



GETTING STARTED WITH THE ARKANSAS RESEARCH PLATFORM: A FRIENDLY TUTORIAL

PRATEEK VERMA, XINTAO WU

7/25/2024

ACKNOWLEDGMENT

Hadi Salman
Doctoral Student



Minh-Hao Van
Doctoral Student



Xintao Wu, PhD
Core Director



David Chaffin, PhD
AHPCC



Pawel Wolinski, PhD
AHPCC



AGENDA

- Introduction to Arkansas Research Platform (ARP)
- Requirements
 - Getting an account on Pinnacle
 - Logging into Open OnDemand portal (OOD)
 - OOD overview
- Running Jupyter Notebooks and Python Programs
- Running shell commands

AGENDA

- Data
 - Google Drive and Github
 - OOD File upload/download
 - SCP terminal command
- Running Jobs
 - OOD
 - Terminal
- Bioimaging Support
 - Traditional machine learning tools: ImageJ and Fiji
 - Deep learning tools: ZeroCostDL4Mic

GETTING AN ACCOUNT ON PINNACLE

Pinnacle (UAF, Fayetteville)

Account application: <https://hpc.uark.edu/hpc-support/user-account-requests/index.php>
Access: https://hpcwiki.uark.edu/doku.php?id=off-campus_access_dmz
Website: <https://hpc.uark.edu/>

supported by



ARKANSAS RESEARCH PLATFORM

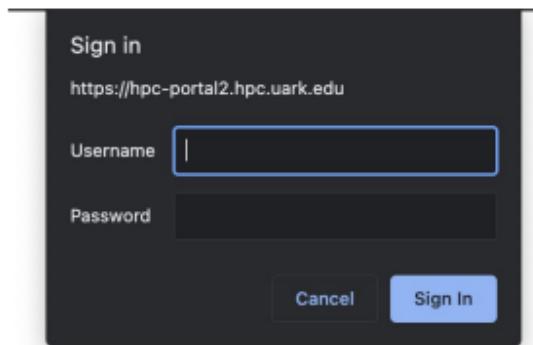


LOGGING INTO OOD

Open OnDemand (OOD) is a web portal interface to HPC resources, which hides the complexities of the HPC scheduling system and Linux/bash command line interface. Also provides a convenient method of displaying desktop/GUI applications running on compute nodes on the users client desktop.

Pinnacle OOD portal

<https://hpc-portal2.hpc.uark.edu/>



The image shows a dark-themed sign-in form titled "Sign in". The URL "https://hpc-portal2.hpc.uark.edu" is displayed above the form. The form has two input fields: "Username" and "Password", both with placeholder text. Below the fields are "Cancel" and "Sign In" buttons. The "Sign In" button is highlighted with a blue border.

Usernames and passwords supplied in the login instruction emails when Pinnacle and Grace accounts were created.

OOD OVERVIEW



Shell access
Start running GUI desktop and apps
Monitor app sessions

Access home directory files
Create and monitor jobs

Files ▾ Jobs ▾ Clusters ▾ Interactive Apps ▾ My Interactive Sessions

PINNACLE

TRESTLES

KARPINSKI

Global Research in Arkansas – Computing Environment

UAMS UNIVERSITY OF ARKANSAS FOR MEDICAL SCIENCES

UNIVERSITY OF ARKANSAS

UNIVERSITY OF ARKANSAS

UNIVERSITY OF ARKANSAS

OnDemand provides an integrated, single access point for all of your HPC resources.

INTERACTIVE JOBS

The screenshot shows the APCC (Arkansas High Performance Computing Center) web interface. The top navigation bar includes links for Files, Jobs, Clusters, Interactive Apps (which is currently selected), and My Interactive Sessions. The 'Interactive Apps' dropdown menu is open, displaying categories for Desktops, GUIs, and Servers, along with specific application icons and names.

- Desktops**
 - Karpinski Desktop (CSCE)
 - Pinnacle Desktop
- GUIs**
 - Comsol
 - Elmer
 - IGV
 - MATLAB
 - Mathematica
 - Ovito
 - PQS
 - R-Studio
 - SAS
- Servers**
 - VM Launch
 - VMD
 - Jupyter Notebook

Two red arrows point from the text instructions below to the 'Pinnacle Desktop' and 'Jupyter Notebook' items in the dropdown menu.

Start a 'Pinnacle Desktop' session

Start a 'Jupyter Notebook' session

INTERACTIVE JOBS EXAMPLE

The screenshot shows the AHPC (Arkansas High Performance Computing Center) web interface. The top navigation bar includes links for Files, Jobs, Clusters, Interactive Apps (which is currently selected), and My Interactive Sessions. The main content area is titled "Interactive Apps" and contains three sections: Desktops, GUIs, and Servers. Under Desktops, there are links for Karpinski Desktop (CSCE) and Pinnacle Desktop. Under GUIs, there are links for Comsol, Elmer, IGV, MATLAB, Mathematica, Ovito, PQS, R-Studio, and SAS. Under Servers, there are links for VM Launch and VMD. A green oval highlights the "Jupyter Notebook" link under Servers. A red arrow points from this highlighted link to the "Jupyter Notebook" section on the right.

Jupyter Notebook

This app will launch a Jupyter Notebook server on one or more nodes.

Queue

comp01

Number of hours

1

Number of nodes

1

I would like to receive an email when the session starts

Launch

INTERACTIVE JOBS EXAMPLE

Jupyter Notebook (1442103) Queued

Created at: 2023-03-08 11:19:11 CST **Delete**

Time Requested: 1 hour

Session ID: [5363de98-2e35-45e5-b718-bb7ce0b0354e](#)

Please be patient as your job currently sits in queue. The wait time depends on the number of cores as well as time requested.



Jupyter Notebook (1442103) 1 node | 1 core | Starting

Created at: 2023-03-08 11:19:11 CST **Delete**

Time Remaining: 59 minutes

Session ID: [5363de98-2e35-45e5-b718-bb7ce0b0354e](#)

Your session is currently starting... Please be patient as this process can take a few minutes.



Jupyter Notebook (1442103) 1 node | 1 core | Running

Host: >_c1401 **Delete**

Created at: 2023-03-08 11:19:11 CST

Time Remaining: 59 minutes

Session ID: [5363de98-2e35-45e5-b718-bb7ce0b0354e](#)

Connect to Jupyter

EXERCISE 1 (5 MINS)

