Introduction to the Amarel Cluster

W. Evan Johnson, Ph.D.
Professor, Division of Infectious Disease
Director, Center for Data Science
Co-Direcor, Center for Biomedical Informatics and Health AI
Rutgers University – New Jersey Medical School

2025-07-27

Introduction to the Amarel HPC Cluster

HPC at Rutgers

About Amarel

In July 2017, the Office of Advanced Research Computing (OARC) unveiled Amarel, a "condominium" style computing environment developed to serve the university's wide-ranging research needs. The Amarel cluster provides a shared platform that optimizes resources for the benefit of all users. Named in honor of **Dr. Saul Amarel**, one of the founders of the Rutgers Computer Science Department and contributor to advanced computing and artificial intelligence research and methodologies, Amarel is designed to suit many different research applications.

Through Aberdeen (late 2023):

- 560 compute nodes
  - 27,872 Intel Xeon cores
  - 224 GPUs
  - Open OnDemand servers
  - InfiniBand FDR & EDR fabric

More information:

- Community-contributed software repository
- Spans three Rutgers data centers (Piscataway, Newark, and Camden) producing a unified compute, data, and storage system
- Rolling node/phase retirement with new node replacement, beginning in spring 2021

Key Features of Amarel

1. **Compute Nodes**: Amarel consists of multiple compute nodes with varying specifications, including CPUs and GPUs.
2. **Parallel File System**: A high-speed parallel file system for storing and accessing large datasets.
3. **Job Scheduler**: A job scheduling system for managing computing resources efficiently.
4. **Software Stack**: A comprehensive collection of software packages and libraries for various scientific computing tasks.
5. **Networking**: High-speed interconnects for fast communication between compute nodes.

Getting Started with Amarel

If you are off campus, you will need access to VPN:

Click here for Rutgers VPN instructions for Windows Users

Click here for Rutgers VPN instructions for Mac Users

Getting Started with Amarel

1. **Request Access**: Contact OARC to request access to Amarel
2. **Training Workshops**: Attend training workshops offered by OARC to learn how to use Amarel effectively.
3. **Documentation and Support**: Explore documentation and seek support from OARC staff and the user community.
4. **Start Small**: Begin with small-scale experiments and gradually scale up as needed.

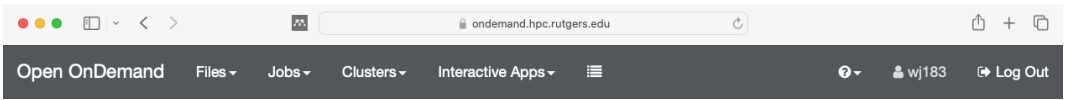
WEJ LAB

Introduction to OnDemand for HPC Clusters

Center for Data Science

What is OnDemand?

- OnDemand is a web-based platform that provides a user-friendly interface for accessing and managing (HPC) clusters.
- It simplifies the process of submitting and monitoring jobs, accessing software, and managing data on HPC systems.
- Go to: http://ondemand.hpc.rutgers.edu (VPN for off campus).



OPEN OnDemand

OnDemand provides an integrated, single access point for all of your HPC resources.

