Using and Installing Custom Software on HARDAC

Part 1 - Environment Modules
Goals of Class

Be able to

• Use cluster-provided software
• Install and use R packages
Class Requirements

Access to the HARDAC cluster

• Email gcb-help@duke.edu to request access if necessary

Complete previous gcb-academy course

• High Performance Computing (HPC) / SLURM Best Practices for HARDAC
Cluster Software Challenges

• Conflicting version and dependency requirements

• Ability to retain all versions of software used in analyses

❌ Global software installation

✔️ Environment Modules
Environment Modules

• Software packages already on the HARDAC cluster
• Must be activated to use
• Commonly used on computing clusters
• Only higher profile software packages are provided
Class Exercises Directory

Create a directory to hold files for this class

```sh
$ mkdir software_class
```

Copy example fasta data files into your new directory

```sh
$ cp /data/common/software_class/*.fasta software_class
```

Change to the new directory

```sh
$ cd software_class
```
Environment Module commands

The `module` command updates your shell environment adding and removing directories containing software packages

- List available modules - `module avail`
- Search modules - `module avail <searchTerm>`
- Activate a module - `module load <moduleName>`
- Deactivate a module - `module unload <moduleName>`
- Deactivate all loaded modules - `module purge`
List Environment Modules

$ module avail

--More--

NOTE: The module command does not work on the login node
$ module avail blast

------------------------ /data/itlab/helmod/modulefiles/Core ------------------------

ncbi-blast/2.2.22-fasrc01       samblaster/0.1.24-gcb01
ncbi-blast/2.3.0+-gcb01       (D)

Where:
(D): Default Module
...
Use an Environment Module

load environment module

```
$ module load ncbi-blast/2.3.0+-gcb01
```

verify command is working as expected

```
$ blastp -h
USAGE
    blastp [-h] [-help] [-import_search_strategy filename]
           [-export_search_strategy filename] [-task task_name] [-db database_name]
```

cleanup - purge environment modules

```
$ module purge
$ blastp -h
bash: blastp: command not found
```
#!/bin/bash
module load ncbi-blast/2.3.0+-gcb01
blastp -query query.fasta -subject subject.fasta -out result.txt

Run blastp.sh using the `sbatch` command

$ sbatch blastp.sh
Submitted batch job 23231383

Look at results (once job completes)

$ cat result.txt
R Environment Module

Comes with many R packages already installed
Allows installing additional R packages
  • Requires a directory to hold the packages
  • Must update the \texttt{R\_LIBS\_USER} environment variable
Create a directory named "envs" in your lab's data directory:

```bash
$ mkdir /data/<LabName>/envs
```

We will store custom environments such as a directory of R packages within this directory.

NOTE: In this course when you see `<LabName>` replace it with your lab's name.
Install an R package

Create a directory to hold R packages

```
$ mkdir /data/<LabName>/envs/r_libs
```

Load R module

```
$ module load R/3.6.1-gcb03
```

Add your directory path to the beginning of R_LIBS_USER

```
$ export R_LIBS_USER=/data/<LabName>/envs/r_libs:$R_LIBS_USER
```

Install the "here" R package

```
$ R
install.packages('here')
q()
```
Verify the R Package

Run R interpreter and test the package

```r
$ R
library(here)
here()
q()
```
Use the R Package - part 1

Create a file named here.R with contents:

```r
library(here)
here()
sessionInfo()
```

Create a file named here.sh with contents:

```bash
#!/bin/bash
module load R/3.6.1-gcb03
export R_LIBS_USER=/data/<LabName>/envs/r_libs:$R_LIBS_USER
Rscript here.R
```
Use the R Package - part 2

Run here.sh using the `sbatch` command

```bash
$ sbatch here.sh
Submitted batch job 23231383
```

Verify expected output

```bash
$ cat slurm-...
R version 3.6.1...
...
[1] "/gpfs/fs1/home/jpb67"
...
Real World Example - Seurat

• Review the software installation instructions
• Look for environment modules that contain the software or it's requirements
• Install Software
• Verify Software
Real World Example - Seurat Documentation

Installation Instructions for Seurat

1. Install R, version 3.6 or greater
2. Recommended: Install R Studio

Install from CRAN

Seurat is available on CRAN for all platforms. To install, run:

```r
# Enter commands in R (or R studio, if installed)
install.packages('Seurat')
library(Seurat)
```
Real World Example - Requirements

Is there a seurat module?

$ module avail seurat

If no seurat module .. is there a R 3.6+ module?

$ module avail R
Real World Example - Install Seurat

Install using the same method as before

```bash
$ module load R/3.6.1-gcb03
$ export R_LIBS_USER=/data/<LabName>/envs/r_libs:$R_LIBS_USER
$ R
install.packages('Seurat')
```

NOTE: Installing/Compiling software can take a while
Real World Example - Verify Seurat

Try first few lines from a Seurat tutorial

```r
library(dplyr)
library(Seurat)
library(patchwork)

# Load the PBMC dataset
pbmc.data <- Read10X(data.dir = "../data/pbmc3k/filtered_gene
# Initialize the Seurat object with the raw (non-normalized data)
```
When Software Installations Fail

Conda Environments
- presented in HARDAC Software Part 2

Singularity Images
- presented in HARDAC Software Part 3
Helpful Resources

HARDAC WIKI

https://wiki.duke.edu/display/HAR/

Requesting Help or Video Feedback

cgb-help@duke.edu