LOG INTO
**PINNACLE DESKTOP (TRES72) AND
JUPYTER NOTEBOOK (GPU06)**

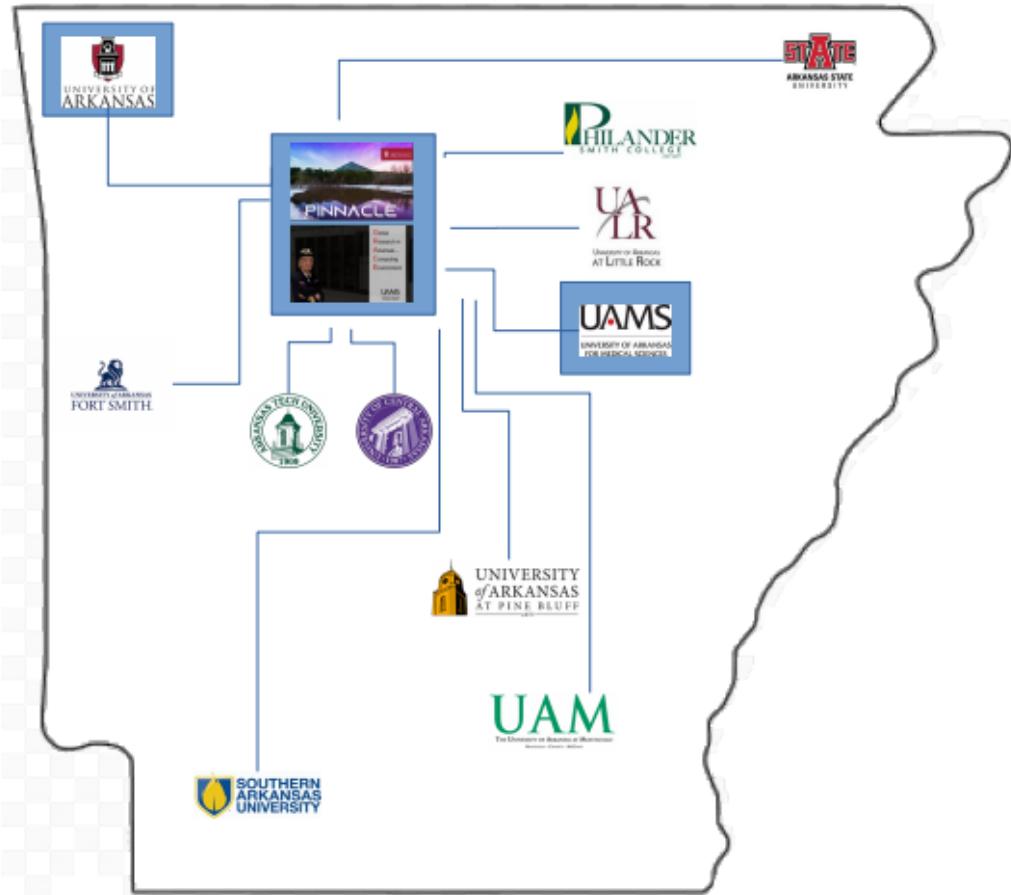
CHAPTER

INTRODUCTION TO ARKANSAS RESEARCH PLATFORM (ARP)



OVERVIEW

Arkansas Research Platform



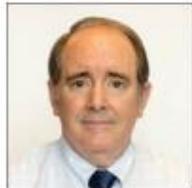
Arkansas Research Platform is a collaboration of higher education institutions within the state of Arkansas which provides computing resources and data storage to all its members (free of charge to students, faculty and staff). The core components are the **Pinnacle cluster** managed by the Arkansas High Performance Computing Center (UAF, Fayetteville) and the **Grace cluster** managed by UAMS High Performance Computing Center (UAMS, Little Rock).

OVERVIEW

Arkansas Research Platform – Support Staff



David Chaffin
dchaffin@uark.edu



Lawrance Tarbox
lrtarbox@uams.edu



Jeff Pummill
jpummil@uark.edu



Kaleb Abram
kzabram@uams.edu



Pawel Wolinski
pwolinsk@uark.edu

hpc-support@listserv.uark.edu

hpcadmin@uams.edu

supported by

UAMS AHPCC
High Performance Computing Arkansas High Performance Computing Center

14

ARKANSAS RESEARCH PLATFORM

ARP

AIMRC
Arkansas Integrative Metabolic Research Center - AIMRC.net

OVERVIEW

Pinnacle Cluster - Fayetteville



ARP RESOURCES

Pinnacle Cluster – Fayetteville

Compute Nodes

100 2x Intel Xeon 6130, 2.1GHz, 32cores, 192GB RAM
50 2x AMD Epyc 7543, 2.8GHz, 64cores, 1024GB RAM
248 other/misc configuration

GPU equipped Nodes: 56

13 4x NVidia A100
1 2x NVidia A100
18 1x NVidia A100
1 2x NVidia V100
20 1x NVidia V100
1 4x NVidia TitanV
2 4x NVidia K80

Large Memory Nodes (>1TB): 4

3 2TB RAM
1 3TB RAM

Storage

2.3 PB main storage (DDN SFA14k exascaler)
0.02PB scratch storage (HPE nvme/lustre)
1 PB archive storage (Dell/lustre)
2 PB object store (Dell/EMC ECS)

Network

Internal 100Gbps Infiniband interconnect
External 100Gbps Ethernet connection

supported by



ARP RESOURCES

Pinnacle Cluster – Fayetteville

Node Count		CPUs	Memory	GPUs	
124	AMD	a6136	32cores	64GB	None
70	Intel	i6130	32cores	192GB	None
55	Intel	i2650v2	16cores	64GB	None
19	Intel	i6130	32cores	192GB	1xNvidia V100
18	AMD	a7543	64cores	1024GB	1xNvidia A100
14	Intel	i2650v4	24cores	128GB	None
13	AMD	a7543	64cores	1024GB	4xNvidia A100
12	Intel	i2620	8cores	32GB	None
10	Intel	i6230	40cores	384GB	None
8	AMD	a7543	32cores	512GB	None
8	AMD	a7543	64cores	256GB	None
7	Intel	i6128	24cores	768GB	None
5	AMD	a7543	64cores	512GB	None
3	AMD	o6378	64cores	512GB	None
3	Intel	i2667v4	16cores	256GB	None
3	AMD	a7543	32cores	1024GB	None
2	Intel	i2650v2	16cores	128GB	4xNvidia K80
2	AMD	a7351	32cores	256GB	None
2	Intel	a7110	64cores	112GB	None
1	Intel	i4166	24cores	192GB	4xNvidia TitanV
1	Intel	i6130	32cores	768GB	2xNvidia V100
1	AMD	a7543	64cores	512GB	2xNvidia A100
1	Intel	i6230	40cores	192GB	1xNvidia V100
1	Intel	i2620	8cores	92GB	None
1	Intel	i2620	8cores	8GB	None
1	Intel	i2620	8cores	82GB	None
1	Intel	i4627v2	32cores	768GB	None
1	Intel	i2620	8cores	72GB	None
1	Intel	i2620	8cores	62GB	None
1	Intel	i2620	8cores	52GB	None
1	Intel	i2650v4	24cores	512GB	None
1	Intel	i2620	8cores	42GB	None
1	Intel	i4860v2	96cores	3072GB	None
1	AMD	a7402	48cores	256GB	None
1	Intel	i4627v2	128cores	2048GB	None
1	AMD	a7543	64cores	2048GB	None
1	AMD	a7402	48cores	2048GB	None
1	Intel	i6128	24cores	192GB	None
1	AMD	a7452	64cores	1024GB	None

Total #of compute nodes: 398
Total #of cores: 12,912

AIMRC SERVER

- Four Dell XE8545 dual CPU quad GPU servers with NVidia A100 GPUs, AMD 7543 CPUs, and 1024 GB of memory.
- Integrated to AHPCC (Arkansas High Performance Computing Center).
- AIMRC researchers have priority access based on the condo computing model.



WHEN TO USE HPC RESOURCES

- NO - word processing
- NO - editing spreadsheets
- NO - editing images
- NO - plotting graphs
- NO - storing personal images
- NO - storing personal movies
- NO - bitcoin mining
- NO - playing multi-player online games
- YES - GPU accelerated programs
- YES - machine learning/AI
- YES - parallel multi-node simulations
- YES - visualizing large data sets
- YES - moving large data sets
- YES - processing large data sets

HPC SOFTWARE

Software on Pinnacle

Molecular dynamics: namd, lammps

Quantum chemistry: quantum espresso, VASP, Gaussian, PQS, NWChem

Multiphysics: COMSOL, Ansys

Data processing programming languages: Python, R

Simulation frameworks: Mathematica, Matlab

Statistics: SAS

Bioinformatics: abyss, bamtools .. spades, samtools, trinity...
(Bioinformatics Software on Pinnacle)

450 application packages
(150 packages installed as
python or R modules)

An **OpenOnDemand** web
portal, Jupyter Notebooks,
RStudio, and other
graphical interfaces in
addition to command line
batch computing.

supported by



ARKANSAS RESEARCH PLATFORM



VOCABULARY

cluster - multiple computers connected by a network, with access to the same file system.

node - a single computer

login node - a gateway node to the cluster. Shared by all users. No computationally intensive tasks are permitted

compute node - node which can run computationally intensive tasks

queueing system - software which organizes and schedules jobs submitted by users to be executed on compute nodes

