>/iops
  - Tier “iops”
  - Low capacity (50G per user by default)
  - High performance in terms of # IO operations
  - Used for calculations that read/write large number of files, and software installation (conda, python, etc)
  - No data purging

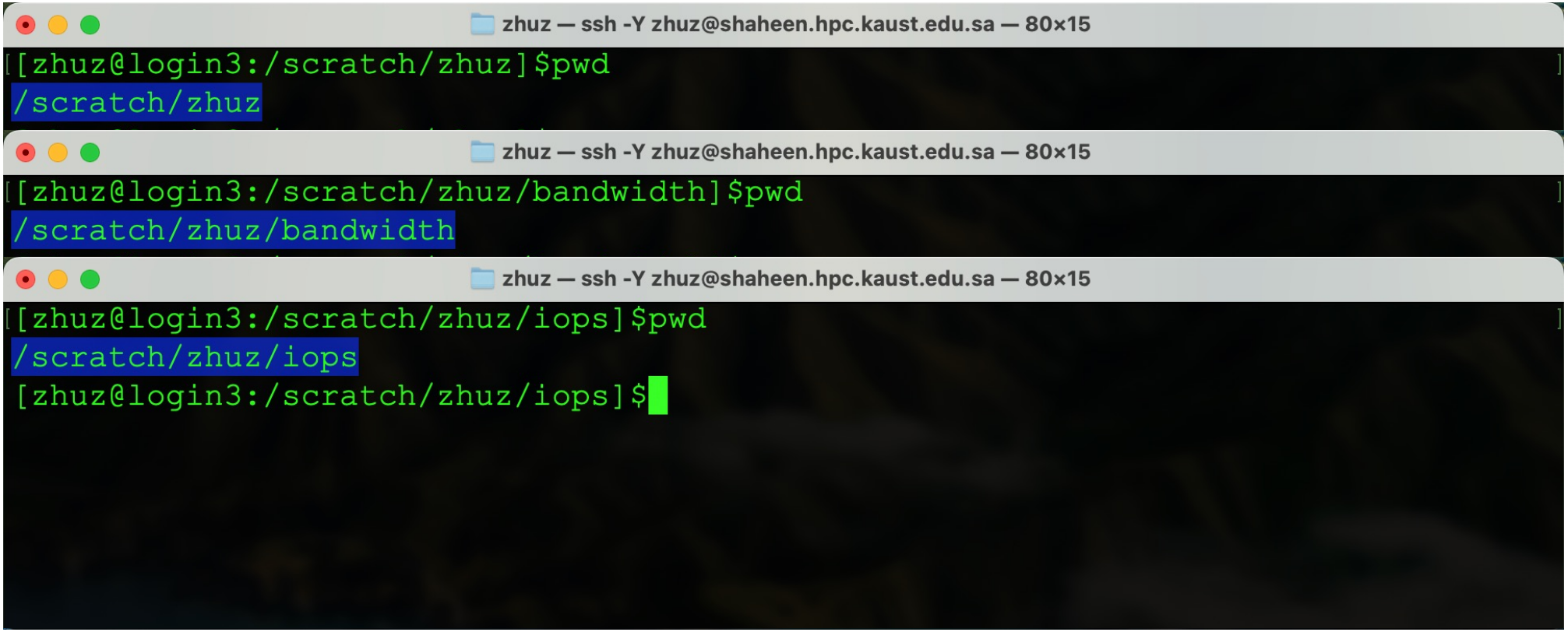