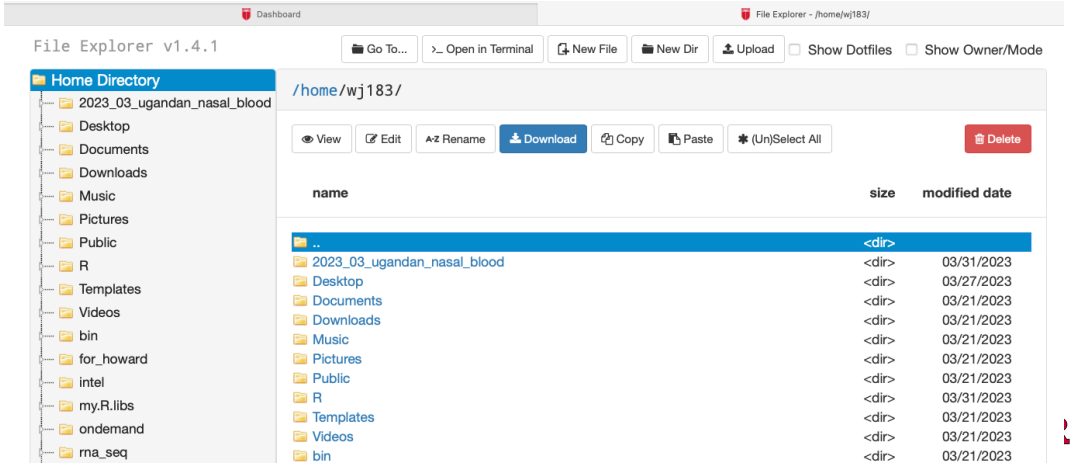
Key Features of OnDemand

1. **Web Interface**: Access HPC resources through a web browser from anywhere with an internet connection.
2. **Job Submission**: Submit and manage computational jobs without needing to use command-line interfaces.
3. **File Management**: Upload, download, and manage files and data on the HPC cluster directly from the browser.
4. **Interactive Sessions**: Launch interactive computing sessions for data analysis and exploration.
5. **Software Environment**: Access a variety of software packages and development tools installed on the HPC cluster.

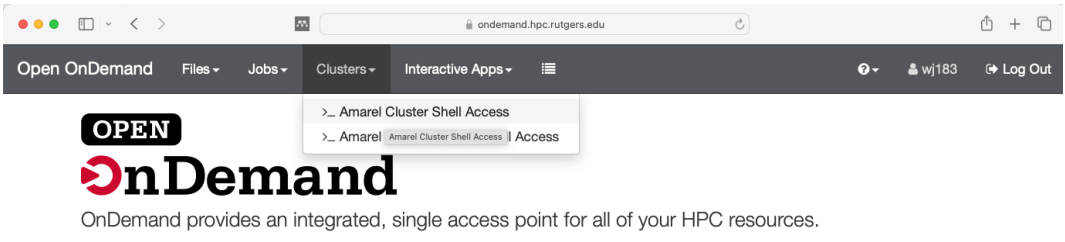
Benefits of OnDemand

- **User-Friendly**: OnDemand provides a simplified interface, making HPC resources more accessible to a wider range of users.
- **Remote Access**: Users can access HPC resources remotely without needing to install any special software.
- **Increased Productivity**: Streamlined workflows and intuitive interfaces help users focus on their research instead of dealing with technical complexities.

