

Lightning Recap from Gateway Week

Bridging the Bench-Machine Learning Gap

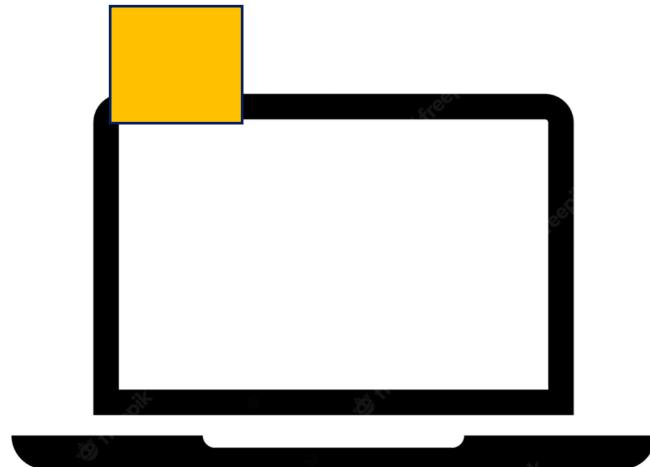
Dr. Emily A. Beck

Dr. Jake Searcy

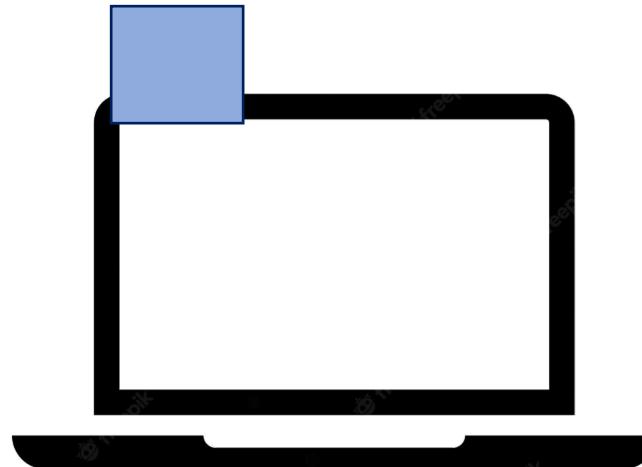
Remember from Gateway Week

We are assuming **no coding experience** so don't stress! It isn't a race to finish.

We will be using a color-coded post-it system during activities



HELP!



All done

Interrupt us!

Ask questions!

Don't be shy!

We will be working in Open OnDemand

Your settings should be saved from Gateway Week.
Interactive Apps -> R/IRkernel

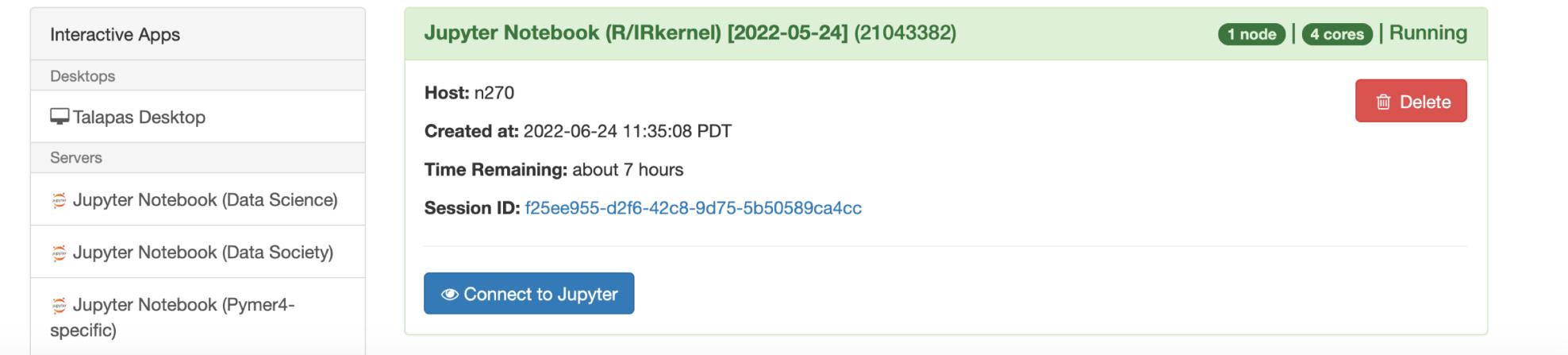


The screenshot shows the top navigation bar of the Open OnDemand interface. It includes links for 'Open OnDemand', 'Files', 'Jobs', 'Clusters', 'Interactive Apps', and a user menu with 'Help', 'Log in as ebeck8', and 'Log Out'.