3 Different Working Directories

- Quota limits
  - kuq



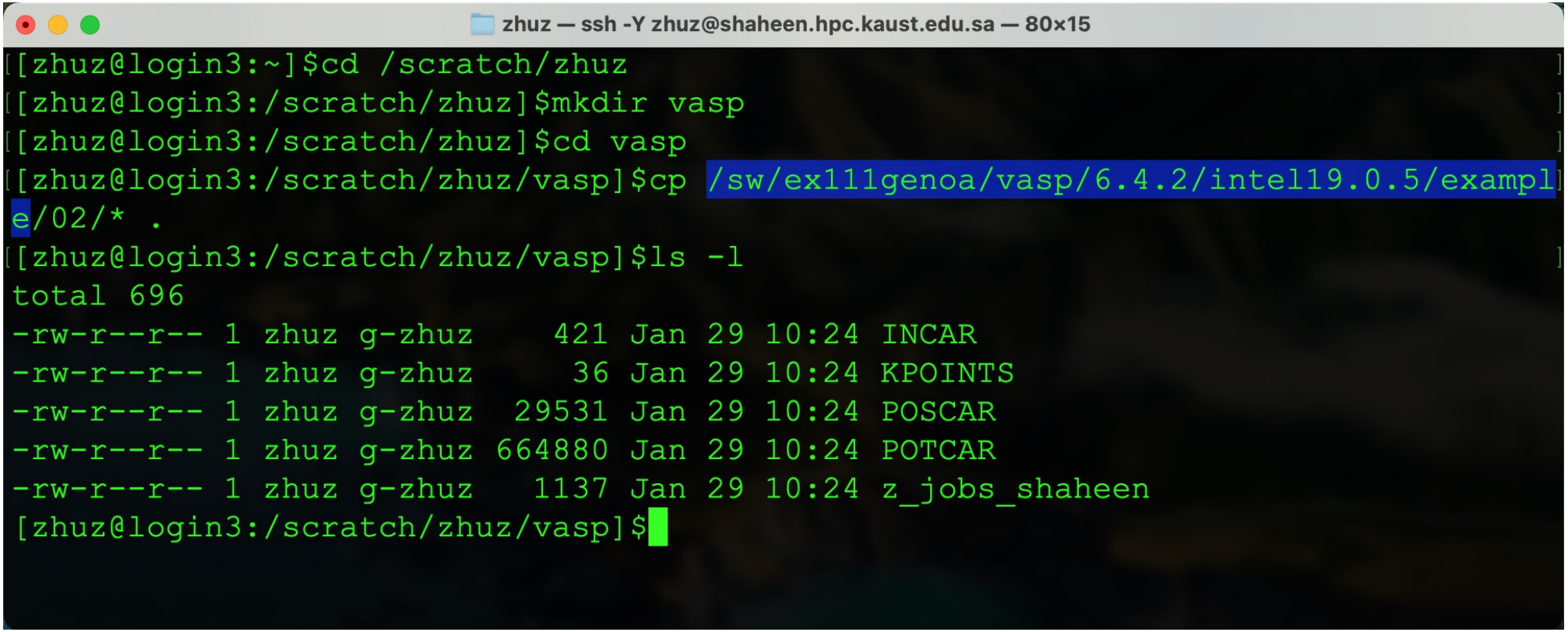
3 Different Working Directories

- Which tier to use? It depends!
  - Do your own tests



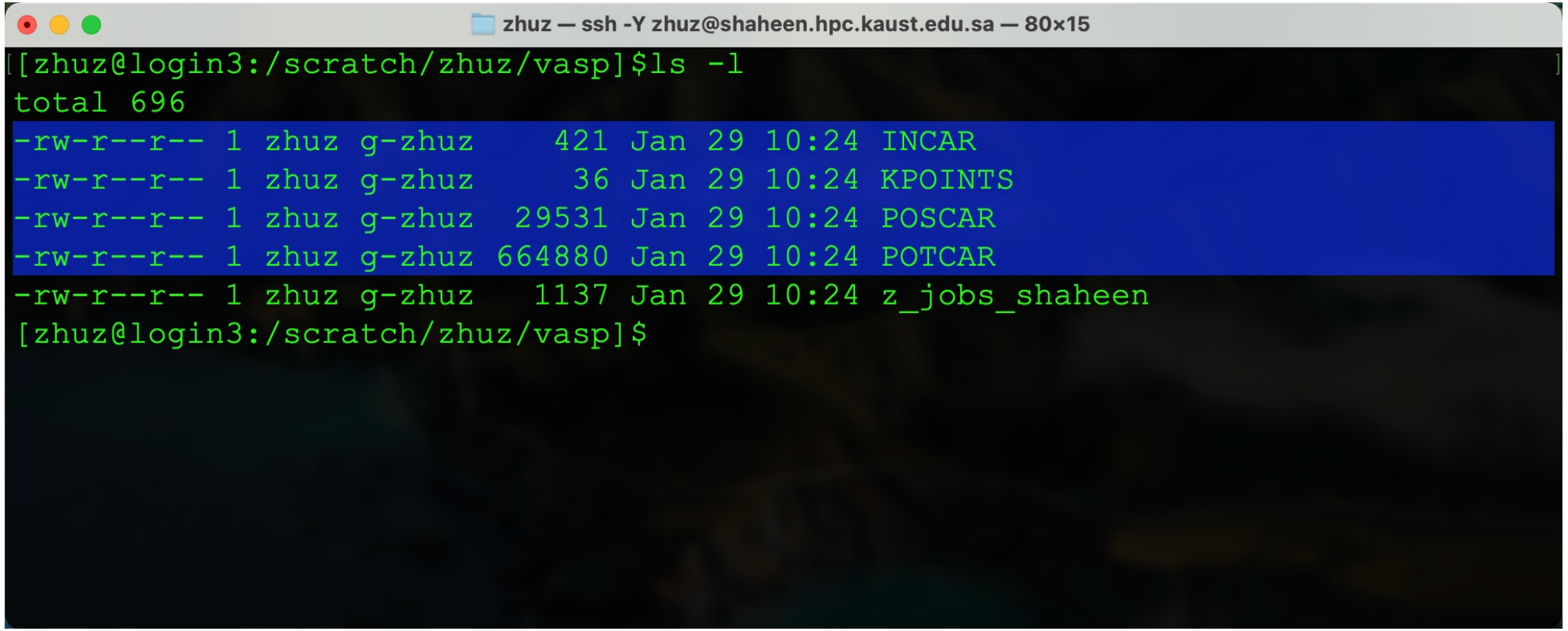
Prepare Input Files

- Examples under Installation Folder
