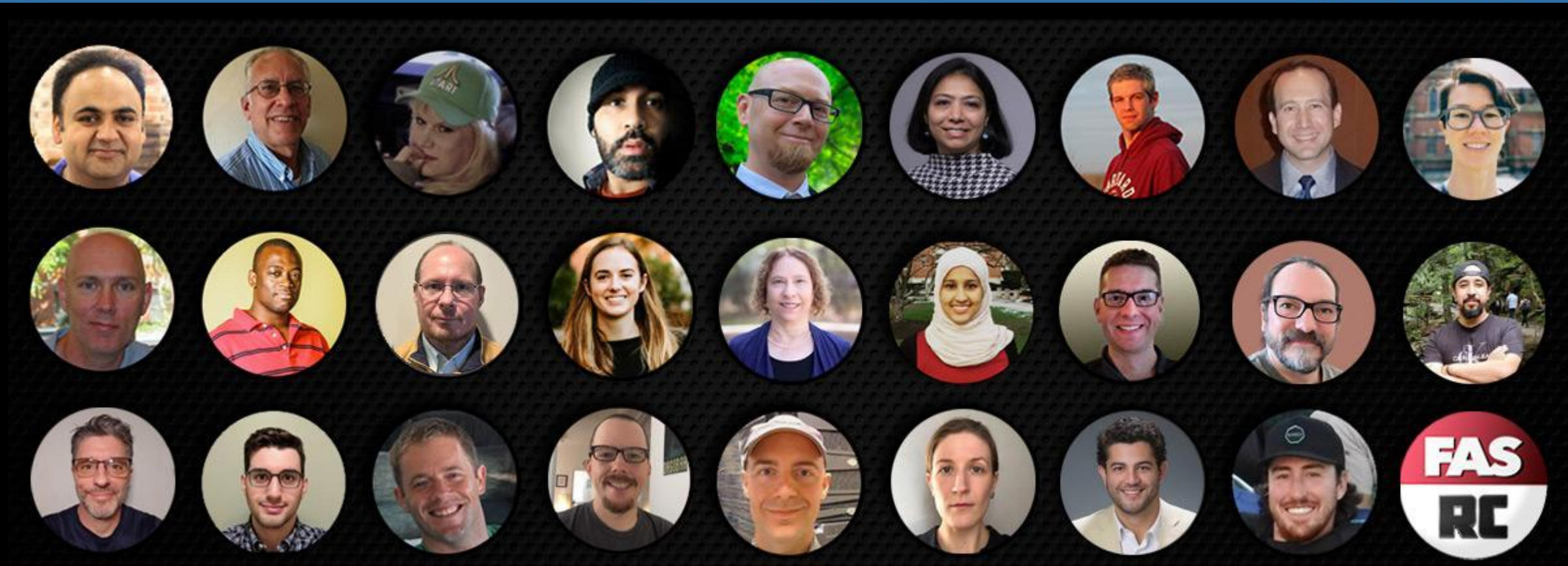
IQSS
The Institute for Quantitative Social Science

Getting Started on the FASRC Clusters with Open OnDemand



Faculty of Arts and Sciences
Research Computing

Learning objectives

- What is Open OnDemand (OOD)?
- Accessing OOD
- Launching apps
- RStudio Server
- Jupyter Notebook
  - Create conda environment (i.e., jupyter kernel)
- Files tab
- Jobs tab
- Remote Desktop
- FASSE proxy

Some definitions

- **OOD**: Open On Demand
- **Cluster**: large group of servers with lots of memory and processors
- **Cannon**: cluster that handles level 2 data. Named after the 19th century Harvard astronomer Annie Jump Cannon.
- **FASSE**: cluster that handles level 3 data. FAS Secure Enclave.

Glossary of these terms: docs.rc.fas.harvard.edu/kb/glossary

What is Open OnDemand (OOD)?

- Open-source web portal to access clusters
- Web-based
  - Uses modern browser like Google Chrome, Mozilla Firefox, or Microsoft Edge.
    - Safari does not support all of OOD's features
  - No software other than a browser needs be installed on your local laptop/desktop
- Easy to learn and use
- Very similar to desktop applications
- The easiest way to run graphical applications remotely on a cluster

OPEN OnDemand

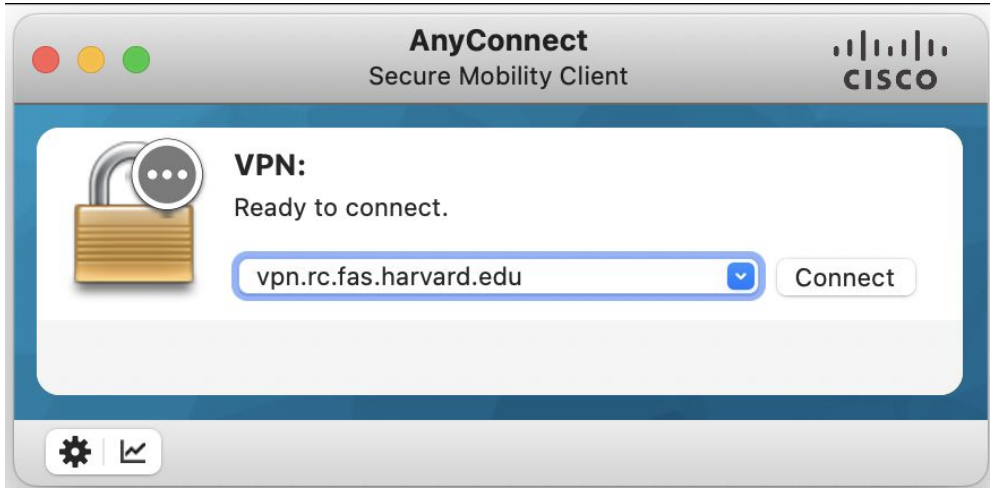
How to access OOD on FASRC Clusters

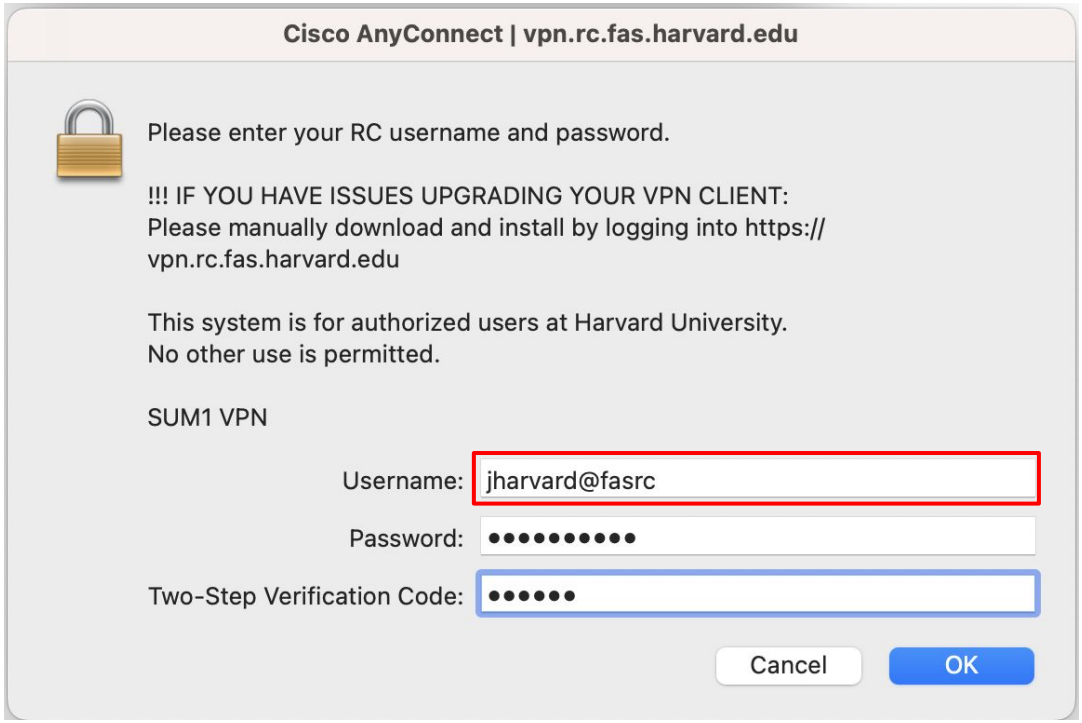
1. Get an account
    - You can choose a username. The default is first initial, last name. I'll use jharvard as an example
    - This is NOT necessarily the same as your HarvardKey username.
2. Log onto the FASRC VPN. This is NOT the generic Harvard VPN.
    - vpn.rc.fas.harvard.edu
    - username is
        - jharvard@fasrc (Cannon)
        - jharvard@fasse (FASSE)
3. Log into Open On Demand for your cluster
    - username is just jharvard, nothing else