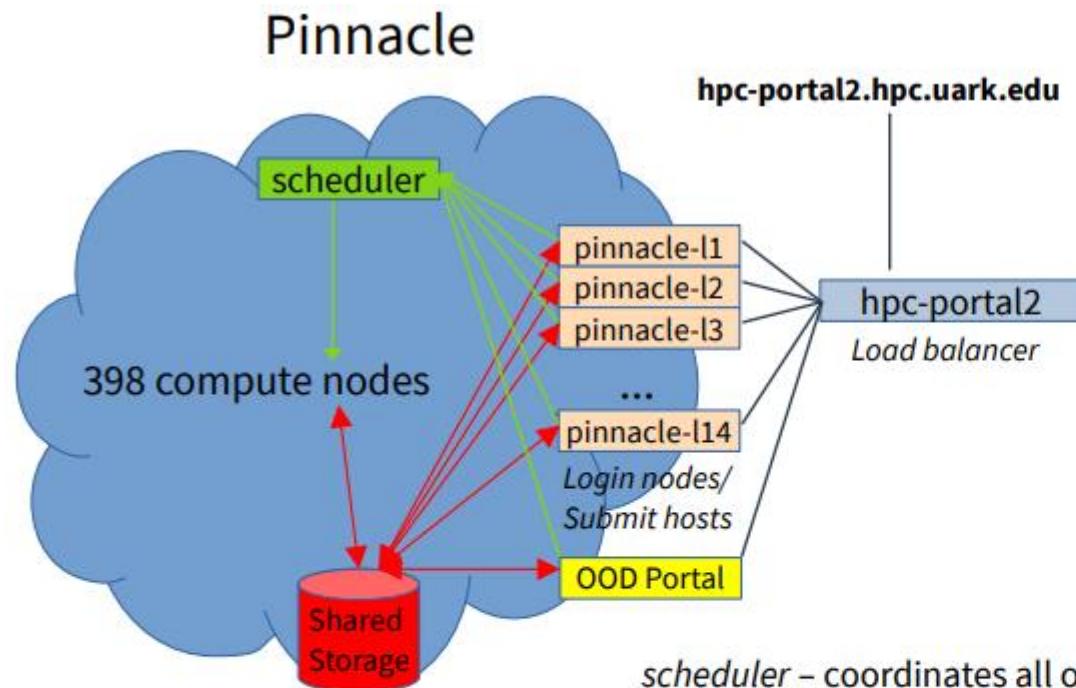
job - a computational task submitted by a user to the queueing system

cpu/core/thread/socket - central prcoossing unit (CPU) is a mircoprocessor on a single integrated circuit with 2 or more processing units (cores). Core can run a single instruction at a time. With Hyperthreading the same core can run 2 instructions at once. Socket is a interface on the motherboard which receives a CPU. HPC motherboards usually have 2 - 4 sockets.

supported by



CLUSTER FUNCTIONAL DIAGRAMS



scheduler – coordinates all of the nodes in the queue (SLURM queueing system)
submit hosts – nodes from which users submit jobs to the queues
login hosts – nodes accessible from the external network
compute nodes – nodes assigned to queues and run jobs

INTERACTIVE JOBS EXAMPLE

Jupyter Notebook (1442103) Queued

Created at: 2023-03-08 11:19:11 CST **Delete**

Time Requested: 1 hour

Session ID: [5363de98-2e35-45e5-b718-bb7ce0b0354e](#)

Please be patient as your job currently sits in queue. The wait time depends on the number of cores as well as time requested.



Jupyter Notebook (1442103) 1 node | 1 core | Starting

Created at: 2023-03-08 11:19:11 CST **Delete**

Time Remaining: 59 minutes

Session ID: [5363de98-2e35-45e5-b718-bb7ce0b0354e](#)

Your session is currently starting... Please be patient as this process can take a few minutes.



Jupyter Notebook (1442103) 1 node | 1 core | Running

Host: >_c1401 **Delete**

Created at: 2023-03-08 11:19:11 CST

Time Remaining: 59 minutes

Session ID: [5363de98-2e35-45e5-b718-bb7ce0b0354e](#)

Connect to Jupyter

EXERCISE 1

(5 MINS)

DESKTOP AND DATA TRANSFER



A screenshot of a web browser window. The address bar contains the URL <https://aimrc.uark.edu/data-science-core-resources/>. The page itself is titled "Data Science Core Resources" and features the AIMRC logo at the top. A navigation menu includes links for Home, Current Research, Cores, Team, News, Opportunities, Contact Us, and a search icon. Below the menu is a large, abstract blue and green background image.

Data Science Core Resources

Upcoming Workshops:

25

July
2024

Getting started with the Arkansas Research Platform

A friendly tutorial on accessing AIMRC high-performance computing resources.

1:00 – 3:00 pm

JBHT 239/ Zoom

TBD

TBD
2024

Open-source deep learning tools for bioimaging

An introduction to available open-source deep learning tools for bio-imaging along with a demo.

TBD

TBD

Office Hours:

Dr. Prateek Verma

- Location: JBHT 351
- Monday: 1:30 – 3:30 pm
- Tuesday: 1:30 – 3:30 pm
- Wednesday: 9:30 – 11:30 am

Important Links

- Our [GitHub repository](#) (look here for code)
- Our [Google Drive](#) (look here for data and other resources)

In this exercise, you learned how to log into the Pinnacle Desktop and how to navigate the system. This exercise was done outside of the Jupyter notebook environment, of course.

Make sure you

1. Logged in to Pinnacle in your browser.
 2. Opened a Pinnacle Desktop session within your browser.
Opened a Jupyter Notebook session within your browser.
 3. Accessed Firefox browser within the Pinnacle Desktop session.
 4. Downloaded the `workshop-hpc-data` folder from the Google Drive link above.
 5. Downloaded the GitHub repository from the link above.
-
6. Extract both downloaded zip files to your home/Downloads directory. Right click in the downloaded zip file and click on 'Extract Here'.
 7. You are free to download/extract your files to any directory you prefer. But if you followed the instructions above, your Home/Downloads directory should look like this:

```
/home/username/Downloads/
    └── aimrc-data-science-core-main
    └── workshop-hpc-data
        ├── aimrc-data-science-core-main.zip
        └── workshop-hpc-data-20240724T162726Z-001.zip
```

CHAPTER

ACCESSING PINNACLE VIA TERMINAL



ON CAMPUS – CHOICE OF SSH CLIENT

Secure Shell (SSH) is a protocol for exchanging information between two nodes on a network: an SSH server and an SSH client. SSH client requests a connection from and SSH server. Client runs only when a connection is established. Server runs continuously waiting for connection requests.

An SSH server is running on each of the login nodes on the ARP clusters (as well on each of the compute nodes).

Linux, Mac

SSH client is built in! (server too)

Windows

Windows Subsystem for Linux (WSL): Open PowerShell and run “wsl --install”
OR

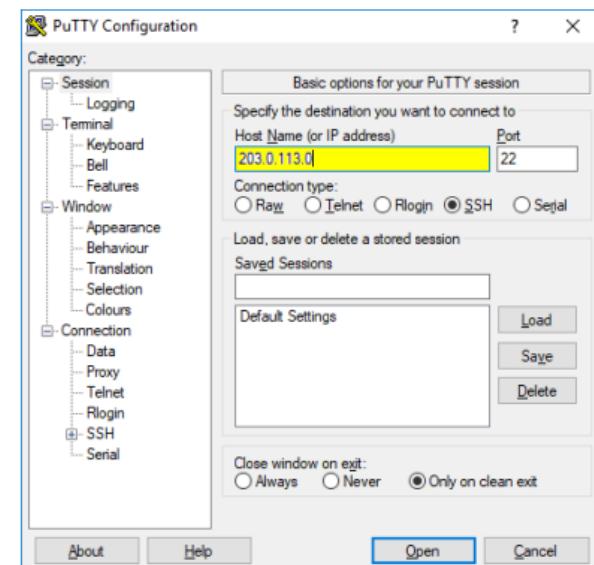
1. please open this page in a browser:

<http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html>
(first link in google search with for keyword “putty”)

2. and right click and save to the desktop:

putty.exe, pscp.exe and plink.exe

<http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html>



ON CAMPUS – CONNECTING TO PINNACLE

ssh user@hostname - log into a remote system

Linux

1. open a terminal and run
2. run:

ssh <username>@hpc-portal2.hpc.uark.edu *Pinnacle*

3. enter your supplied password when prompted.

