

PIE Lab Manual

The PIE lab

2019-07-15

Contents

	5
1 People	7
1.1 Lab	7
1.2 PIs	7
2 Resources	9
2.1 R	9
3 Talapas Overview	11
3.1 Using Open OnDemand Interactive Access	11
3.2 Using Talapas for Batch Jobs	14
4 Code of Conduct	15

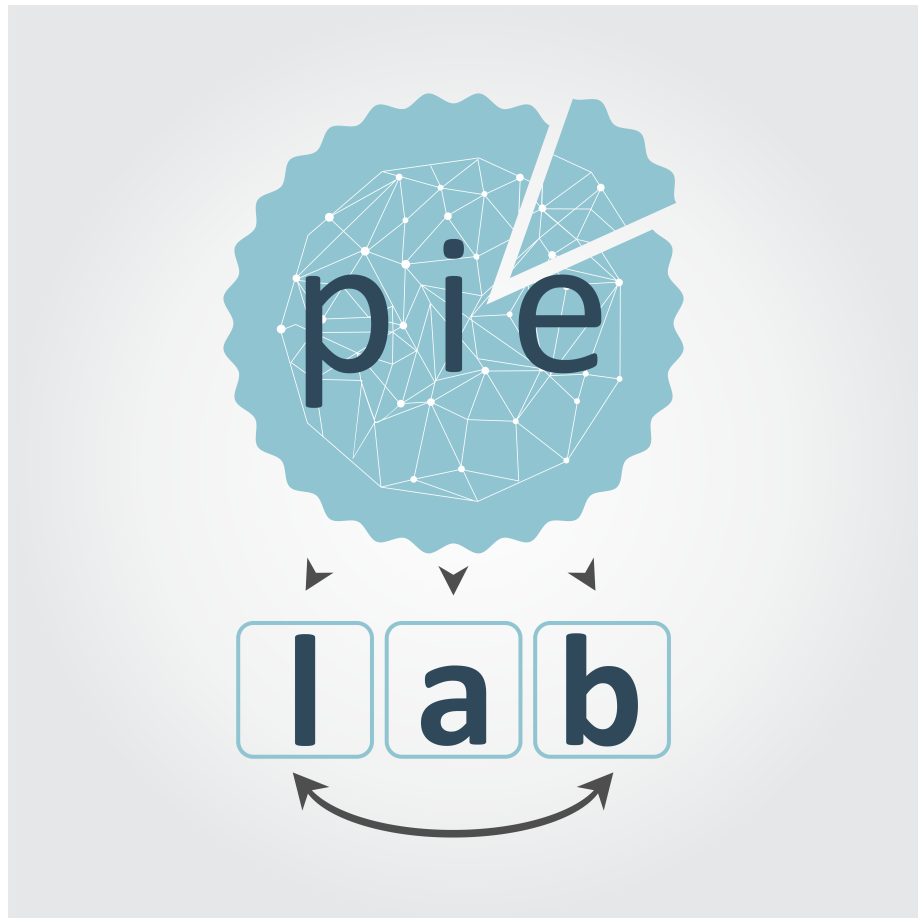


Figure 1:

Chapter 1

People

1.1 Lab

PIE lab – Straub 417 and 419

1.2 PIs

1.2.1 David M. Condon

- Office: Straub 323
- Email: dcondon@uoregon.edu

1.2.2 Sara J. Weston

- Office: Straub 325
- Email: sweston2@uoregon.edu

Chapter 2

Resources

2.1 R

2.1.1 Learning R

- Learning Statistics with R by Danielle Navarro
- R for Data Science by Hadley Wickham

2.1.2 Reference guides

- Cheatsheets

Chapter 3

Talapas Overview

Talapas is the high performance computing cluster managed by Research Advanced Computing Services (aka RACS) at the University of Oregon. Since it is possible (and maybe even best) to get started with Talapas without knowing too much about the capabilities of the system, we will simply point to the main website and the Talapas Knowledge Base for learning more.

In this document, the numbered sections below will deal with separate tasks that users in the lab might want to perform on Talapas. If you learn how to do something that is not documented below, please add it to the list, even if you are unsure that others will find it useful. Or, just point to the documentation available online.

3.1 Using Open OnDemand Interactive Access

The RACS team recommends using an incognito tab in Chrome or Firefox before logging in to Talapas (Safari will not work), as this may help to limit the potential for someone to misuse your credentials. Bear in mind that computing services can be costly so there is more at risk than your personal info (and data).