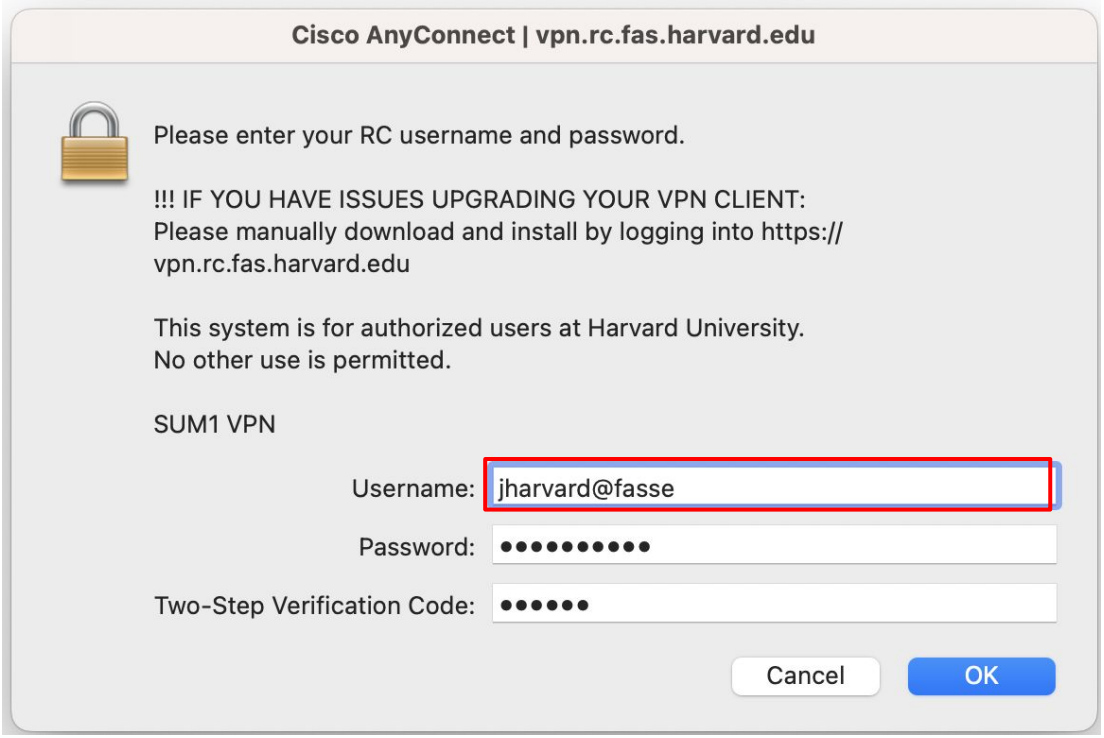
      Cannon link: https://rcood.rc.fas.harvard.edu

      FASSE link: https://fasseood.rc.fas.harvard.edu

Connecting to VPN







Signing in to the OOD Dashboard

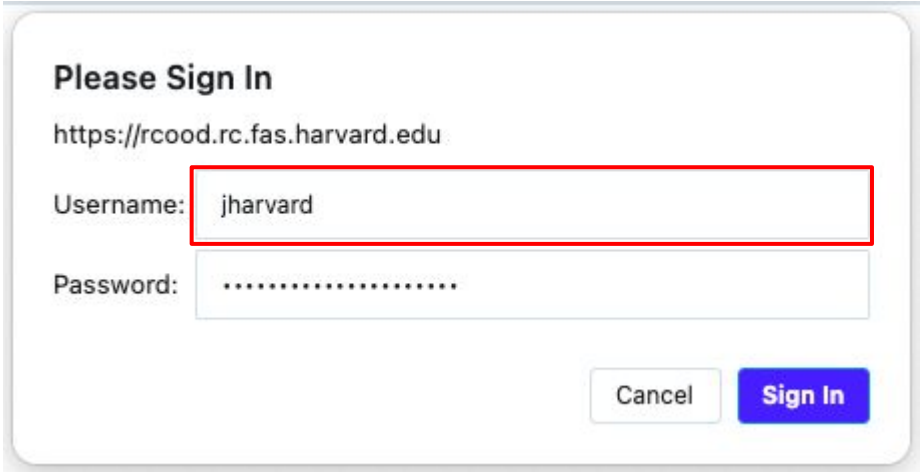
When you load the site

https://rcood.rc.fas.harvard.edu

or

https://fasseood.rc.fas.harvard.edu

You will be prompted to log in.