Windows

1. double click on **putty.exe** on the desktop

2. specify

hpc-portal2.hpc.uark.edu *Pinnacle*

as “Hostname” and click on “Open” button.

3. enter your username (w/o ‘@uark.edu’ or ‘@uams.edu’)
4. enter your supplied password

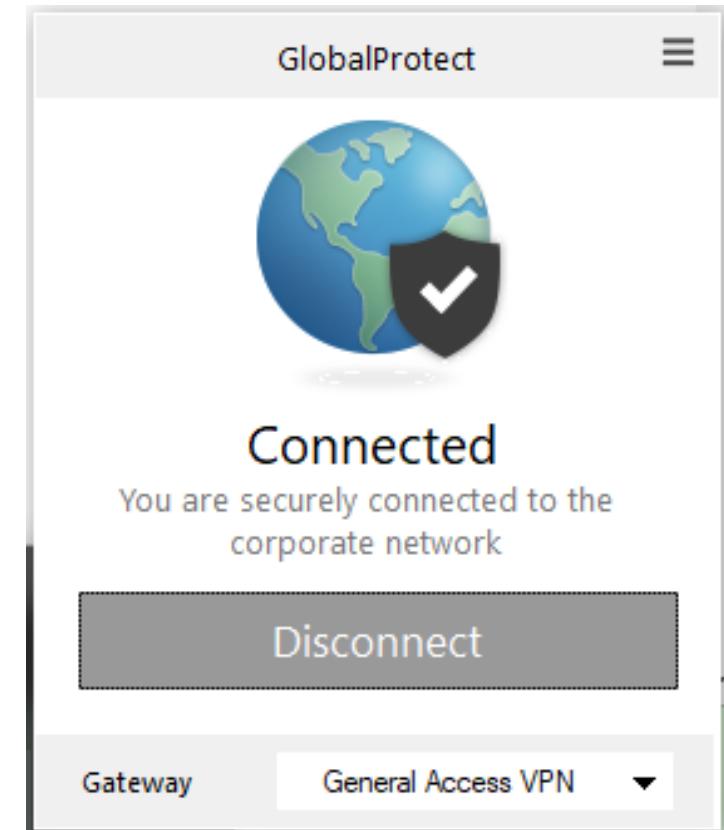
ssh <username>@hpc-portal2.hpc.uark.edu *Pinnacle*

supported by



OFF CAMPUS VPN

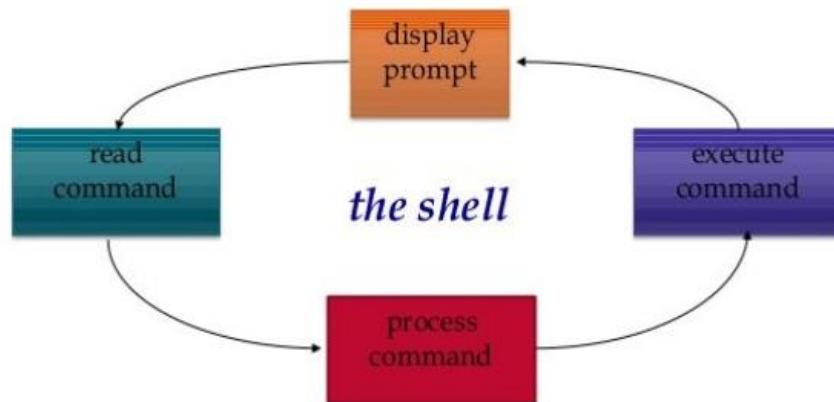
- For off-campus access, use GlobalProtect VPN as instructed in the following link
 - <https://its.uark.edu/network-access/vpn/>
- Once connected, your computer is now part of the university network and pinnacle can be accessed as described on the on-campus process.



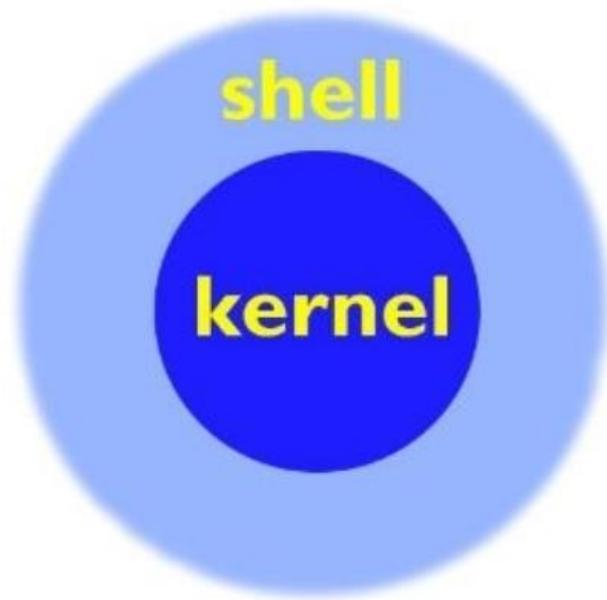
BASH SHELL

Kernel – core part of the operating system. Contains all the drivers necessary to communicate with the underlying hardware. Manages the CPUs, memory, I/O operations from all peripheral devices (disk, keyboard, mouse, printers, scanners etc.)

Shell – provides the Command Line Interface (CLI) for interaction with the kernel and maintains the environment, a set of variables. Bash is the most commonly used shell on Linux.



Kernel receives commands from the shell and performs them. There may be multiple shells interacting with the kernel at one time (multiple users logged in at once, each with multiple shells).



BASH SHELL - ENVIRONMENTS

env – shows all of the environmental variables that are defined in your shell

- SHELL – the shell you are running
- HOSTNAME - name of the host
- PWD - present working directory
- USER - user name
- HOME - path of the user's home directory
- PATH – list of directories to search when user types in a command
- PS1 – configures prompt

BASH SHELL - COMMANDS

env – shows all of the environmental variables that are defined in your shell

echo - display text string or contents of an environmental variable

who – list users currently logged on

whoami

hostname

date

pwd – show present working directory

man - show a manual page (help page) about any command

apropos - show any manual pages which reference a given string

history - show listing of your commands

nano – basic text editor

vi – powerful text editor

FILE SYSTEMS

df - disk free **mount** - show mounted files systems and mount points

Filesystem	1K-blocks	Used	Available	Use%	Mounted on
devtmpfs	8114528	0	8114528	0%	/dev
tmpfs	8133140	32	8133108	1%	/dev/shm
tmpfs	8133140	271720	7861420	4%	/run
tmpfs	8133140	0	8133140	0%	/sys/fs/cgroup
/dev/mapper/centos-root	8374272	7613856	760416	91%	/
/dev/vdal	1038336	174084	864252	17%	/boot
172.17.27.1@o2ib,172.17.27.21@o2ib:172.17.27.2@o2ib,172.17.27.22@o2ib:/scrfs	2455426280448	1640220406744	790355060004	68%	/scrfs
172.17.27.13@o2ib:/scrl	11468687248	4939330684	5950685576	46%	/scrl
172.17.27.14@o2ib:/scr2	11471001300	7212051640	3680170900	67%	/scr2
tmpfs	1626632	0	1626632	0%	/run/user/0
tmpfs	1626632	0	1626632	0%	/run/user/4948
tmpfs	1626632	0	1626632	0%	/run/user/5430
tmpfs	1626632	0	1626632	0%	/run/user/5194
tmpfs	1626632	0	1626632	0%	/run/user/5060
tmpfs	1626632	0	1626632	0%	/run/user/5282
tmpfs	1626632	0	1626632	0%	/run/user/5632
tmpfs	1626632	0	1626632	0%	/run/user/4924