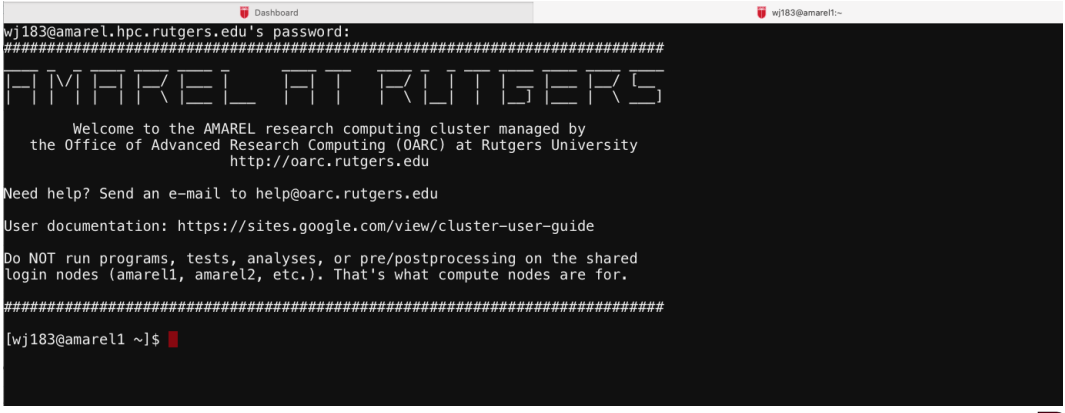
Accessing Files



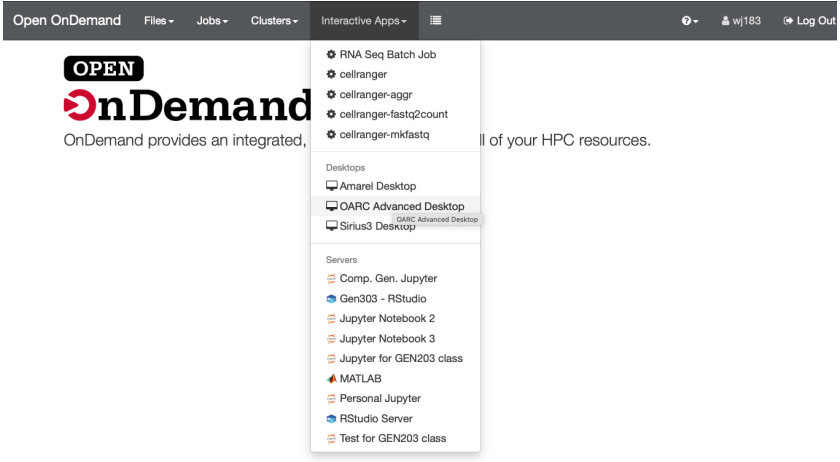
Accessing the Terminal



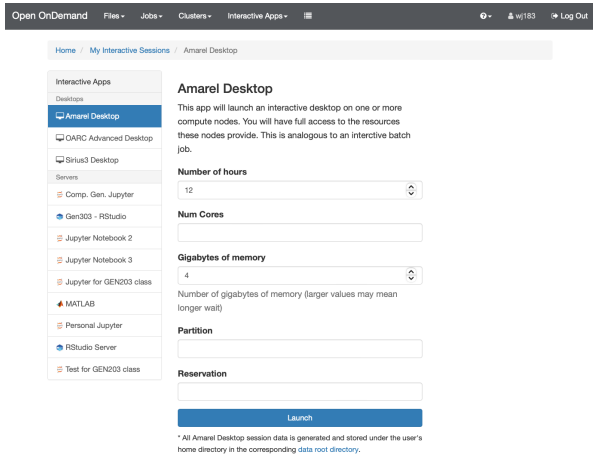
Accessing the Terminal



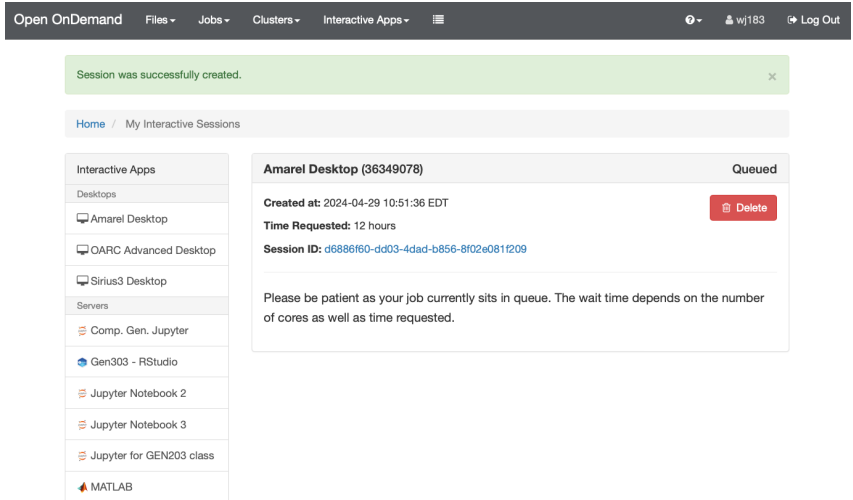
Accessing the Amarel Desktop through OnDemand



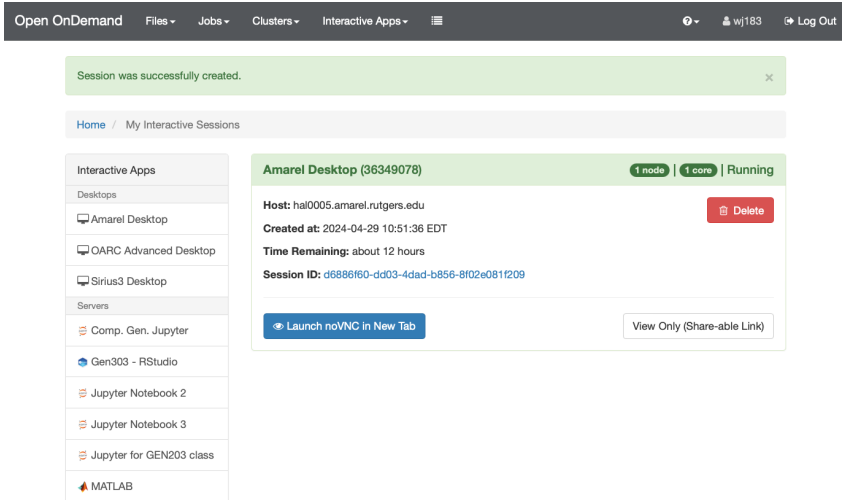
Accessing the Amarel Desktop through OnDemand