  - /sw/ex111genoa/*code*/*ver*/*compilation*/example
  - Inputs for application; Jobscript for Slurm



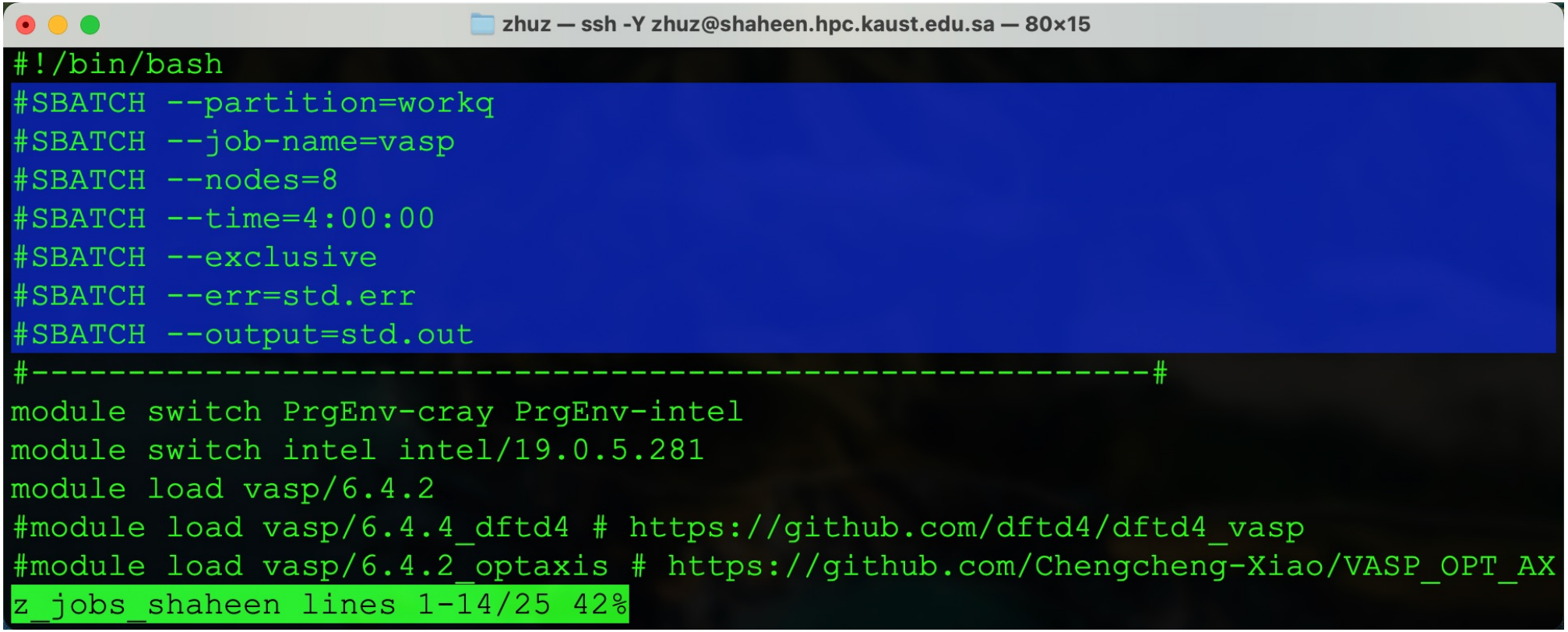
Prepare Input Files

- VASP Input Files
  - Upload from your own personal workstations
  - Modifying existing input files