Filesystem	Size	Used	Avail	Use%	Mounted on
devtmpfs	7.8G	0	7.8G	0%	/dev
tmpfs	7.8G	32K	7.8G	1%	/dev/shm
tmpfs	7.8G	266M	7.5G	4%	/run
tmpfs	7.8G	0	7.8G	0%	/sys/fs/cgroup
/dev/mapper/centos-root	8.0G	7.3G	743M	91%	/
/dev/vdal	1014M	171M	844M	17%	/boot
172.17.27.1@o2ib,172.17.27.21@o2ib:172.17.27.2@o2ib,172.17.27.22@o2ib:/scrfs	2.3P	1.5P	736T	68%	/scrfs
172.17.27.13@o2ib:/scrl	11T	4.7T	5.6T	46%	/scrl
172.17.27.14@o2ib:/scr2	11T	6.8T	3.5T	67%	/scr2
tmpfs	1.6G	0	1.6G	0%	/run/user/0
tmpfs	1.6G	0	1.6G	0%	/run/user/4948
tmpfs	1.6G	0	1.6G	0%	/run/user/5430
tmpfs	1.6G	0	1.6G	0%	/run/user/5194
tmpfs	1.6G	0	1.6G	0%	/run/user/5060
tmpfs	1.6G	0	1.6G	0%	/run/user/5282
tmpfs	1.6G	0	1.6G	0%	/run/user/5632
tmpfs	1.6G	0	1.6G	0%	/run/user/4924

FILE SYSTEMS

ls - list files **ls -l** - long listing **ls -al** - show all files (including hidden files, starting with ".")

```
pinnacle-112:hs028:~$ ls -l
total 116
lrwxrwxrwx 1 root root 34 Jan 20 2022 ondemand -> /scrfs/storage/ondemand-data/hs028
drwxr-xr-x 4 hs028 jzhan 4096 Jan 26 2022 python-packages
drwxr-xr-x 8 hs028 jzhan 4096 Jan 27 2022 R01
drwxr-xr-x 7 hs028 jzhan 4096 Feb 22 2021 RITHM-master
-rw-r--r-- 1 hs028 jzhan 97944 Apr 18 2022 RITHM-master.zip
drwxr-xr-x 6 hs028 jzhan 4096 Feb 16 2022 TWITTER_DATA
drwxr-xr-x 13 hs028 jzhan 4096 Jan 20 2022 Twitter_Old
```

	user	group	Size in bytes	Modification date	File name
--	------	-------	---------------------	----------------------	-----------

user group others

file / directory / link
read / write / execute

EXERCISE 2

(5 MINUTES)

(SSH, FILESYSTEM)



Type the following into Windows Command Prompt:
ssh username@hpc-portal2.hpc.uark.edu

Practice running the following commands

```
env  
who  
whoami  
hostname  
date  
df  
cd /home/username/Downloads  
ls -al
```

CHAPTER

JUPYTER NOTEBOOKS



INTERACTIVE JOBS EXAMPLE

The screenshot shows a Jupyter Notebook interface. At the top, there is a navigation bar with tabs for "Files", "Running", "IPython Clusters", and "Nbextensions". On the right side of the header, there are "Quit" and "Logout" buttons. Below the header, a message says "Select items to perform actions on them." A file list table is displayed, showing the following entries:

File/Folder	Last Modified	Size
ondemand	2 years ago	
privatemodules	2 days ago	
python-packages	a year ago	
R01	a year ago	
RITHM-master	3 days ago	123 B
test	2 days ago	
TWITTER_DATA	a year ago	
Twitter_Old	a year ago	
hady.sh	3 days ago	97.9 kB
RITHM-master.zip	a year ago	
slurm-1435917.out	3 days ago	118 B
slurm-1435920.out	3 days ago	118 B

On the right side of the interface, there is a "New" button with a dropdown menu. The "Notebook" option in the dropdown is highlighted with a green oval and has a green arrow pointing to it from below. The "Notebook" option is labeled "Python 3 (ipykernel)". Other options in the dropdown include "Text File", "Folder", and "Terminal".

EXERCISE 3

(5 MINUTES)

JUPYTER NOTEBOOK – GET SYSTEM INFO

CHAPTER

FILE MANAGEMENT, MOVING DATA



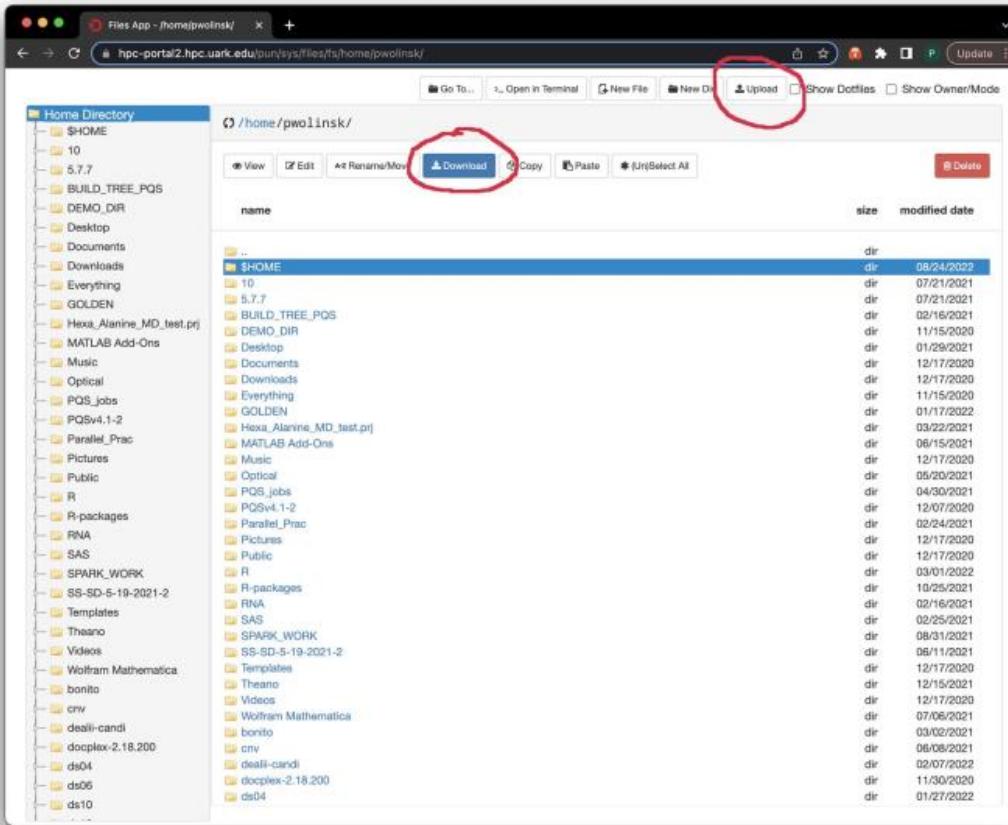
OOP FILE UPLOAD/DOWNLOAD

File dialog window

- Top Menu: **Files** → **Home Directory**
New browser tab opens with the contents of your home directory
- **Upload** button - select file from your local machine and upload it to Pinnacle
- **Download** button - select file from your Pinnacle account and download it to your local machine

Using those two buttons:

- 1) download file etc/redhat-release to your laptop
- 2) Upload any text file to your home directory on Pinnacle



supported by



ARKANSAS RESEARCH PLATFORM

ARP AIMRC
Arkansas Integrative Metabolic Research Center - AIMRC.net

OOD SCP TERMINAL COMMAND

scp – secure copy uses SSH protocol to move files between hosts on network. Unlike wget, scp allows to move files both ways: **download** from remote host and **upload** to remote host. **-r** (recursive) option transfers directories and all its contents.

`scp <source> <destination>`

Source and destination can be of from `username@host:file_path`

Recursive copy: `scp -r <source> <destination>`

```
scp [OPTIONS] [[user@]src_host:]file [[user@]dest_host:]file
```

The most common options:

- p** Preserves modification times, access times, and modes from the original file.
- P port** Specifies the port to connect to on the remote host.
- r** Recursively copy entire directories. Note that **scp** follows symbolic links encountered in the tree traversal.
- q** Quiet mode: disables the progress meter as well as warning and diagnostic messages.
- C** Compression enable. Passes the **-C** flag to `ssh(1)` to enable compression.