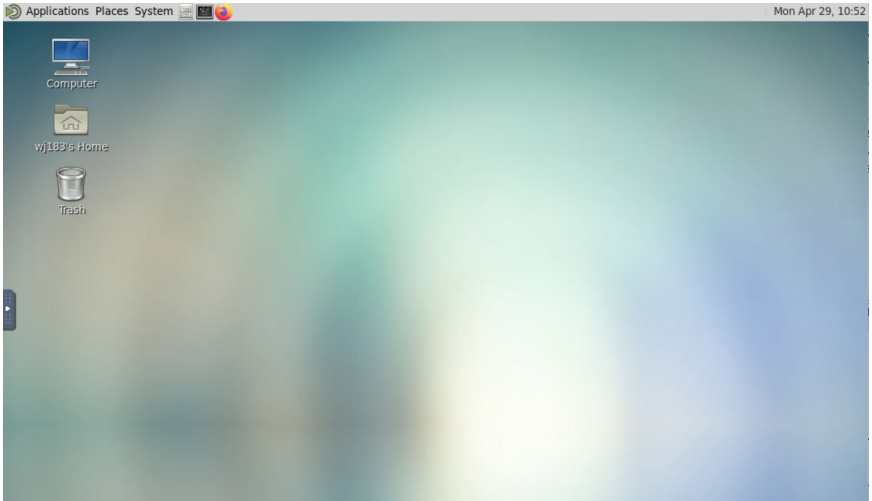
Accessing the Amarel Desktop through OnDemand



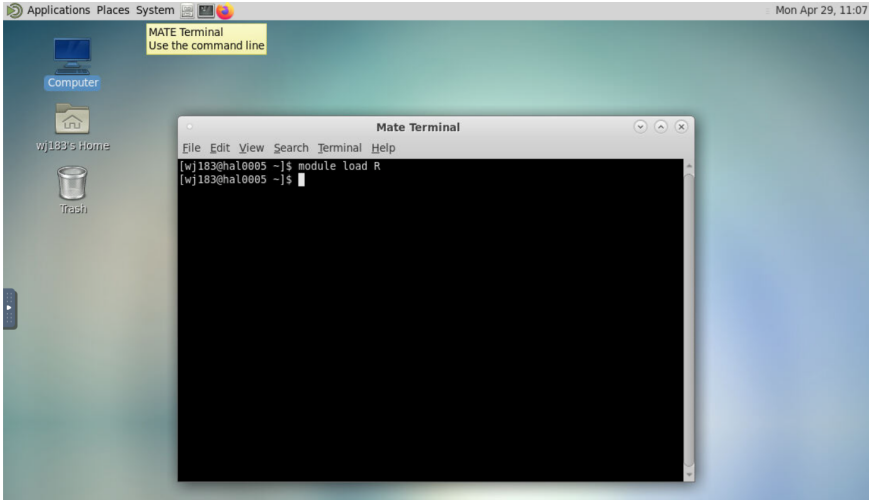
Accessing the Amarel Desktop through OnDemand