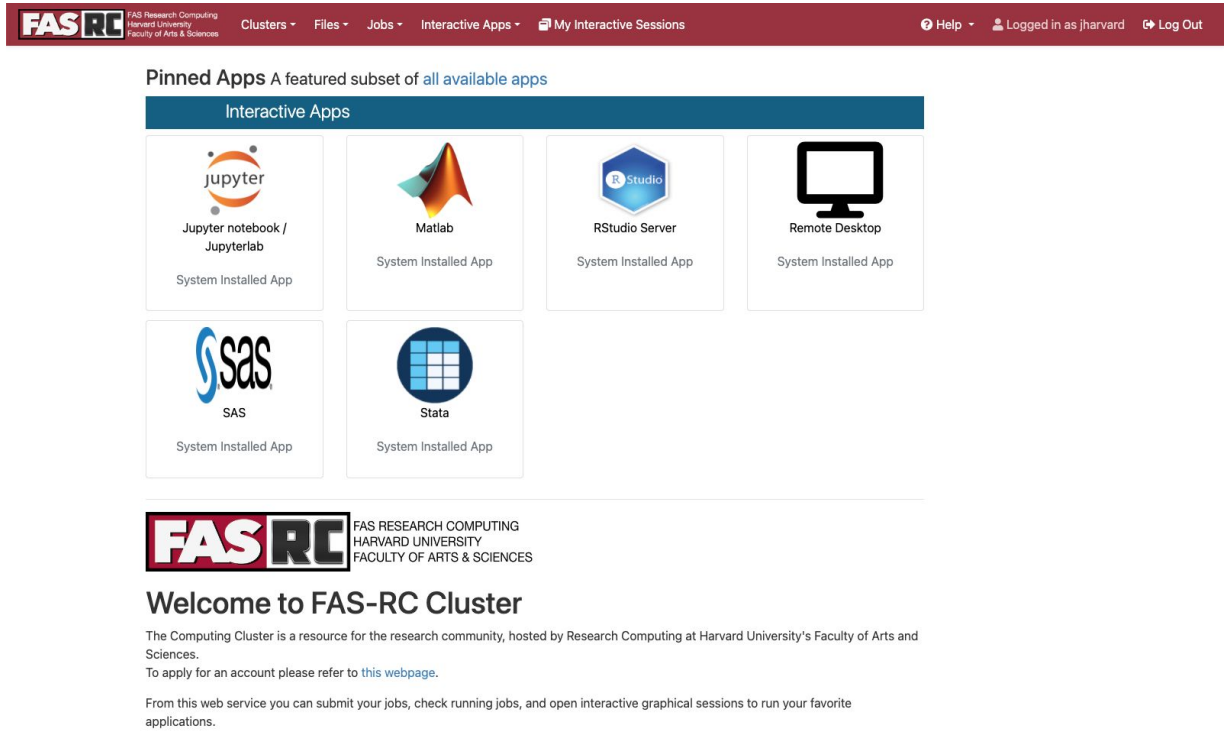


Unlike when you log in to the VPN, on the website you need to use

ONLY your username, NOT username@cluster

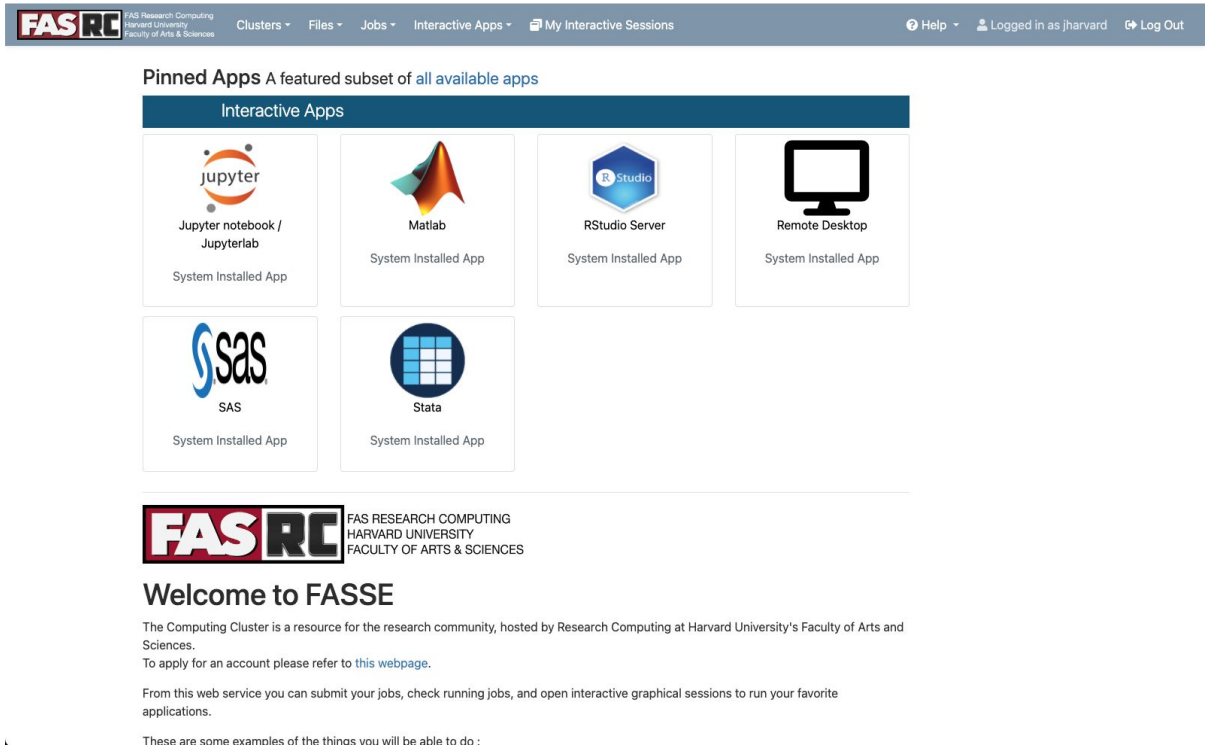
OOD dashboard on Cannon and FASSE

Cannon



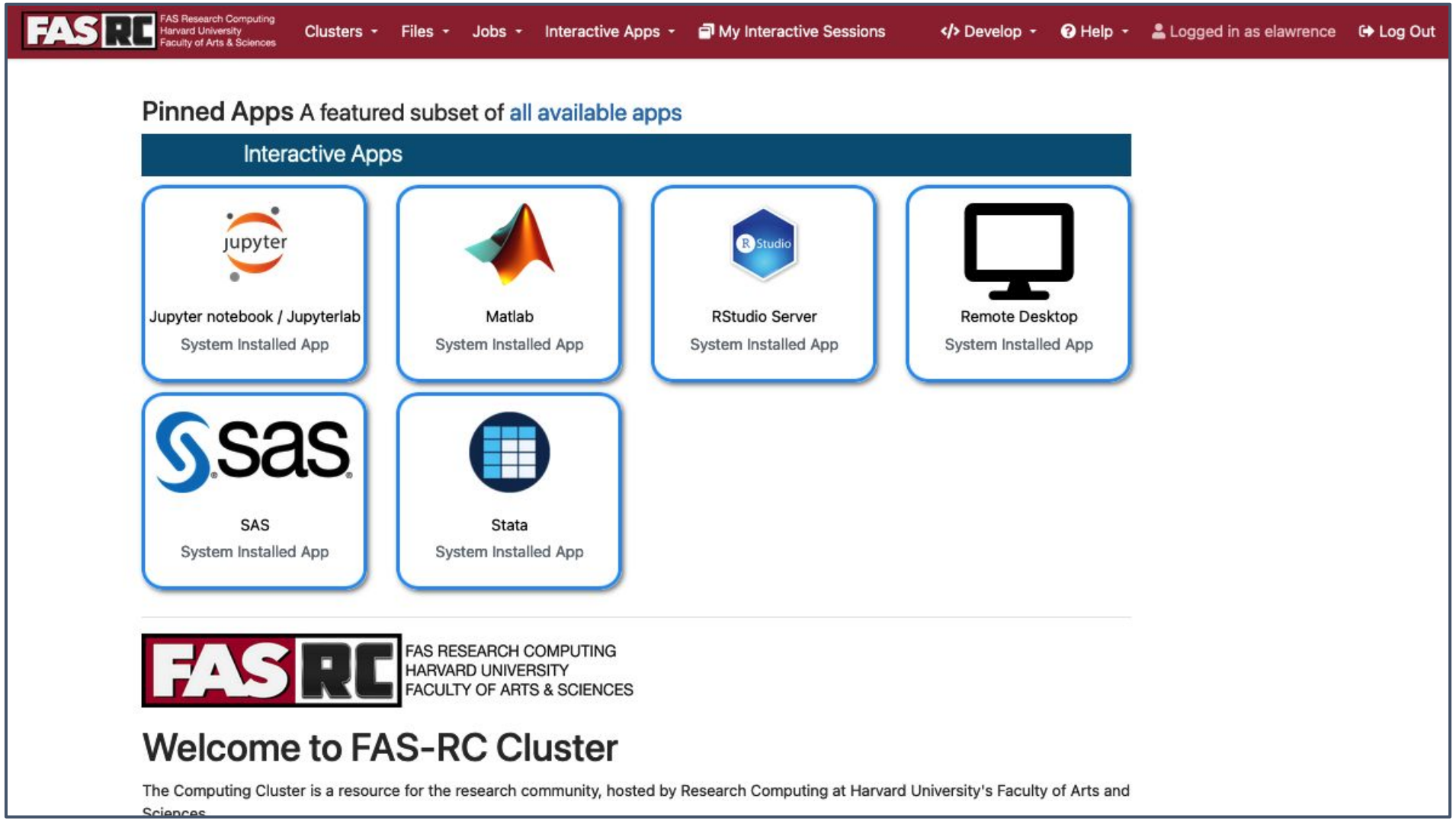
https://rcood.rc.fas.harvard.edu

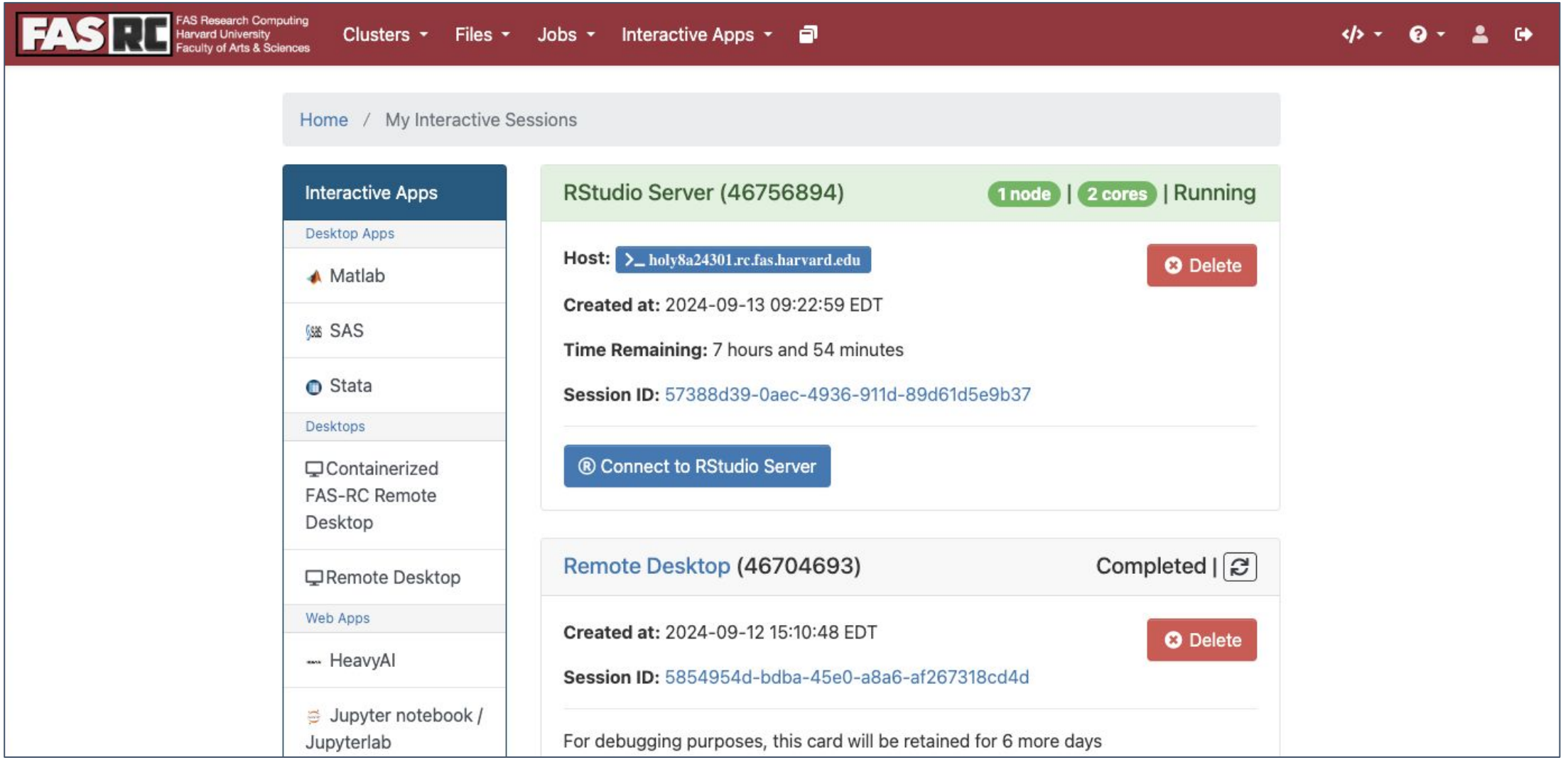
FASSE



https://fasseood.rc.fas.harvard.edu

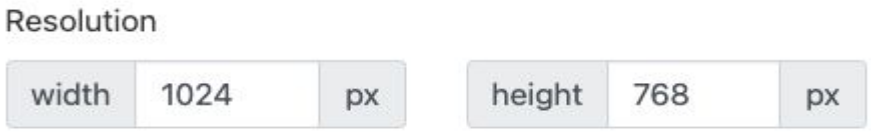
Launching an app from the Dashboard




FAS RC
FAS Research Computing
Harvard University
Faculty of Arts & Sciences
Clusters
Files
Jobs
Interactive Apps
Home / My Interactive Sessions
Interactive Apps
Desktop Apps
Matlab
SAS
Stata
Desktops
Containerized FAS-RC Remote Desktop
Remote Desktop
Web Apps