EXERCISE 4

(5 MINUTES)

A SIMPLE PYTHON PROGRAM



CHAPTER

SOFTWARE MODULES AND PYTHON/ANACONDA



SOFTWARE MODULES

Hundreds of applications, compilers and scripting languages, some with multiple versions, are installed on Grace and Pinnacle. They cannot all be available to the users at once because they interfere with each other.

The Environment Modules system is a tool to help users manage the shell environment and provide access to specific applications and versions on demand. It does this by manipulating the values of shell environmental variables. The most important of those are **\$PATH** and **\$LD_LIBRARY_PATH**.

```
pinnacle-112:hs028:~$ ls /share/apps/
abinit           espresso    intel2020-2          obsoletemodulefiles      ShengBTE
accumulo        etsf_io     intelpython2        oldspackmodulefiles    siesta
acml            examples     JAGS                openblas               silvaco
alamode         fcron       java                OpenBLAS              singularity
amber           FEBio       jeigen              opencv                 slurm
amber16tools17 ffmpeg      jna                 OpenFOAM              smartmontools
amdappsdk       fftw        julia               openmpi               smrtlink
ansys          Fiji.app    kafka              opt                  spack
aocl            fluidix    knem               oracle_developer_studio spackmodulefiles

pinnacle-112:hs028:~$ ls /share/apps/bioinformatics/
abyss          bcftools    braker3             cutadapt      fastqc      guppy      kraken2
admixture      beagle      broccoli            DataAnalysis fastStructure HipSTR     Ktrim
aegean         beagle-java busco               datamash     FastTree     hisat2     last
agalma         beast       bwa                 dDocent      fastx       hmmer     lib
allpaths-lg    bedops      bwa-meth            deepTools   filecount   htslib     libBigWi
angsd         bedtools2  cadnano            diamond     filecount.sh idba      list1
annovar        bern       CAFE                dock       flash       ie.sh      listls
astalavista   bfc        canu               edirect     Flye        igb       lordec
astral         bin        CAP3                eigen      freebayes  IGV       lorma
ASTRAL         bismark    cdhit              emboss     gappa      impute    lsc
atria          blas       Celera             EnTAP      GATK       infernal  MACSE
```

SOFTWARE MODULES

module avail - list all available modules

module list – list modules currently loaded

module load – load a module

module remove – remove a loaded module

```
pinnacle-15:hs028:~$ module avail
```

----- /share/apps/modulefiles -----			
BayesAss3-SNPs/1.0	cassandra/3.11.6	gctf/1.06	intelcompiler/22.2.0
BerkeleyGW/2.1	ccl/latest	gd/2.2.2	intelcompiler/2023.0.0
BerkeleyGW/3.0.1	(D) ccl/2021.8.0	(D) gd/2.3.2	interproscan/5.17.56.0
CAFE/4.2.1	cdhit/4.6.5	gdal/1.11.1	interproscan/5.57.90.0 (D)
CERN/6.24	cdhit/4.6.8	gdal/2.2.4	iqtree/1.6.12
EnTAP/0.8.1	cdhit/4.8.1	(D) gdal/3.3.1	isown/1.0
LIGGGHTS/1	cdo/2.0.5	gemma/0.94.1	itac/latest
METABOLIC/4.0	changa/2.0	gemma/0.97	(D) itac/2021.8.0 (D)
MUMmer/3.23	cistem/1.0.0-beta	genemark/2.6-3	jags/4.1.0
MaSuRCA/3.4.2	clang/5.0.1	genemark_es/4	jags/4.3.0 (D)
MetaGeneMark/3.38	clck/latest	genometools/1.5.8	java/ibmjdk_1.8.0
PAML/4.6	clck/2021.7.2	(D) geos/3.6.2	java/openjdk_1.7.0
PAML/4.8	clustalw/2.1	git/2.28.0	java/openjdk_1.8.0_20
PAML/4.9e	cmake/3.5.2	git/2.35.1	(D) java/openjdk_1.8.0

```
pinnacle-15:hs028:~$ module list
```

```
Currently Loaded Modules:  
1) os/el7
```

```
pinnacle-15:hs028:~$ module load git  
pinnacle-15:hs028:~$ module list
```

```
Currently Loaded Modules:  
1) os/el7 2) git/2.35.1
```

```
pinnacle-15:hs028:~$ module remove git/2.35.1  
pinnacle-15:hs028:~$ module list
```

```
Currently Loaded Modules:  
1) os/el7
```

SOFTWARE MODULES (ANACONDA)

Activate base environment

```
pinnacle-15:hs028:~$ source /share/apps/bin/conda-3.10.sh
(base) pinnacle-15:hs028:~$ conda env list
# conda environments:
#
private-env          /home/hs028/.conda/envs/private-env
base                 * /share/apps/python/anaconda-3.10
STAR                 /share/apps/python/anaconda-3.10/envs/STAR
bioperl-3.10          /share/apps/python/anaconda-3.10/envs/bioperl-3.10
braker2tabix-3.10    /share/apps/python/anaconda-3.10/envs/braker2tabix-3.10
busco-3.10            /share/apps/python/anaconda-3.10/envs/busco-3.10
conda-3.10-perl       /share/apps/python/anaconda-3.10/envs/conda-3.10-perl
cudnn-3.10            /share/apps/python/anaconda-3.10/envs/cudnn-3.10
deeplab2-3.10          /share/apps/python/anaconda-3.10/envs/deeplab2-3.10
deformetrica-3.10     /share/apps/python/anaconda-3.10/envs/deformetrica-3.10
htseq-3.10             /share/apps/python/anaconda-3.10/envs/htseq-3.10
mpi4py-mvapich-3.10   /share/apps/python/anaconda-3.10/envs/mpi4py-mvapich-3.
10                   /share/apps/python/anaconda-3.10/envs/mpi4py-openmpi-3.
10
mpi4py-openmpi-3.10    /share/apps/python/anaconda-3.10/envs/p4vasp-3.10
p4vasp-3.10            /share/apps/python/anaconda-3.10/envs/phonopy-3.10
phonopy-3.10           /share/apps/python/anaconda-3.10/envs/poolsnp-3.10
poolsnp-3.10           /share/apps/python/anaconda-3.10/envs/poppunk-3.10
poppunk-3.10           /share/apps/python/anaconda-3.10/envs/prokka-3.10
prokka-3.10            /share/apps/python/anaconda-3.10/envs/pymor-3.10
pymor-3.10              /share/apps/python/anaconda-3.10/envs/qiime2-2022.2-3.1
qiime2-2022.2-3.10     /share/apps/python/anaconda-3.10/envs/repeatmodeler-3.1
repeatmodeler-3.10     /share/apps/python/anaconda-3.10/envs/scoary-3.10
scoary-3.10             /share/apps/python/anaconda-3.10/envs/terra-3.10
terra-3.10              /share/apps/python/anaconda-3.10/envs/thirdorder-3.9
thirdorder-3.9
```

Load anaconda module

```
pinnacle-15:hs028:~$ module load python/3.10-anaconda
pinnacle-15:hs028:~$ module list

Currently Loaded Modules:
 1) os/el7    2) python/3.10-anaconda
```

Check module loaded

```
pinnacle-15:hs028:~$ which python
/share/apps/python/anaconda-3.10/bin/python
```

CHAPTER

JOBS – INTERACTIVE, BATCH



RUNNING JOBS – OPEN ONDEMAND (OOD)



JOB COMPOSER

The image shows the AHPCC OnDemand interface. At the top, there is a navigation bar with the AHPCC logo, followed by menu items: Files, Jobs, Clusters, and Interactive Apps. The 'Jobs' item is highlighted with a red circle and has a dropdown menu. The dropdown menu contains two items: 'Active Jobs' and 'Job Composer'. A red arrow points from the text 'Click on 'Job Composer' to create, start, or stop jobs' to the 'Job Composer' button in the dropdown menu. Below the navigation bar, there is a large image of a landscape with a blue sky and green trees. Dashed lines connect the 'Job Composer' button to a second navigation bar below it. This second bar also has the AHPCC logo and the same four menu items. To the right of the menu items, there is a link labeled 'My Interactive Sessions'. Below this bar, there are three large cards representing different resources: 'PINNACLE' (a snowy mountain landscape), 'TRESTLES' (a close-up of a wave), and 'KARPINSKI' (a server room). Each card features the University of Arkansas logo in the top right corner.