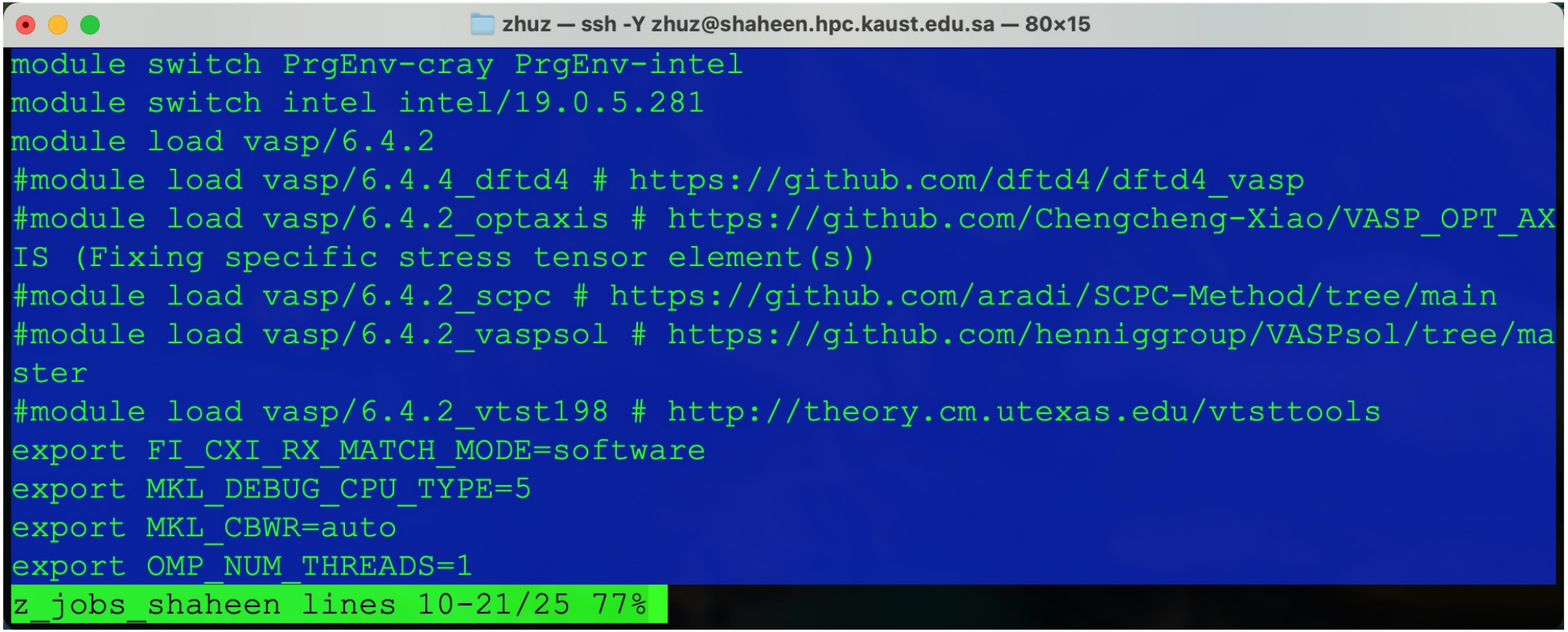
Prepare Input Files

- Slurm Jobscript
  - SLURM directives



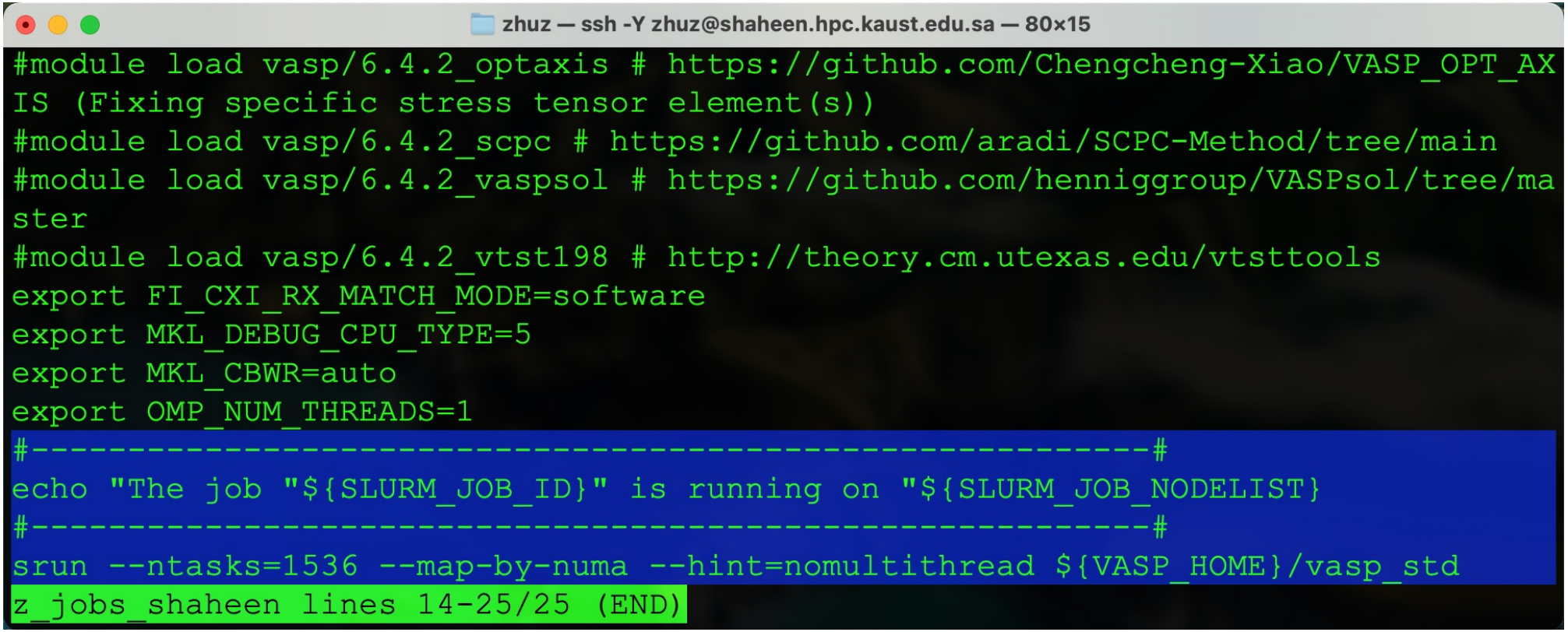
Prepare Input Files

- Slurm Jobscript
  - Environments settings



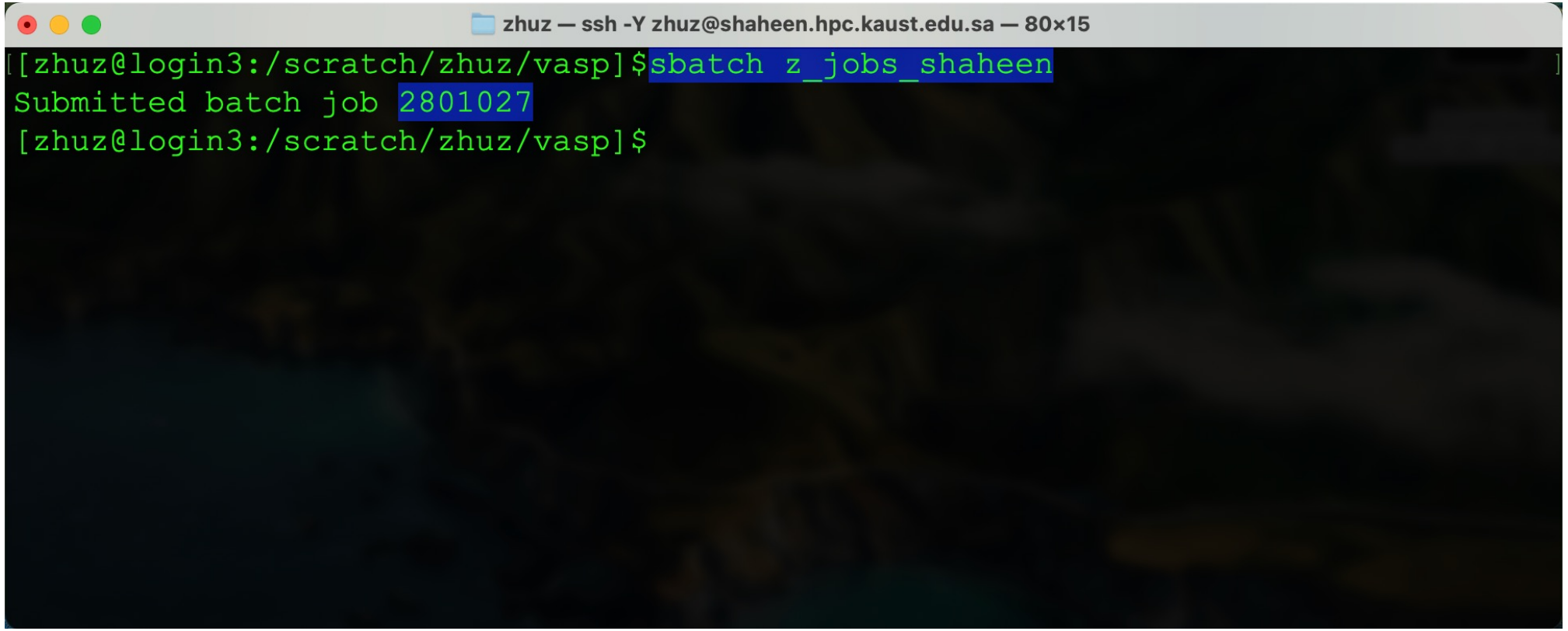
Prepare Input Files

- Slurm Jobscript
  - Commands to run



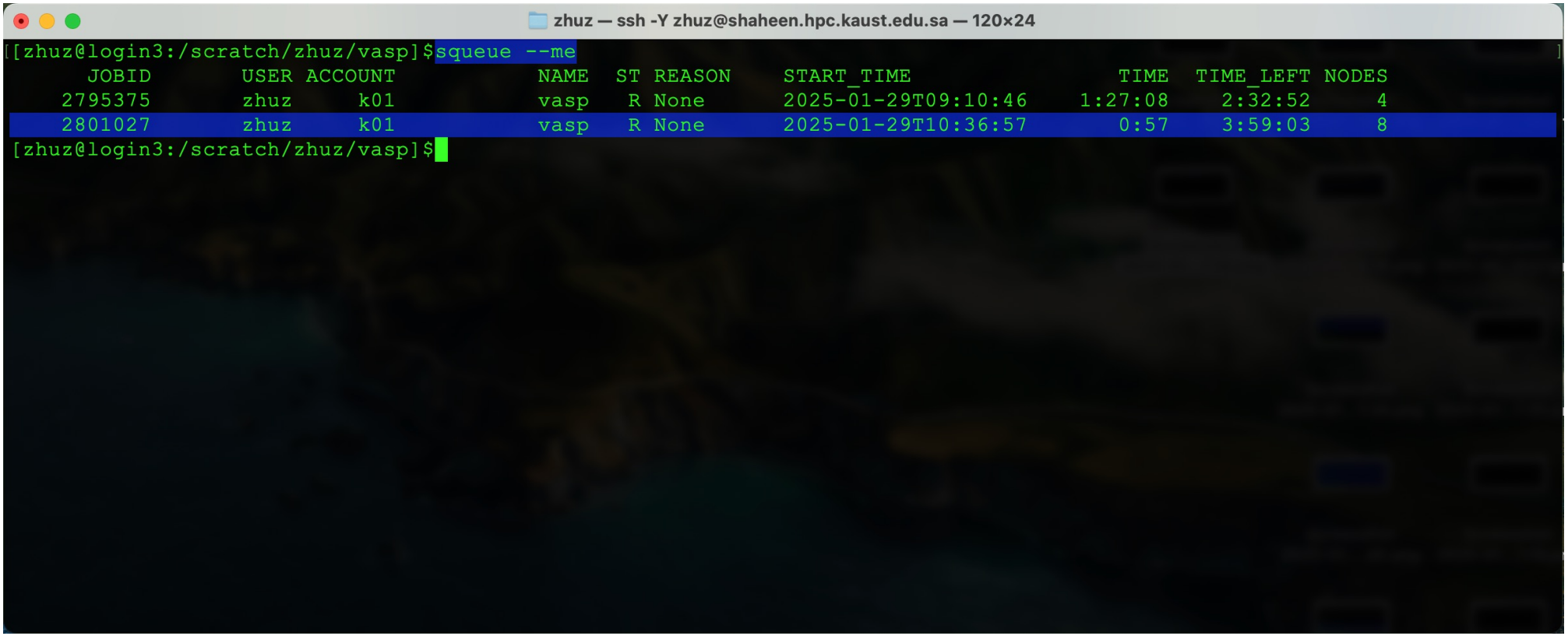
Job Submission

- sbatch
  - Submit jobs