HeavyAI
Jupyter notebook / Jupyterlab
RStudio Server (46756894)
1 node | 2 cores | Running
Host: holy8a24301.rc.fas.harvard.edu
Delete
Created at: 2024-09-13 09:22:59 EDT
Time Remaining: 7 hours and 54 minutes
Session ID: 57388d39-0aec-4936-911d-89d61d5e9b37
Connect to RStudio Server
Remote Desktop (46704693)
Completed |
Created at: 2024-09-12 15:10:48 EDT
Delete
Session ID: 5854954d-bdba-45e0-a8a6-af267318cd4d
For debugging purposes, this card will be retained for 6 more days

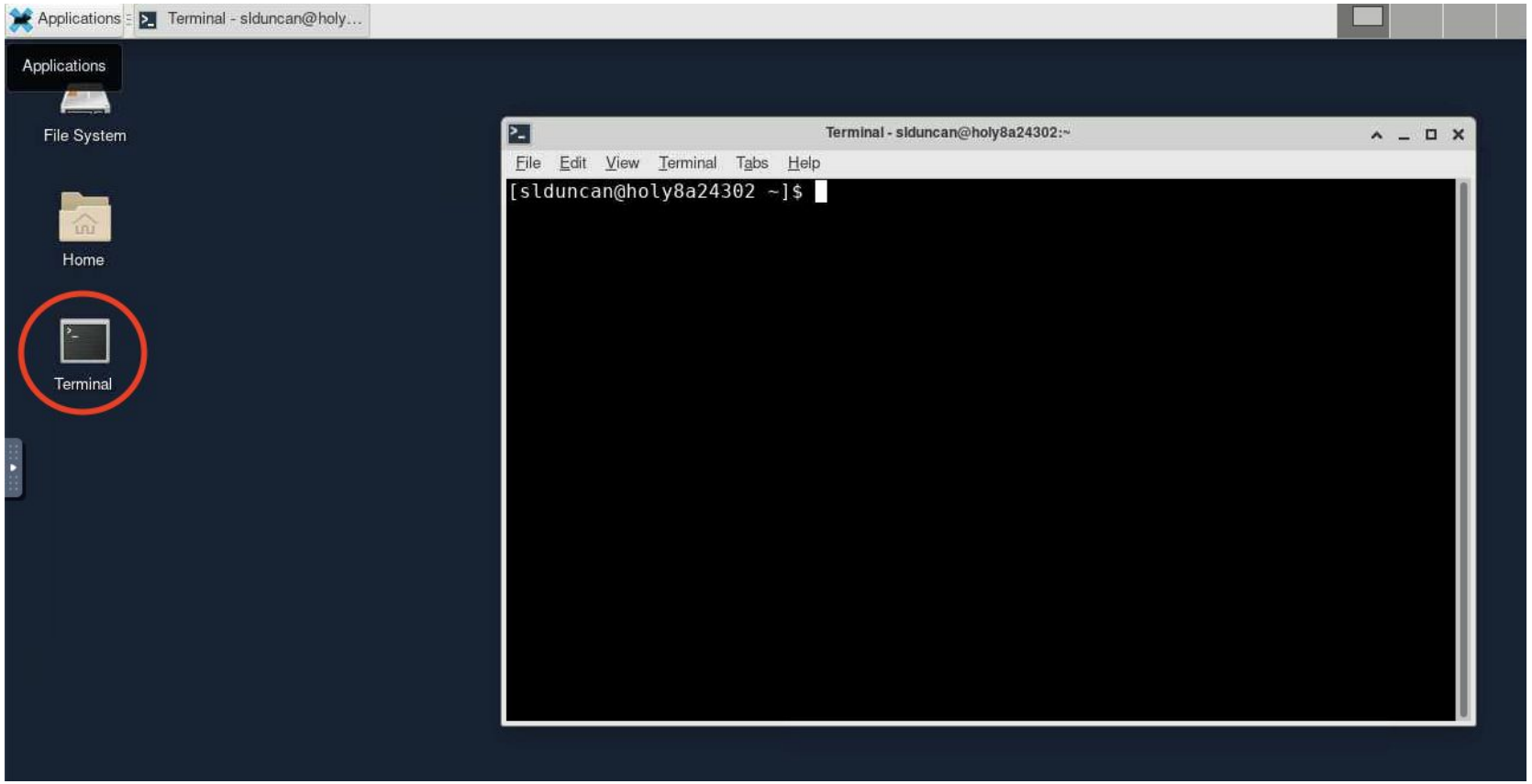
Remote Desktop

- Not as necessary as it used to be
- **For running long Jupyter Notebook sessions**
- Terminal
- Can also be used to open multiple applications in a single window
- Choose the defaults for resolution



