

# R Workflows on ManeFrame II (M2)

---

Robert Kalescky

HPC Applications Scientist

August 31, 2021

Research and Data Sciences Services

Office of Information Technology

Center for Research Computing

Southern Methodist University



Research Support

ManeFrame II (M2)

Accessing ManeFrame II (M2)

Slurm

Reproducibility and Environments

Additional Resources

## Research Support

---



- Provide research computing support, consultations, and collaborations
- Data Science - Dr. Eric Godat
- High-Performance Computing - Dr. Robert Kalescky & Dr. John LaGrone
- Machine Learning / Artificial Intelligence - Dr. Tue Vue
- Custom Devices (IOT, wearables, etc.) - Guillermo Vasquez



- Maintains our primary shared resource for research computing, ManeFrame II (M2), in collaboration with OIT
- Provides research computing tools, support, and training to all faculty, staff, and students using research computing resources
- [www.smu.edu/crc](http://www.smu.edu/crc) has documentation and news
- [help@smu.edu](mailto:help@smu.edu) or [rkalescky@smu.edu](mailto:rkalescky@smu.edu) or [jlagrone@smu.edu](mailto:jlagrone@smu.edu) for help
- Request an account at [www.smu.edu/crc](http://www.smu.edu/crc)

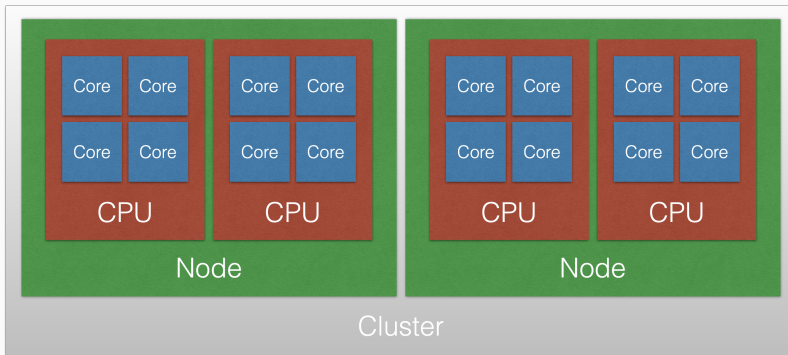


Date	Time	Workshop
February 2	2-4	ManeFrame II (M2) Introduction
February 9	2-4	Workflows in R
February 15	3-5	Finding and Preparing Text Data Sets for Mining
February 16	1-4	Machine Learning with Python Part 1
February 17	12-1	AI for the Non-Expert
February 18	12-1	Introduction to GitHub
February 23	2-4	Containers and Spack
March 2	2-4	ManeFrame II (M2) Introduction
March 3	1-4	Data Science Workflow with R
March 8	3-6	Introduction to Python for Text Mining
March 9	1-4	Machine Learning with Python Part 2
March 22	3-6	Getting Support for Text Mining
March 23	2-4	Shared Memory Parallelism
March 30	1-4	Deep Learning with Python Part 1
April 6	2-4	ManeFrame II (M2) Introduction
April 13	2-4	Accelerator Libraries and APIs
April 20	1-4	Deep Learning with Python Part 2
April 27	2-4	MPI/NCCL/SHMem
May 4	2-4	ManeFrame II (M2) Introduction

**Table 1:** Workshops will be held each Wednesday from 2:00 to 4:00 PM. Sessions will typically be recorded and posted along with session materials. Register on the Library Workshop Calendar <https://libcal.smu.edu/calendar/libraryworkshops>

## ManeFrame II (M2)

---



**Figure 1:** A cluster is a collection of individual computers networked together. Applications can be configured to run on all available compute resources.





Type	Quantity	Cores	Memory [GB]	Additional Resources
Standard-Memory	176	36	256	
Medium-Memory-1	35	36	768	
Medium-Memory-2	4	24	768	3 TB SSD local scratch
High-Memory-1	5	36	1,536	
High-Memory-2	6	40	1,536	3 TB SSD local scratch
GPGPU-1	36	36	256	NVIDIA P100 GPU has 3,584 CUDA cores and 16 GB CoWoS
MIC-1	36	64	384	16 GB of high bandwidth (400 GB/s) stacked memory
VDI	5	36	256	NVIDIA Quadro M5000 GPU
v100x8	3	36	768	8 NVIDIA V100 GPUs with 5,120 CUDA cores and 32 GB CoWoS
Faculty Partner Nodes	3			Various research specific NVIDIA GPU configurations
ManeFrame II	354	11,276	120 TB	2.8 PB storage and InfiniBand network



Partition	Duration	Cores	Memory [GB]
development	2 hours	various	various
htc	1 day	1	6
standard-mem-s	1 day	36	256
standard-mem-m	1 week	36	256
standard-mem-l	1 month	36	256
medium-mem-1-s	1 day	36	768
medium-mem-1-m	1 week	36	768
medium-mem-1-l	1 month	36	768
medium-mem-2	2 weeks	24	768
high-mem-1	2 weeks	36	1538
high-mem-2	2 weeks	40	1538
mic	1 week	64	384
gpgpu-1	1 week	36	256
v100x8	1 week	1	20
fp-gpgpu-2	various	24	128
fp-gpgpu-3	various	40	384



- \$HOME**
- Default file system when logging into M2, e.g. `/users/$USER`.
  - Space should be used to write, edit, compile programs, and job submission scripts, etc.
  - Restricted by quotas (200 GB) and backed-up.