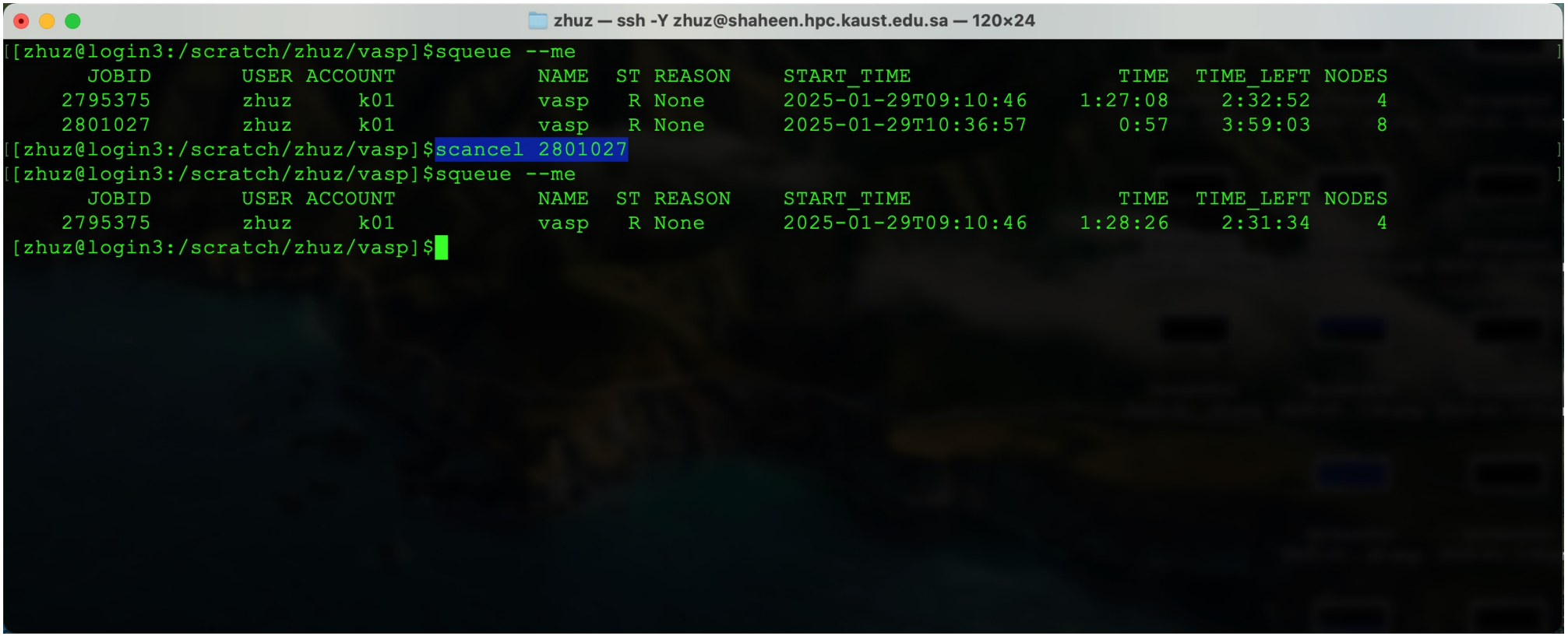
Job Submission

- squeue
  - Check job status



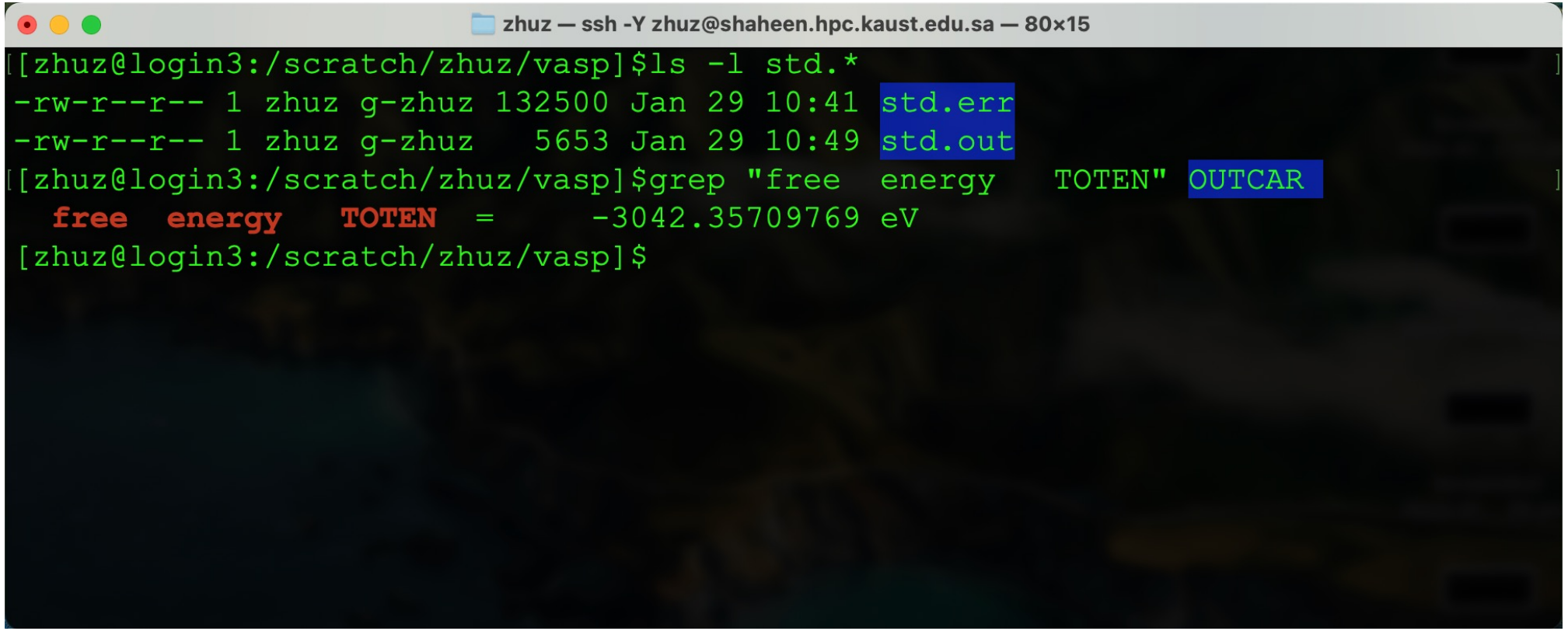
Job Submission

- scancel
  - Cancel jobs



Check Output Files

- Successful or Not? If yes, Analyze Results
  - Standard output/error files:std.out/std.err)
  - Application output files: OUTCAR



Tips

- Do not run directly on the login nodes
  - login nodes are shared
- It won't work to submit jobs from /home
  - /home is not seen on the compute nodes
- Backup important data from /scratch to /project (or /home, or your local computers)
  - Files in /scratch are not backed up, and are deleted automatically after 60 days
  - Do not confuse /scratch/project and /project

Tips

- Licensed software
  - Need license for commercial software
  - Different software have different license terms
    - VASP: you can use it on Shaheen, as long as you have your own license
    - Gaussian: you cannot use it on Shaheen, even if you have your own license
    - Contact us if you have any questions

Thank You!

help@hpc.kaust.edu.sa

Agenda

- 8:30am Welcome
- 8:35am Shaheen III Hardware Overview
- 8:55am How to apply on Shaheen III
- 9:05am Getting Started on Shaheen III
- 9:15am Software Environment
- 9:35am Job Scheduling
- 10:00am Coffee Break
- 10:15am Storage overview & Best practices
- 10:30am Applications software example: VASP workflow
- **10:50 am Applications software example: CFD applications**
- 11:10 am Applications software example: Bio informatics workflow
- 11:20-11.30am Q&A and Open Discussion

Shaheen III Survey