Starting a terminal

- Double click on the icon that looks like a computer screen
- From there you can type in any commands you want to



Running Jupyter Notebook in Remote Desktop

- Very important: Jupyter Notebook will not continue to run if you close the Jupyter notebook page! The cell that is running will lose the data and output files will not be written

1. Solution: run Remote Desktop app and launch Jupyter Notebook from within Remote Desktop
2. Documentation: https://docs.rc.fas.harvard.edu/kb/ood-remote-desktop-how-to-open-software/#Jupyter_Notebook

```
#Jupyter Notebook
[jharvard@holy7c02111 ~]$ module load python
[jharvard@holy7c02111 ~]$ mamba activate OOD_env
[jharvard@holy7c02111 ~]$ jupyter notebook
```

Creating your own kernel for Jupyter Notebook

- You can create your own kernels, but some command line needed
- Note: a kernel is the same as a conda, python, or mamba environment

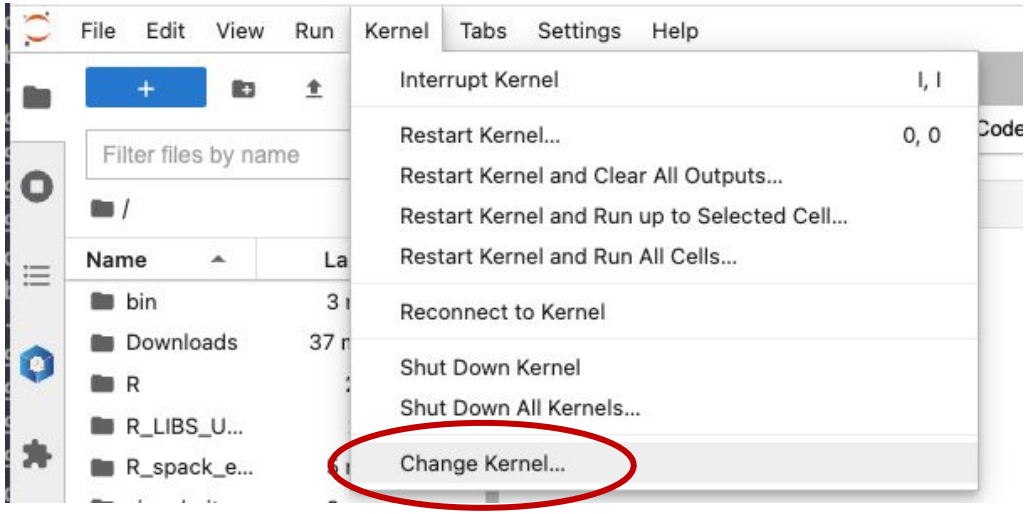
1. Open a terminal in the ”Remote Desktop” app
   ⇒Don’t create mamba environments inside Jupyter Notebook/Lab!
2. Create mamba environment and install package `ipykernel` and `nb_conda_kernels`
3.

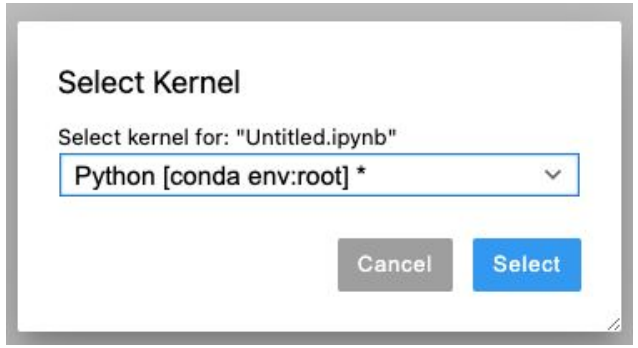
```
[jharvard@holy7c02111 ~]$ module load python
[jharvard@holy7c02111 ~]$ mamba create -n OOD_env python=3.11 pip wheel numpy
[jharvard@holy7c02111 ~]$ mamba activate OOD_env
(OOD_env)[jharvard@holy7c02111 ~]$ mamba install ipykernel nb_conda_kernels
```

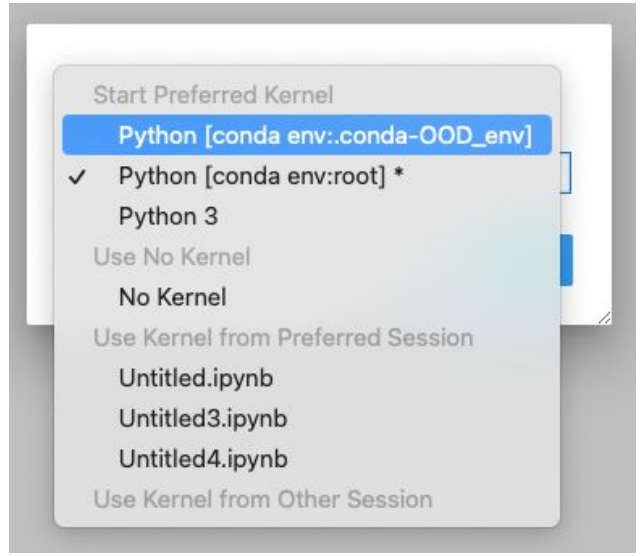
https://docs.rc.fas.harvard.edu/kb/python-package-installation/#Use_mamba_environment_in_Jupyter_Notebooks

Using your new environment in Jupyter Notebook

3. Launch **new** Jupyter Notebook session (existing session will not work!)
4. Select newly created mamba environment as the kernel
   a. Open a notebook
   b. On the top menu, click Kernel -> Select Kernel -> Click on OOD_env







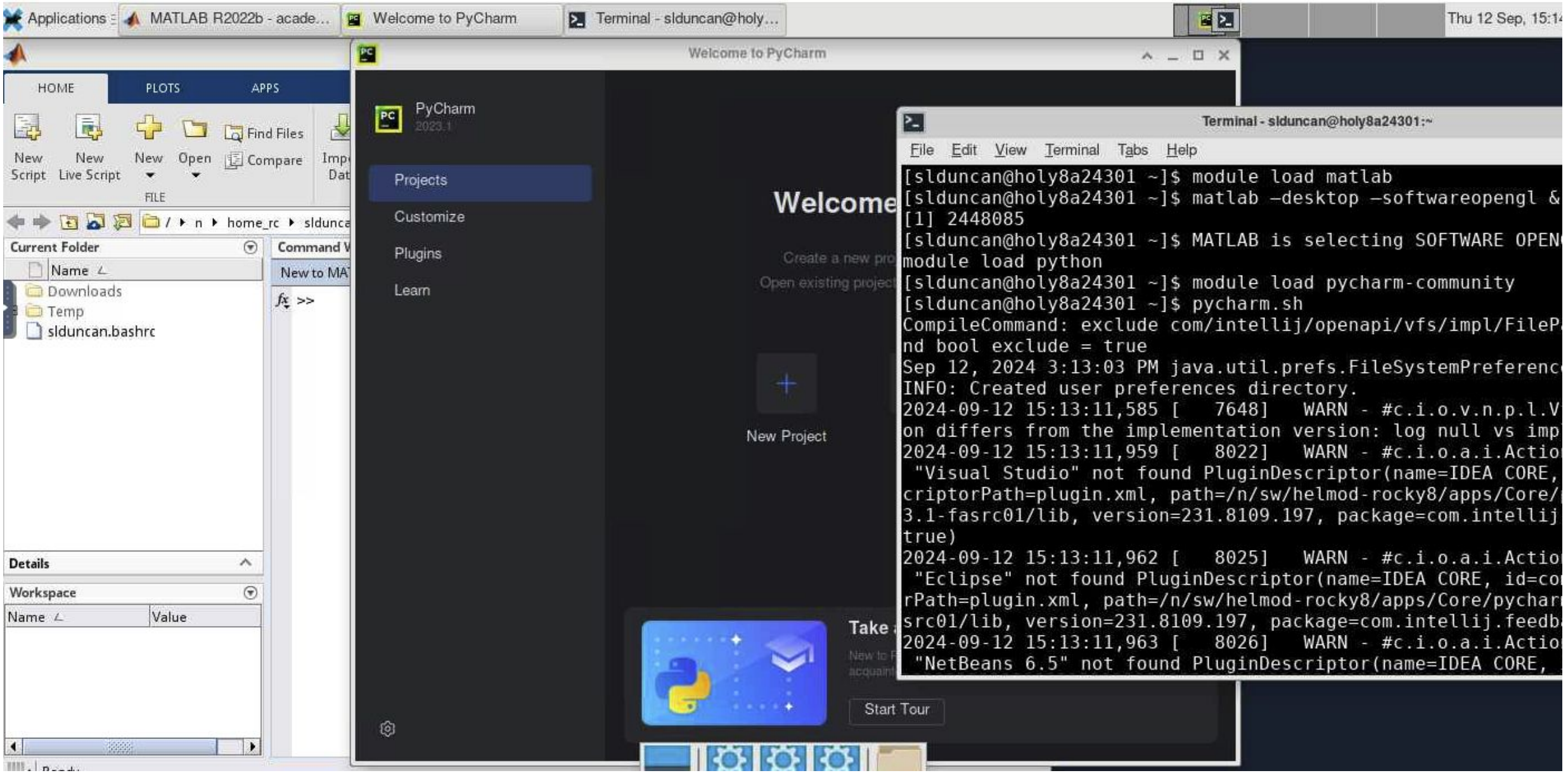
Opening Multiple Applications in Remote Desktop