In the browser window, go to: <https://talapas-ln1.uoregon.edu>

Enter your username and password. If you don't know your credentials (or if you're not sure you have credentials), visit this page to request access. This will require the approval of a lab PI (Condon or Weston). Since there are fees associated with using the computing cluster, your request will be evaluated based on need and the availability of other options.

There is a brief overview of the OnDemand service worth reading — see here.

3.1.1 Loading files for Interactive Access

Once you have logged in, it's fairly intuitive to navigate the the dropdown menus at the top of the browser window. The 'Files' dropdown gives several options and each will open a new window. For shared projects, consider using the '/projects/pielab/shared' directory to store your files; otherwise use the directory associated with your username *within the pielab directory* (in other words, this one: '/projects/pielab/[username]'). Within the correct folder, you will want to create a new folder that is specific to each project as this is where you'll store the input and output files, just as you would if working on your own hard drive.

3.1.2 To launch an interactive session

From the OnDemand landing page, click 'Interactive Apps'/'Talapas Desktop'. This will load a form that requires entry of the PIRG name (pielab) and a few other fields. There are several things to keep in mind when filling out the form to launch the virtual desktop on Talapas. First, the 'partition' field requires selection of one of several options. For a complete list, with associated specs, limitations, and restrictions, see [here](#). For many (most?) jobs, the default 'interactive' partition should be fine. For more intense jobs, consider using 'short' (gives up to 24 hours) or 'long' (up to 2 weeks). Note that 'interactive' has a max of 4 CPUs, which equates to about 16G RAM. For reference, my souped up laptop has 32G and a standard MacBook Pro has 16G or maybe 8G if older. Analyses of SAPA data from 2017 or later will routinely kill RStudio with only 16G of RAM (necessitating use of the 'short' or 'long' partitions).

For the number of hours, a good practice is to use the lesser of the max amount of time (this depends on the partition) or a generous guess at the length of time needed. If the analysis runs for a long while and the times out, you will have no output and will have spent the funds anyway. Better to have a chance of getting what you need than running out of time. However, you should also get into the practice of deleting instances that are no longer in use, as this will save money! Otherwise, the expenses will add up until the full time allotment has expired. For the number of cores and total memory, the standard seems to be 4G of memory for each CPU — choose the CPUs accordingly (as best you can). Note however that you can also use GPUs, which seem to have more memory. See [here](#) for more info about guessing the right specifications. (For SAPA processing, I have been experimenting with 1 GPU, 160 GB, 18 cores, on the short partition for 16 hours, but this is much larger than is probably needed for most things).

3.1.3 To determine the memory usage of jobs previously run

This has to be done through an ssh interface with Talapas. Open a shell and enter this at the prompt:

```
ssh username@talapas-ln1.uoregon.edu
```

followed by your password.

To see the memory usage of recent jobs, enter:

```
sacct --format='JobID,Elapsed,MaxRSS'
```

To get more information on an individual job (recommended), enter:

```
seff <JobID>
```

That last command will return something like:

```
$ seff 9609990

Job ID: 9609990
Cluster: mycluster
User/Group: username/talapas
State: CANCELLED (exit code 0)
Nodes: 1
Cores per node: 8
CPU Utilized: 00:23:25
CPU Efficiency: 3.10% of 12:36:16 core-walltime
Job Wall-clock time: 01:34:32
Memory Utilized: 29.12 GB
Memory Efficiency: 91.00% of 32.00 GB
```

This job was cancelled after repeated efforts to compile. Note the high ratio of Memory Utilized to Memory Available.

3.1.4 To run a live session of RStudio

Once you have begun the session (i.e., submitted the form), a new page will load. From here, click the blue button that says 'Launch noVNC in New Tab'. This button may not be visible right away as you sit in the queue, waiting for the necessary resources to become available. When you launch the tab, a virtual desktop will launch with a dropdown menu in the upper left called 'Applications'. Choose 'Education'/'Talapas RStudio'. Presumably you can figure out how to manage your workflow from here.

3.1.5 Ending your interactive session

Your analyses will continue running until completed, timed-out, or stopped for some other reason (coding errors, etc.). When your analyses are done, you should return to the Open OnDemand main page and click the red ‘Delete’ button. This will stop your usage of the system and associated fees.

3.2 Using Talapas for Batch Jobs

Haven’t done this yet! But see [here](#) if you want to give it a go.

Chapter 4

Code of Conduct

Write code of conduct here.