Accessing the Amarel Desktop through OnDemand



Accessing the Amarel Desktop through OnDemand

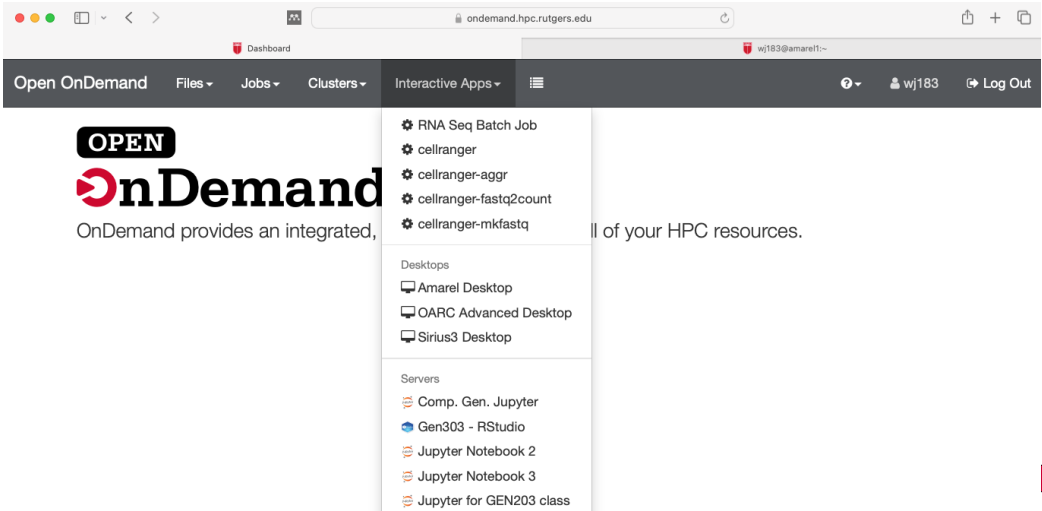


Accessing R through the terminal

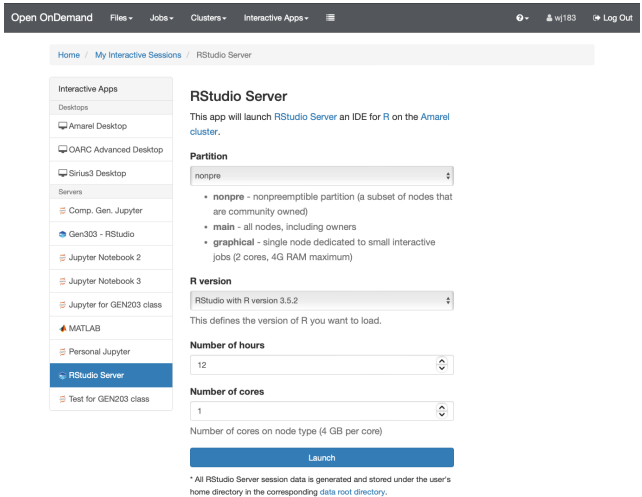
You first need to attach R to your session (using Amarel Desktop):

```
[wj183@hal0005 ~]$ module load R
[wj183@amarel1 ~]$ R
```