Documentation: https://docs.rc.fas.harvard.edu/kb/ood-remote-desktop-how-to-open-software/

- It can be used to launch most GUI applications
  1. Load module
  2. Set environmental variables (if needed)
  3. Launch software
- You can have multiple applications open

```
# Matlab
[jharvard@holy7c02111 ~]$ module load matlab
[jharvard@holy7c02111 ~]$ matlab -desktop -softwareopengl &

#PyCharm
[jharvard@holy7c02111 ~]$ module load python
[jharvard@holy7c02111 ~]$ module load pycharm-community
[jharvard@holy7c02111 ~]$ pycharm.sh
```


Applications
MATLAB R2022b - acade...
Welcome to PyCharm
Terminal - slduncan@holy...
Thu 12 Sep, 15:14
HOME
PLOTS
APPS
New Script
New Live Script
New
Open
Find Files
Compare
FILE
Current Folder
Name
Downloads
Temp
slduncan.bashrc
Details
Workspace
Name
Value
Welcome to PyCharm
PyCharm
2023.1
Projects
Customize
Plugins
Learn
Welcome
New Project
Start Tour
Terminal - slduncan@holy8a24301:~
File Edit View Terminal Tabs Help
[slduncan@holy8a24301 ~]$ module load matlab
[slduncan@holy8a24301 ~]$ matlab -desktop -softwareopengl &
[1] 2448085
[slduncan@holy8a24301 ~]$ MATLAB is selecting SOFTWARE OPEN
module load python
[slduncan@holy8a24301 ~]$ module load pycharm-community
[slduncan@holy8a24301 ~]$ pycharm.sh
CompileCommand: exclude com/intellij/openapi/vfs/impl/FileP
nd bool exclude = true
Sep 12, 2024 3:13:03 PM java.util.prefs.FileSystemPreferenc
INFO: Created user preferences directory.
2024-09-12 15:13:11,585 [7648] WARN - #c.i.o.v.n.p.l.V
on differs from the implementation version: log null vs imp
2024-09-12 15:13:11,959 [8022] WARN - #c.i.o.a.i.Actio
"Visual Studio" not found PluginDescriptor(name=IDEA CORE,
criptorPath=plugin.xml, path=/n/sw/helmod-rocky8/apps/Core/
3.1-fasrc01/lib, version=231.8109.197, package=com.intellij
true)
2024-09-12 15:13:11,962 [8025] WARN - #c.i.o.a.i.Actio
"Eclipse" not found PluginDescriptor(name=IDEA CORE, id=co
rPath=plugin.xml, path=/n/sw/helmod-rocky8/apps/Core/pychar
src01/lib, version=231.8109.197, package=com.intellij.feedb
2024-09-12 15:13:11,963 [8026] WARN - #c.i.o.a.i.Actio
"NetBeans 6.5" not found PluginDescriptor(name=IDEA CORE,

Inactivity lock out

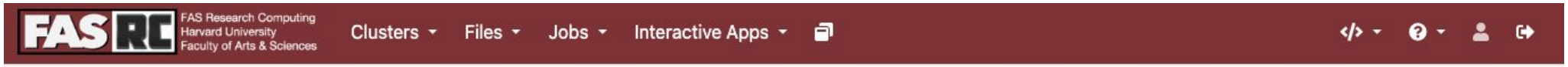
- It may lock out due to inactivity
- Use your FASRC password to unlock

Filling out a form to launch an app

- Request the resources that you need
  (If you don't know for a first trial run, use similar resources as your laptop/desktop)
  - Partition (Name): depends on Cannon vs FASSE
  - Memory (RAM): amount of memory in GB
  - Number of cores: recommended at least 2
  - Number of GPUs: if >= 1, make sure you **select** a gpu partition
  - Allocated time: time you would like your session to run.

  } the minimum and/or maximum values of each field depends on the selected partition

  - Email for status notification: to know when job starts, ends
  - Reservation: if you have a special reservation (this requires approval from FASRC)
  - Account: use this if you have more than one PI_lab affiliation



Home / My Interactive Sessions / RStudio Server

Interactive Apps

Desktop Apps

- Matlab
- SAS
- Stata

Desktops

- KNIME
- Containerized FAS-RC Remote Desktop
- Remote Desktop

Web Apps

- HeavyAI
- Jupyter notebook / Jupyterlab
- RStudio Server

RStudio Server

This app will launch an RStudio Server instance on a FAS RC compute node. This app provides a common software environment for FAS Informatics workshops and general-purpose single-node RStudio Server jobs.

See RStudio Server OOD app for more information.

- User-installed R libraries will be installed in *~/R/ifxrstudio/<IMAGE_TAG>*.

Partition

test

sbatch -p, --partition=<partition_names>

Slurm partition name (e.g., **shared**), or comma-separated list of partition names (e.g., **shared,test**)

Memory Allocation in GB

32

sbatch --mem=<size>G

Number of CPUs to allocate

2

Slurm Partitions

Partition is the term that Slurm uses for queues. Partitions can be thought of as a set of resources and parameters around their use (See also: Convenient Slurm Commands). You can find out what partitions you have access to using the `spart` command. FASSE has different partitions than Cannon.

Note: In the case where no resources have been requested explicitly, default resources that get allocated to a job on Cannon or FASSE are, serial_requeue for the partition, 10 minutes for the time, 1 core, and 100 MB for the memory.

Partition	Nodes	Cores per Node	CPU Core Types	Mem per Node (GB)	Time Limit	Max Jobs	Max Cores	GPU Capable?	/scratch size (GB)
sapphire	186	112	Intel "Sapphire Rapids"	990	3 days	none	none	No	396
shared	310	48	Intel "Cascade Lake"	184	3 days	none	none	No	68
bigmem	4	112	Intel "Sapphire Rapids"	1988	3 days	none	none	No	396
bigmem_intermediate	3	64	Intel "Ice Lake"	2000	14 days	none	none	No	396
gpu	36	64	Intel "Ice Lake"	990	3 days	none	none	Yes (4 A100/node)	396
intermediate	12	112	Intel "Sapphire Rapids"	990	14 days	none	none	No	396
unrestricted	8	48	Intel "Cascade Lake"	184	none	none	none	No	68
test	18	112	Intel "Sapphire Rapids"	990	12 Hours	5	112	No	396
gpu_test	14	64	Intel	487	12	5	64	Yes (8 A100 MIG	172

Allocated Time (expressed in MM , or HH:MM:SS , or DD-HH:MM)

02-08:00

sbatch -t, --time=<time>

R version

R 4.3.3 (Bioconductor 3.18, RStudio 2023.09.1)

This defines the version of R/Bioconductor you want to use.

☐ I would like to receive an email when the session starts

(optional) email address for job status notification

(optional) Slurm Reservation

sbatch --reservation=<name>

Leave blank if you do not have a Slurm reservation

☑ Start rstudio with a new configuration

Checking this box will start rstudio with a fresh configuration.