- Be sure to read the [Talapas-specific documentation for Open OnDemand](#).
- This site has all of the power of your Talapas account, so please use only computers you trust when accessing it. We strongly suggest using "private browsing" or "incognito" windows--closing the window guarantees you get logged out when you're done.
- **Please let us know if you encounter any problems.**

Session was successfully created. X

Home / My Interactive Sessions



The screenshot shows the 'My Interactive Sessions' page. On the left, a sidebar lists 'Interactive Apps' under 'Desktops': 'Talapas Desktop', 'Jupyter Notebook (Data Science)', 'Jupyter Notebook (Data Society)', and 'Jupyter Notebook (Pymer4-specific)'. The main area displays a session card for a 'Jupyter Notebook (R/IRkernel)' session. The card details include: Host: n270, Created at: 2022-06-24 11:35:08 PDT, Time Remaining: about 7 hours, Session ID: f25ee955-d2f6-42c8-9d75-5b50589ca4cc, and a status of 1 node | 4 cores | Running. A red 'Delete' button is located in the top right corner of the card. At the bottom of the card is a blue 'Connect to Jupyter' button.

Remember: you can also visualize your environment on Open OnDemand

The screenshot shows the Open OnDemand web interface at talapas-ln1.uoregon.edu/pun/sys/dashboard/batch_connect/sessions. The user is logged in as ebeck8.

File Explorer: A sidebar on the left shows the user's home directory structure:

- Home Directory
- /projects/datascience/ebeck8
- /projects/maplab/ebeck8
- /projects/datascience/shared
- /projects/maplab/shared

Session Confirmation: A green notification bar at the bottom left states "Session was successfully deleted."

Interactive Session Overview: The main area displays a summary of a running session:

- Jupyter Notebook (R/IRkernel) [2022-05-24] (20766575)**
- Host:** n270
- Created at:** 2022-06-06 11:31:54 PDT
- Time Remaining:** about 7 hours
- Session ID:** 35b67b5e-3e1b-4ea2-9660-76cf0f7b031
- Status:** 1 node | 4 cores | Running
- Delete** button

Connect Options: A blue button labeled "Connect to Jupyter" is visible at the bottom of the session summary.

Remember: you can also visualize your environment on Open OnDemand

File Explorer

/home/ebeck8/

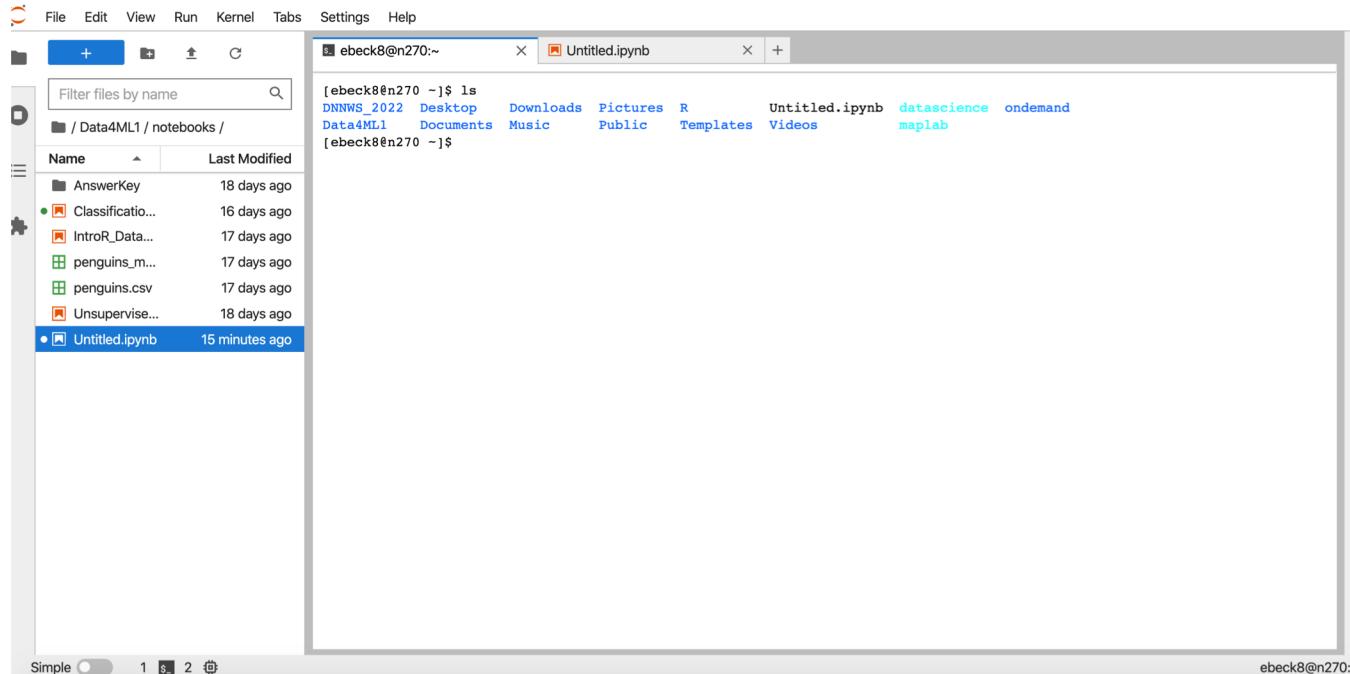
Go To... Open in Terminal New File New Dir Upload Show Dotfiles Show Owner/Mode