- \$WORK**
- Long term storage at `/work/users/$USER`.
  - Restricted by quotas (8 TB) and not backed-up.

- \$SCRATCH**
- Scratch space at `/scratch/users/$USER`.
  - Treat \$SCRATCH as a volatile file system that is not backed-up.

## Accessing ManeFrame II (M2)

---



- SSH
- Provides secure command line access to the login nodes.
  - Graphical applications require X11-forwarding.
  - Five login nodes all accessible via `m2.smu.edu`.
  - [M2 SSH Instructions](#)

- HPC Portal
- Provides web-based access to M2.
    - File access
    - Terminal access
    - JupyterLab (Notebooks)
    - RStudio
    - Remote Desktops

- SFTP
- Transfer files to and from M2 using an SFTP client.
  - [M2 SFTP Instructions](#)



- Provides an integrated single access point for HPC resources on the ManeFrame II (M2) supercomputer
- Accessing the Portal:
  - Access to the HPC portal requires an existing M2 account
  - Go to **hpc.smu.edu**
  - Sign in using your SMU ID and SMU password
- HPC Portal Documentation
- HPC Portal Walkthrough Video



Slurm

---



```
1  monte_carlo_pi <- function(n) {  
2    x=runif(n)  
3    y=runif(n)  
4    z=sqrt(x^2+y^2)  
5    length(which(z<=1))*4/n  
6  }  
7  
8  args = commandArgs(trailing=TRUE)  
9  monte_carlo_pi(strtoi(args[1]))  
10
```

Listing 1: Serial algorithm to estimate the value of Pi.





```
1  library(parallel)
2
3  monte_carlo_pi <- function(n) {
4    z = sqrt(runif(1)^2+runif(1)^2)
5    return(if(z<=1) 1 else 0)
6  }
7
8  args = commandArgs(trailing=TRUE)
9  n = strtoi(args[1])
10 c = strtoi(args[2])
11 cl = makeCluster(getOption("cl.cores", c))
12 result = clusterApply(cl, 1:n, fun=monte_carlo_pi)
13 sum(unlist(result))*4/n
14 stopCluster(cl)
15
```

Listing 2: Parallel algorithm to estimate the value of Pi.



```
1 module load r/4.0.2
2 srun -p htc --mem=6G --pty $SHELL
3
```

Listing 3: Using `srun` to log into a compute node to run commands interactively.



```
1 module load r/4.0.2
2 srun -p htc --mem=6G Rscript pi_monte_carlo_serial.R 1000
3
```

Listing 4: Using `srun` to run commands directly on a compute node.



```
1 module load r/4.0.2
2 sbatch -p htc --mem=6G --wrap "sleep 30; time Rscript pi_monte_carlo_serial.R
  ↪ 1000"
3
```

Listing 5: Using `sbatch --wrap` wrap a commands in an `sbatch` script that is then submitted to the queue can run non-interactively.



```
1  #!/bin/bash
2  #SBATCH -J R
3  #SBATCH -o R_%j.out
4  #SBATCH -p htc
5  #SBATCH --mem=6G
6
7  module purge
8  module load r/4.0.2
9
10 time Rscript pi_monte_carlo_serial.R 1000
11
```

Listing 6: Using `sbatch` run serial computations via an `sbatch` script.



```
1  #!/bin/bash
2  #SBATCH -J pi
3  #SBATCH -o pi_%j.out
4  #SBATCH -p development
5  #SBATCH --cpus-per-task=2
6  #SBATCH --mem=6G
7
8  module purge
9  module load r/4.0.2
10
11  time Rscript pi_monte_carlo_parallel.R 10000000 ${SLURM_CPUS_PER_TASK}
12
```

Listing 7: Using `sbatch` run parallel computations via an `sbatch` script.



```
1  #!/bin/bash
2  #SBATCH -J pi_array
3  #SBATCH -o pi_array_%a-%A.out
4  #SBATCH --array=1-4%2
5  #SBATCH -p development
6  #SBATCH --mem=6G
7
8  module purge
9  module load r/4.0.2
10
11  time Rscript pi_monte_carlo_serial.R $((100**${SLURM_ARRAY_TASK_ID}))
12
```

Listing 8: Using `sbatch --array` run parallel jobs via a single `sbatch` script.

## Reproducibility and Environments

---





- Use Git for version control
- Use **renv** for package versioning



- GitHub (<https://github.com>) is a great tool for working with Git repositories
- Use or add your SMU email address to your GitHub account to access all GitHub Enterprise features



- renv::init()** Initialize new project-local environment with a private R library
- renv::snapshot()** Save the current state of the project library
- renv::restore()** Restore the previously saved state of the project library, e.g. reloading the environment on another machine



1. Create new Git repository on GitHub
2. Clone the new repository on M2
3. Create new RStudio project in RStudio via M2's Portal
4. Create clean package environment via `renv::init(bare = TRUE)`
5. Install custom packages, e.g. `renv::install("mcmc")`
6. Create environment snapshot via `renv::snapshot(packages="mcmc")`
7. Commit and push changes to the Git repository

## Additional Resources

---



- [Advanced R](#)
- [renv](#)
- [RStudio Environments](#)



Need help or have questions?

[rkalescky@smu.edu](mailto:rkalescky@smu.edu)

[jlagrone@smu.edu](mailto:jlagrone@smu.edu)

[help@smu.edu](mailto:help@smu.edu) (include HPC in the subject line)