This is useful if you need to run different instances at the same time with different configurations.
Leave the box unchecked if you want rstudio to start with the default configuration you have in your ~/.rstudio

(optional) Slurm Account

sbatch -A, --account=<account>

Leave blank if you do not have multiple Slurm accounts

Additional Slurm options

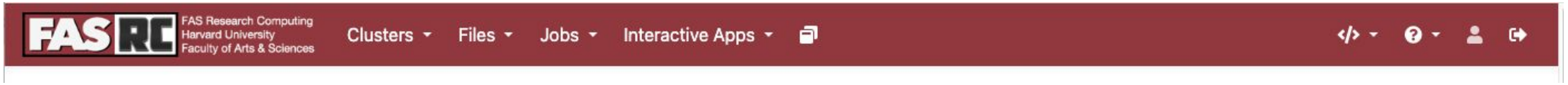
additional slurm options (**long format option only**)
Example : --constraint=intel --exclusive

Please make sure the options you select are compatible with other fields in the form.
Also, please make sure you use **long format option only**
(e.g. --nodelist=holy7c24502 instead of -w holy7c24502)

Launch

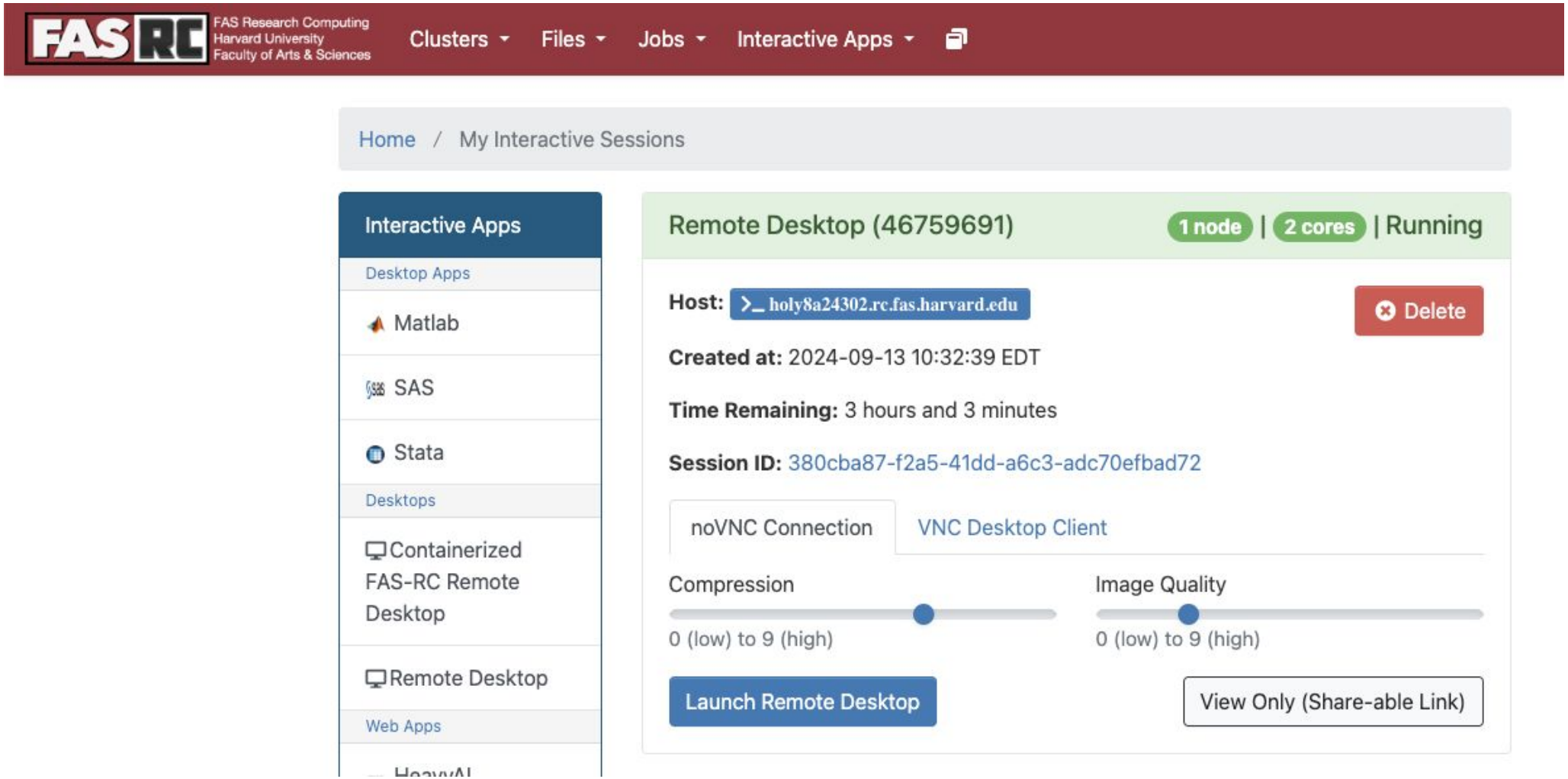
* The RStudio Server session data for this session can be accessed under the data root directory.

Menu bar



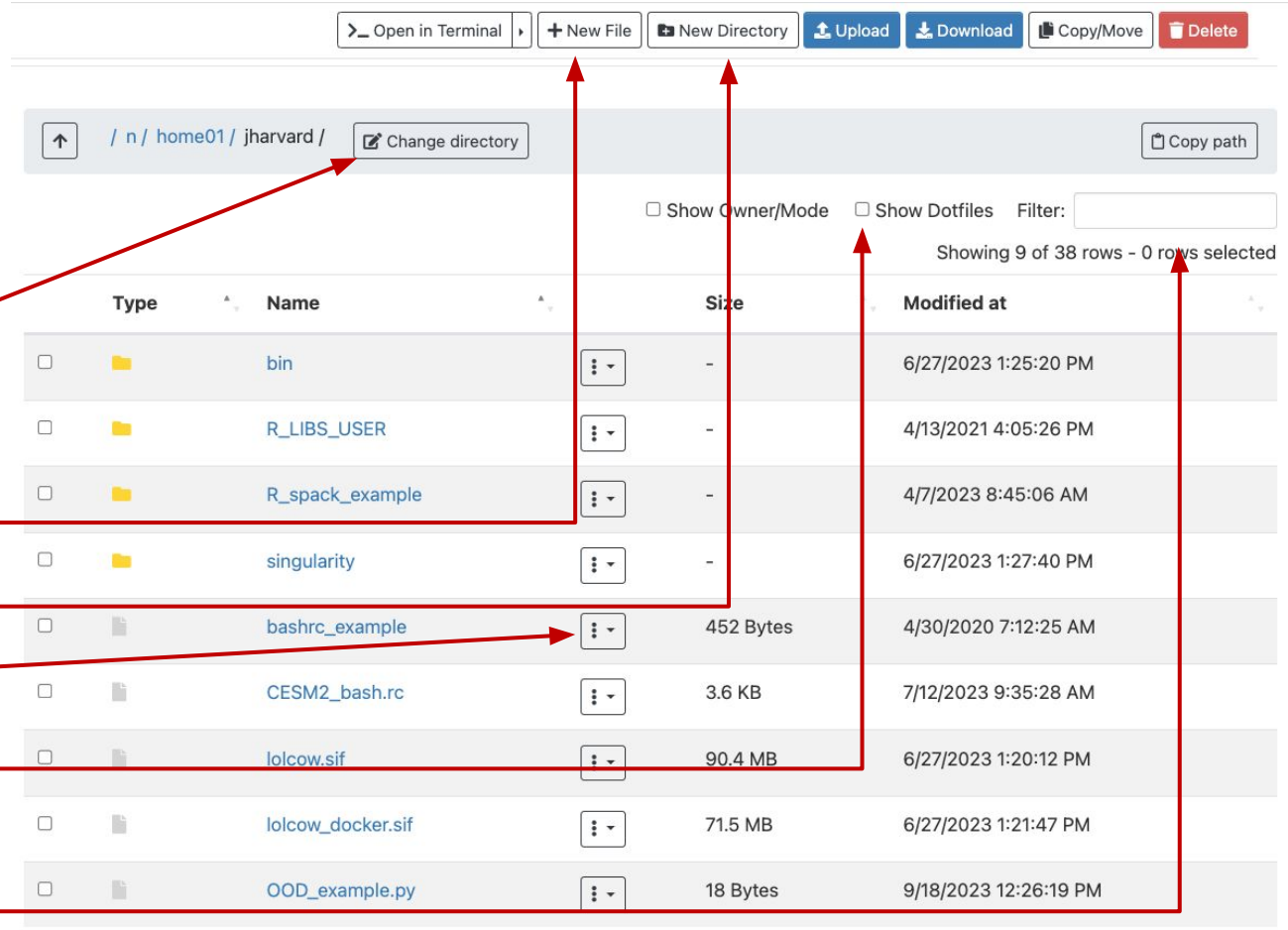
- My Interactive Sessions
- Files
- Jobs
- Interactive apps