View Edit A-Z Rename Download Copy Paste (Un)Select All Delete

name	size	modified date
..	<dir>	01/10/2022
DNNWS_2022	<dir>	06/06/2022
Data4ML1	<dir>	05/25/2022
Desktop	<dir>	05/25/2022
Documents	<dir>	05/25/2022
Downloads	<dir>	05/25/2022
Music	<dir>	05/25/2022
Pictures	<dir>	05/25/2022
Public	<dir>	05/25/2022
R	<dir>	06/06/2022
Templates	<dir>	05/25/2022
Videos	<dir>	05/25/2022
datascience	<dir>	04/24/2020
maplab	<dir>	02/04/2021
ondemand	<dir>	01/10/2022
Untitled.ipynb	53.51kb	05/31/2022

During *Gateway Week* we learned

To navigate the terminal:



The screenshot shows a Jupyter Notebook interface. On the left, there's a file browser with a sidebar containing a search bar and a list of files. The main area shows a terminal window with the command `ls` and its output, which includes several directories and files like `DNNWS_2022`, `Desktop`, `Downloads`, etc. A specific file, `Untitled.ipynb`, is highlighted in the file list.

cd- “change directory” primary command for moving between directories

mkdir- “make directory” create a new folder or directory

pwd- “print working directory” tells you where you are!

ls- “list” lists all the files in your current directory

cp- “copy” makes a copy of a file

mv- “move” moves a files

During *Gateway Week* we learned

To view large files in the terminal

more	head	tail	cat
view a text file one screen full at a time	view the top 15 lines of a file	view the last 15 lines of a file	spit the whole file at once
space-bar: scroll q: quit	-n num controls the number of lines	-n num controls the number of lines	

During *Gateway Week* we learned
To search large files using *grep*

grep is the simplest of these three commands and differs from awk and sed in that it cannot be used to add, modify, or delete text.

Similar to the **search** function ctrl-f that many are familiar with.

It can be used to **print** or **count** the desired text using various flags.

Useful grep flags:

grep -c (count)

grep -v (invert/everything but)

grep -n (show line numbers)

grep -i (ignore case)

How to use:

function -flag term <file>

grep -c ATGC sequence.fasta

Count the number of times “ATGC” appears in my sequence file

During *Gateway Week* we learned

(1) To search and replace in large files using *sed*

`sed 's/pattern/replace/' [input_file] > output_file`

(2) To subset large files using *sed*

`sed -n '1p;/pattern/p' [input_file] > output_file`

This prints the first row in a file then perform the search and print function

WARNING: *sed* will overwrite your existing file!

Using `> output_file` writes a new file with the changes I made

During *Gateway Week* we learned
To subset large files using *awk*

Example:

If I have a large tsv file and I only want columns 1, 3, and 7 I can use awk

```
awk '{print $1, $3, $7}' large_file.tsv > subset.tsv
```

If we are working with other file types (for example csv) we need to note what the field separator is:

–F fs where fs is the field separator

Csv example:

```
-F ,
```

During *Gateway Week* we learned

Lots of tips and tricks!

Tip #1: Save yourself a headache and avoid special characters

Using underscores in place of spaces makes things much easier

Tip #2: Are you typing? You are doing it wrong!

Tab-completion:

- Tab once to complete uniquely
- Tab twice to see all possible completions

Up-arrow:

- Find your most recently used commands

Tip #3: Moving groups of files using *

Work with all files with a common name simultaneously using *

`mv RawReads* /working_directory` (moves all my files beginning with “RawReads”)

Want to spend more time going over these useful Linux command?

Come to office hours and we can work through examples together!