Click on 'Job Composer' to create, start, or stop jobs

UNIVERSITY OF ARKANSAS

PINNACLE

TRESTLES

UNIVERSITY OF ARKANSAS

KARPINSKI

UNIVERSITY OF ARKANSAS

Global Research in Arkansas – Computing Environment

UAMS UNIVERSITY OF ARKANSAS FOR MEDICAL SCIENCES

OnDemand provides an integrated, single access point for all of your HPC resources.

NEW JOB

Jobs

Create a new job using template file

Create a new job using custom file

Create a new job using existing job

Create a template file to be used for new jobs

The screenshot shows a user interface for managing jobs. At the top left, there's a 'New Job' button with a dropdown arrow. Below it are three options: 'From Default Template', 'From Specified Path', and 'From Selected Job'. The 'From Selected Job' option is highlighted with a red arrow pointing to it. To the right of these options are 'Open Terminal', 'Submit', and 'Stop' buttons. Further down, there's a table header with columns for 'Created', 'Name', 'ID', 'Cluster', 'Status', and sorting arrows for each. A 'Delete' button is also visible. At the bottom, there's a search bar labeled 'Search:' and navigation buttons for 'Previous' and 'Next'.

NEW JOB EXAMPLE

Create New "Hadi_Template"

Change these notes by editing the manifest.yml in this template's directory

Job Name: Hadi_Template

Cluster: Pinnacle

Script Name: main_job.sh

Create New Job **Reset**

Selected Template Details

Template location: /home/hs028/ondemand/data/sys/myjobs/templates/hadi_template/

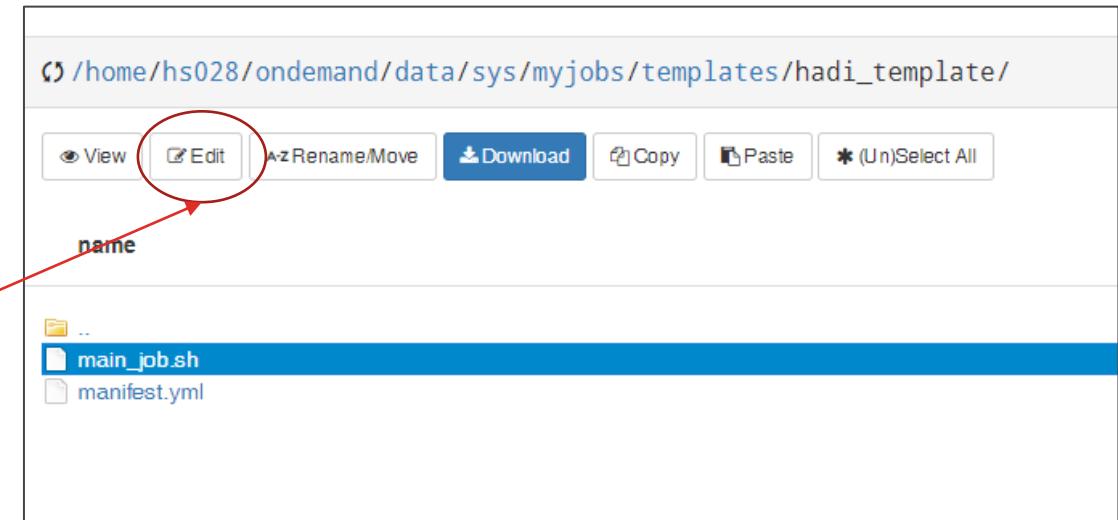
Folder Contents:

- main_job.sh
- manifest.yml

Open Dir

Click to create job

Click to navigate and modify the main_job.sh file



```
$ myjob.sh  X
aimrc-data-science-core > workshops > hpc > $ myjob.sh
1 #!/bin/bash
2 module load python/3.10-anaconda
3 source /share/apps/bin/conda-3.10.sh
4 python crop_images.py
```

NEW JOB EXAMPLE

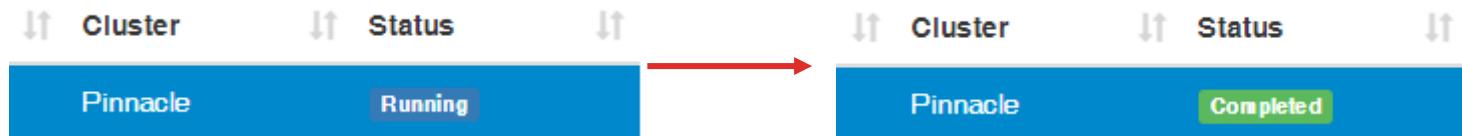
Jobs

Click to submit job

After submit status will turn to running

The screenshot shows a user interface for managing jobs. At the top, there's a header with a 'New Job' button, 'Edit Files', 'Job Options', 'Open Terminal', 'Submit' (highlighted in green), and 'Stop' buttons. To the right are 'Create Template' and 'Delete' buttons, and a search bar. Below this is a table with columns: Created, Name, ID, Cluster, and Status. The first row shows a job named 'Hadi_Template' created on 'February 28, 2023 9:47am' on 'Pinnacle' cluster with 'Not Submitted' status. The second row shows a job named '(default) Simple Sequential Job' created on 'February 28, 2023 9:22am' on 'Karpinski' cluster with 'Not Submitted' status. At the bottom, it says 'Showing 1 to 2 of 2 entries' and has 'Previous', '1', and 'Next' buttons.

Created	Name	ID	Cluster	Status
February 28, 2023 9:47am	Hadi_Template		Pinnacle	Not Submitted
February 28, 2023 9:22am	(default) Simple Sequential Job		Karpinski	Not Submitted



BATCH JOBS

Open Ondemand Portal – First HPC job

Submit batch job

Top Menu: Jobs→Job Composer

- +New Job button
- Edit Files button
-
- Select 'myjob.sh'
- View button

```
#!/bin/bash
# JOB HEADERS HERE

echo "Hello World"
```

- Close browser tab
- Submit button

View results

Wait for "completed status"

- Edit Files button
- Select 'slurm-?????.out'
- View button

Hello World

supported by



RUNNING JOBS – TERMINAL

QUEUEING SYSTEMS

```
pinnacle-112:hs028:~$ sinfo
PARTITION AVAIL TIMELIMIT NODES STATE NODELIST
condo      up   infinite    1  comp* cl306
condo      up   infinite    1  down* c2002
condo      up   infinite   14    mix c[1427,1611,1720,1913,1918-191
condo      up   infinite   31    alloc c[1311-1314,1421,1716,1914-191
condo      up   infinite  122    idle c[1301-1305,1307-1310,1315-133
pcon06     up   6:00:00     1  comp* cl306
pcon06     up   6:00:00     1  down* c2002
pcon06     up   6:00:00     14    mix c[1427,1611,1720,1913,1918-191
pcon06     up   6:00:00     31    alloc c[1311-1314,1421,1716,1914-191
pcon06     up   6:00:00    120    idle c[1301-1305,1307-1310,1315-133
comp01*    up   1:00:00     1    mix c1410
comp01*    up   1:00:00    35    alloc c[1403-1409,1501-1520,1603-161
comp01*    up   1:00:00     2    idle c[1401-1402]
comp06     up   6:00:00     1    mix c1410
comp06     up   6:00:00    24    alloc c[1403-1409,1411-1415,1417-142
comp72     up   3:00:00:00   1    mix c1410
comp72     up   3:00:00:00   43    alloc c[1404-1409,1411-1415,1417-142
acomp06    up   6:00:00     1    idle c2111
gpu06      up   6:00:00    19    alloc c[1612-1615,1701-1715]
gpu72      up   3:00:00:00   19    alloc c[1612-1615,1701-1715]
agpu06     up   6:00:00     8    mix c[1907,1910,2005,2008,2106,210
agpu06     up   6:00:00    10    idle c[1905-1906,1908-1909,1911-191
agpu72     up   3:00:00:00   8    mix c[1907,1910,2005,2008,2106,210
agpu72     up   3:00:00:00   8    idle c[1908-1909,1911-1912,2006-200
qgpu06     up   6:00:00     4    idle c[1901-1904]
qgpu72     up   3:00:00:00   4    idle c[1901-1904]
himem06    up   6:00:00     1  down* c1423
himem06    up   6:00:00     1  alloc c1422
himem06    up   6:00:00     4  idle c[1424-1426,1428]
himem72    up   3:00:00:00   1  down* c1423
himem72    up   3:00:00:00   1  alloc c1422
himem72    up   3:00:00:00   4  idle c[1424-1426,1428]
cloud72    up   3:00:00:00   3  idle c[1331,1601-1602]
csce72     up   3:00:00:00   14  idle cs[2204-2216,2218]
csce-k2-72 up   3:00:00:00   7  idle cs[2013-2019]
cscloud72  up   3:00:00:00   2  idle cs[2202-2203]
tres72     up   12:00:00:00  1  drain* tres0731
tres72     up   12:00:00:00  5  down* tres[0132,0210,0823,1004-1005]
tres72     up   12:00:00:00  1  mix tres0732
tres72     up   12:00:00:00  1  alloc tres0203
tres72     up   12:00:00:00 102  idle tres[0114,0116,0120,0125-0127,
```