My Interactive Sessions



Files tab

- Default options: home directory and holyscratch
- Click on “Change directory” to go to a lab share at /n/holylabs/LABS
- Create new file
- Create new directory (i.e., folder)
- Click on three dots for options
- Check “Show Dotfiles” to see hidden files
- Filter to find files or directories in current directory



Active Jobs and their details in the Jobs tab

Active Jobs

Show 50 entries Filter:

	ID	Name	User	Account	Time Used	Queue	Status	Cluster	Actions
>	2469887	.fasrcood/sys/dashboard/sys/RemoteDesktop	jharvard	jharvard_lab	01:35:49	serial_requeue	Completed	Cannon Cluster	
v	2474168	.fasrcood/sys/dashboard/sys/Jupyter	jharvard	jharvard_lab	00:09:37	test	Running	Cannon Cluster	

Running .fasrcood/sys/dashboard/sys/Jupyter 2474168

Cluster	Cannon Cluster
Job Id	2474168
Job Name	.fasrcood/sys/dashboard/sys/Jupyter
User	jharvard
Account	jharvard_lab
Partition	test
State	RUNNING
Reason	None
Total Nodes	1
Node List	holy7c02412
Total CPUs	2
Time Limit	2:00:00
Time Used	9:39
Memory	8192M

A failed job

Matlab (2474322) Undetermined

Created at: 2023-09-18 15:28:06 EDT

Delete

Time Requested: 1 hour

Session ID: 0847d7b8-1d3f-4a61-877d-582272b74ec0

Your session has entered a bad state. Feel free to contact support for further information.

Active Jobs

Show 50 entries Filter:

	ID	Name	User	Account	Time Used	Queue	Status	Cluster	Actions
>	2469887	.fasrcood/sys/dashboard/sys/RemoteDesktop	jharvard	jharvard_lab	01:35:49	serial_requeue	Completed	Cannon Cluster	
>	2474322	.fasrcood/sys/dashboard/sys/Matlab	jharvard	jharvard_lab	00:02:27	test	Undetermined	Cannon Cluster	
>	2474168	.fasrcood/sys/dashboard/sys/Jupyter	jharvard	jharvard_lab	00:15:45	test	Running	Cannon Cluster	

Diagnosing a failed job

2474322	.fasrcood/sys/dashboard/sys/Matlab	jharvard	jharvard_lab	00:02:27	test	Undetermined	Cannon Cluster	

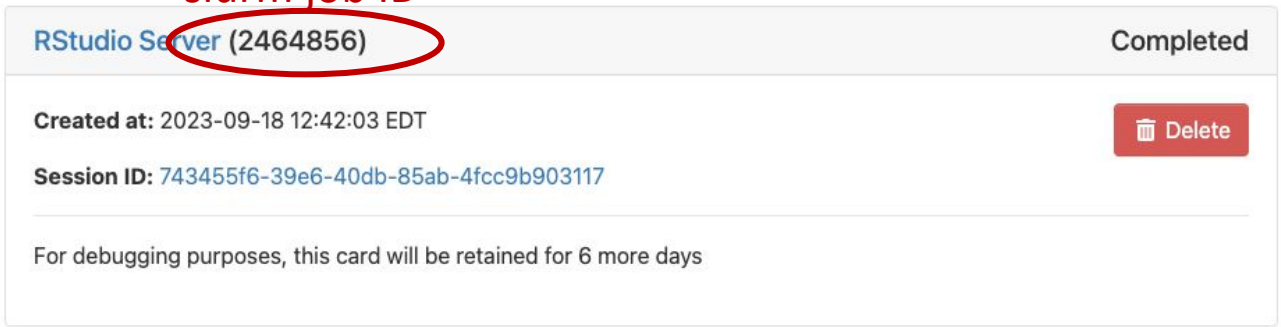
Undetermined **.fasrcood/sys/dashboard/sys/Matlab** **2474322**

Cluster	Cannon Cluster
Job Id	2474322
Job Name	.fasrcood/sys/dashboard/sys/Matlab
User	jharvard
Account	jharvard_lab
Partition	test
State	OUT_OF_MEMORY
Reason	OutOfMemory
Total Nodes	1
Total CPUs	2
Time Limit	1:00:00
Time Used	2:27
Memory	4096M

Detailed job information from the command line

If job no longer appears on “Active Jobs”, check job status from command line with slurm job ID

slurm job ID




```
[jharvard@boslogin01 ~]$ sacct -j 2464856
JobID           JobName  Partition    Account  AllocCPUS      State ExitCode
------------ ---------- ---------- ---------- ---------- ---------- --------
2464856      .fasrcood+       test jharvard_+          2    TIMEOUT      0:0
2464856.bat+      batch            jharvard_+          2  CANCELLED     0:15
2464856.ext+     extern            jharvard_+          2  COMPLETED      0:0

[jharvard@holy7c02111 ~]$ sacct -j 2471535
JobID           JobName  Partition    Account  AllocCPUS      State ExitCode
------------ ---------- ---------- ---------- ---------- ---------- --------
2471535      .fasrcood+       test jharvard_+          2 OUT_OF_ME+    0:125
2471535.bat+      batch            jharvard_+          2 OUT_OF_ME+    0:125
2471535.ext+     extern            jharvard_+          2  COMPLETED      0:0
```

Closing running OOD windows/tabs

- In most OOD apps, you can close the browser tab while the code is running, and the code will continue to run on the background
- Jupyter Notebook will not! The cell that is running will lose the data and output files will not be written
  - Solution: run Remote Desktop app and launch Jupyter Notebook from within Remote Desktop
  - Documentation: https://docs.rc.fas.harvard.edu/kb/ood-remote-desktop-how-to-open-software/#Jupyter_Notebook
- Because closing tabs does not end the application, it is important to cancel your job when you are done using it. Otherwise it will be charged to your lab's fairshare.

FASSE proxy

Documentation: https://docs.rc.fas.harvard.edu/kb/proxy-settings/

- You may need to set FASSE proxy on
  - RStudio server if you are unable to reach CRAN and download R packages
  - Stata if you are unable to load libraries via http
  - Firefox (web browsing)
  - Jupyter Notebook
  - Access Github
  - (Basically, anything outside of FASSE)

Quickstart Guides for using the FASRC Clusters

- Cannon Quickstart Guide
  - https://docs.rc.fas.harvard.edu/kb/iqss-cannon-quickstart-guide
- FASSE Quickstart Guide
  - https://docs.rc.fas.harvard.edu/kb/iqss-fasse-quickstart-guide

- Quickstart guides have more than just information on OOD
  - how to do text based access
  - office hours, training, tickets

FASRC documentation

- FASRC docs: https://docs.rc.fas.harvard.edu/
- GitHub User_codes: https://github.com/fasrc/User_Codes/
- Getting help
  - Office hours: https://www.rc.fas.harvard.edu/training/office-hours/
  - Ticket
    - Email: rchelp@rc.fas.harvard.edu to open a ticket with us

Additional trainings

Training calendar: https://www.rc.fas.harvard.edu/upcoming-training/

Training materials: https://docs.rc.fas.harvard.edu/kb/training-materials/

Training session evaluation

Please, fill out our training session evaluation. Your feedback is essential for us to improve our trainings!!

https://tinyurl.com/FASRCworkshop

Thank you :)

FAS Research Computing