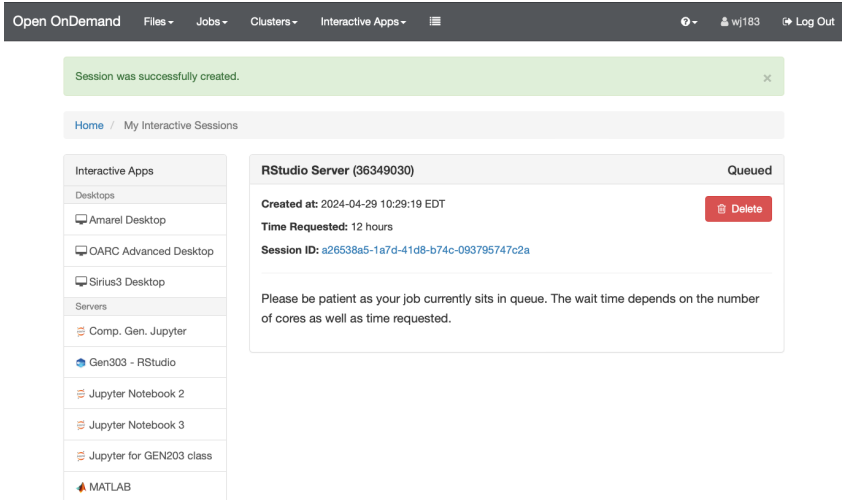
Accessing the RStudio through OnDemand



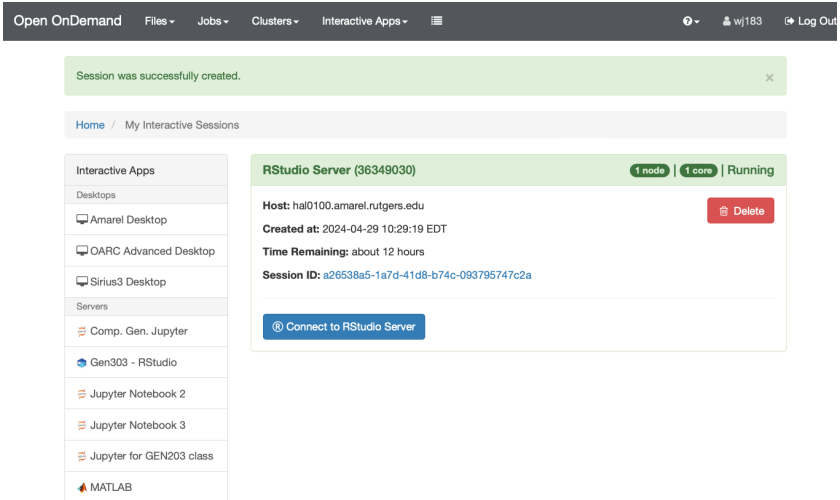
Accessing the RStudio through OnDemand



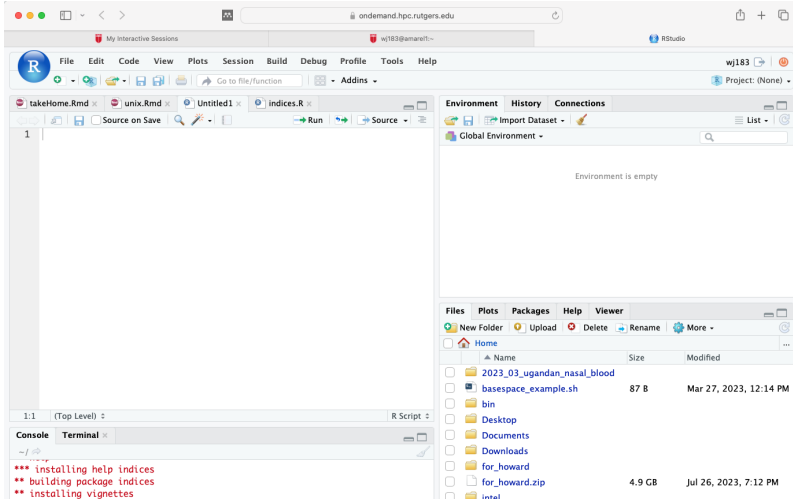
Accessing the RStudio through OnDemand



Accessing the RStudio through OnDemand



Accessing the RStudio through OnDemand



WEJ LAB

Introduction to Cluster Batch Job Scheduling with SLURM

Center for Data Science

What is SLURM?

- SLURM (Simple Linux Utility for Resource Management) is an open-source job scheduler and resource manager used on many high-performance computing (HPC) clusters.
- It allows users to submit, manage, and monitor batch jobs on computing clusters efficiently.

Key Concepts

1. **Partition**: A partition is a logical grouping of compute nodes with similar hardware characteristics or usage policies.
2. **Job Script**: A job script is a text file containing instructions for the scheduler, including job parameters, resource requests, and executable commands.
3. **Resource Allocation**: Users specify the required resources for their jobs, such as CPUs, memory, and runtime, in the job script.
4. **Job Status**: Monitor the status of their submitted jobs using SLURM commands or monitoring tools.

Submitting a Job with SLURM

1. **Create a Job Script**: Write a job script using a text editor, specifying job parameters and commands.
2. **Submit the Job**: Use the `sbatch` command to submit the job script to SLURM.
3. **Monitor Job Status**: Check the status of the submitted job using commands like `squeue` or monitoring tools provided by SLURM.
4. **Retrieve Results**: Once the job completes, retrieve the results and data generated by the job from the designated output directory.

Example Job Script

```
#!/bin/bash
#SBATCH --job-name=my_job
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=4
#SBATCH --mem=8G
#SBATCH --time=1:00:00

# Execute your command here
./my_executable input_file.txt > output_file.txt
```

WEJ LAB
Center for Data Science

Best Practices

- Resource Requests: Specify accurate resource requirements to ensure optimal resource allocation and job performance.
- Job Prioritization: Understand the partition structure and prioritize jobs accordingly to optimize resource utilization.
- Job Dependencies: Use job dependencies to sequence jobs and ensure proper execution order when necessary.
- Error Handling: Include error handling and logging mechanisms in job scripts to facilitate troubleshooting and debugging.

Conclusion

SLURM is a powerful tool for managing batch job scheduling and resource allocation on cluster computing systems. By understanding key concepts and following best practices, users can effectively utilize SLURM to submit, manage, and monitor computational tasks on HPC clusters. Explore documentation and seek support from the cluster administrators or user community to learn more about SLURM and optimize your workflow.

Session info

```
sessionInfo()
```

```
## R version 4.5.1 (2025-06-13)
## Platform: aarch64-apple-darwin20
## Running under: macOS Sequoia 15.5
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/lib/libRlapack.dylib;  LAPACK version 3.12.1
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## loaded via a namespace (and not attached):
##  [1] compiler_4.5.1    fastmap_1.2.0     cli_3.6.5         tools_4.5.1
##  [5] htmltools_0.5.8.1 rstudioapi_0.17.1 yaml_2.3.10       rmarkdown_2.29
##  [9] knitr_1.50        xfun_0.52         digest_0.6.37     rlang_1.1.6
## [13] evaluate_1.0.4
```