sinfo – show node/queue assignment

QUEUEING SYSTEMS (CONT'D)

JOBID	PARTITION	NAME	USER	ST	TIME	NODES	NODELIST (REASON)
1381997	condo	b3lyp	yingyuan	CG	0:00	1	c1306
1431879	condo	maker_Ch	mussmann	PD	0:00	1	(Resources)
1431880	condo	maker_Ch	mussmann	PD	0:00	1	(Priority)
1431881	condo	maker_Ch	mussmann	PD	0:00	1	(Priority)
1431882	condo	maker_Ch	mussmann	PD	0:00	1	(Priority)
1431883	condo	maker_Ch	mussmann	PD	0:00	1	(Priority)
1431884	condo	maker_Ch	mussmann	PD	0:00	1	(Priority)
1431885	condo	maker_Ch	mussmann	PD	0:00	1	(Priority)
1431886	condo	maker_Ch	mussmann	PD	0:00	1	(Priority)
1431887	condo	maker_Ch	mussmann	PD	0:00	1	(Priority)
1431888	condo	maker_Ch	mussmann	PD	0:00	1	(Priority)
1431878	condo	maker_Ch	mussmann	R	2-01:15:35	1	tres0734
1431876	condo	maker_Ch	mussmann	R	2-21:53:56	1	tres0735
1431877	condo	maker_Ch	mussmann	R	2-21:53:56	1	tres0736
1433538	condo	CNT	aleaton	R	1-21:25:57	4	c[3209-3212]
1432270	condo	C18_150	aleaton	R	2-14:52:31	4	c[3209-3212]
1432268	condo	C18_120	aleaton	R	2-14:54:03	4	c[3209-3212]
1432267	condo	CCar	aleaton	R	2-14:58:38	4	c[3205-3208]
1432264	condo	Car	aleaton	R	2-15:00:40	2	c[3229-3230]
1432263	condo	CCr	aleaton	R	2-15:02:12	4	c[3201-3204]

Pending

Running

squeue - jobs in the queues

RUNNING INTERACTIVE JOBS

Running
on
random
node

srun – submit an interactive job to the queue

```
pinnacle-112:hs028:~$ srun -p comp01 --pty /bin/bash
Currently Loaded Modules:
 1) os/el7

cl410:hs028:~$ squeue -u hs028
      JOBID PARTITION      NAME      USER ST      TIME  NODES NODELIST(REASON)
      1435908    comp01      bash    hs028  R      0:38      1 cl410
cl410:hs028:~$ exit
exit
pinnacle-112:hs028:~$
```

RUNNING INTERACTIVE JOBS

srun – submit an interactive job to the queue

Running
on a
selected
server
'aimrc'

Select server name

```
pinnacle-15:hs028:~$ srun --partition pcon06 --constraint 'aimrc' --nodes=1 --pty /bin/bash
Currently Loaded Modules:
 1) os/el7

c2003:hs028:~$ squeue -u hs028
      JOBID PARTITION      NAME      USER ST      TIME  NODES NODELIST(REASON)
      1450802    condo      bash    hs028 R      1:23      1 c2003
c2003:hs028:~$ exit
exit
pinnacle-15:hs028:~$
```

RUNNING BATCH JOBS

Running
on
random
server

sbatch – submit a batch job to the queue

```
pinnacle-112:hs028:~$ sbatch -p condo hady.sh
Submitted batch job 1435917
pinnacle-112:hs028:~$ squeue -u hs028
      JOBID PARTITION      NAME      USER ST          TIME  NODES NODELIST(REASON)
      1435917    condo  hady.sh    hs028 PD      0:00      1 (Priority)
pinnacle-112:hs028:~$ ls
hady.sh  ondemand  python-packages  R01  RITHM-master  RITHM-master.zip  slurm-1435917.out
pinnacle-112:hs028:~$ cat slurm-1435917.out
Script is executing on:
c1427
My present working directory is /home/hs028
Time and date
Sun Mar  5 12:56:52 CST 2023
pinnacle-112:hs028:~$
```

Running
on a
selected
server
'aimrc'

Select server name

nodes

```
pinnacle-112:hs028:~$ sbatch --partition pcon06 --constraint 'aimrc' --nodes=1 hady.sh
Submitted batch job 1435920
pinnacle-112:hs028:~$ squeue -u hs028
      JOBID PARTITION      NAME      USER ST          TIME  NODES NODELIST(REASON)
      1435920    condo  hady.sh    hs028 PD      0:00      1 (Priority)
pinnacle-112:hs028:~$ squeue -u hs028
      JOBID PARTITION      NAME      USER ST          TIME  NODES NODELIST(REASON)
pinnacle-112:hs028:~$
```

EXERCISE 5 (10 MINUTES) RUNNING JOBS





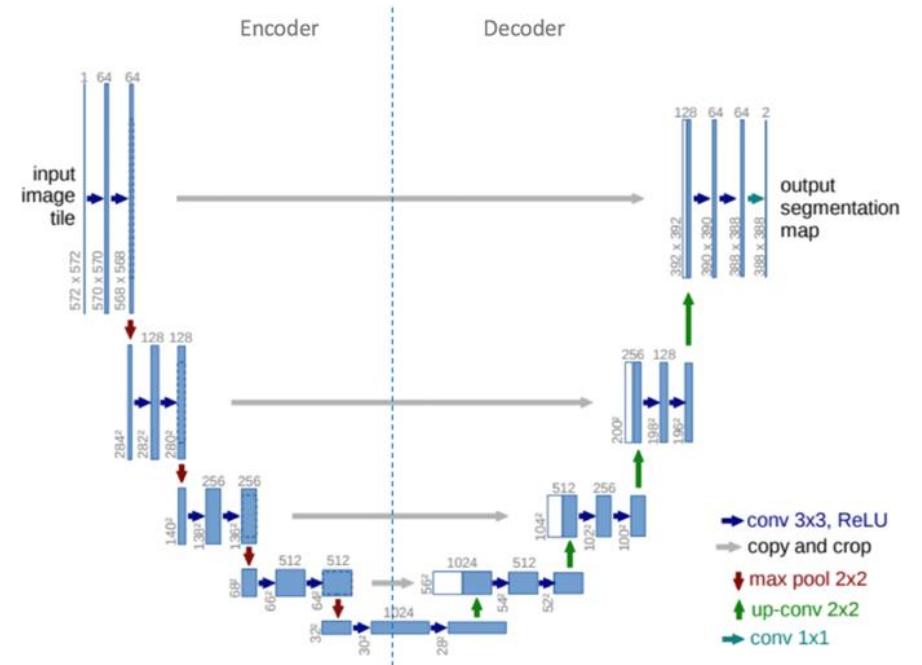
CHAPTER

BIOIMAGING SUPPORT

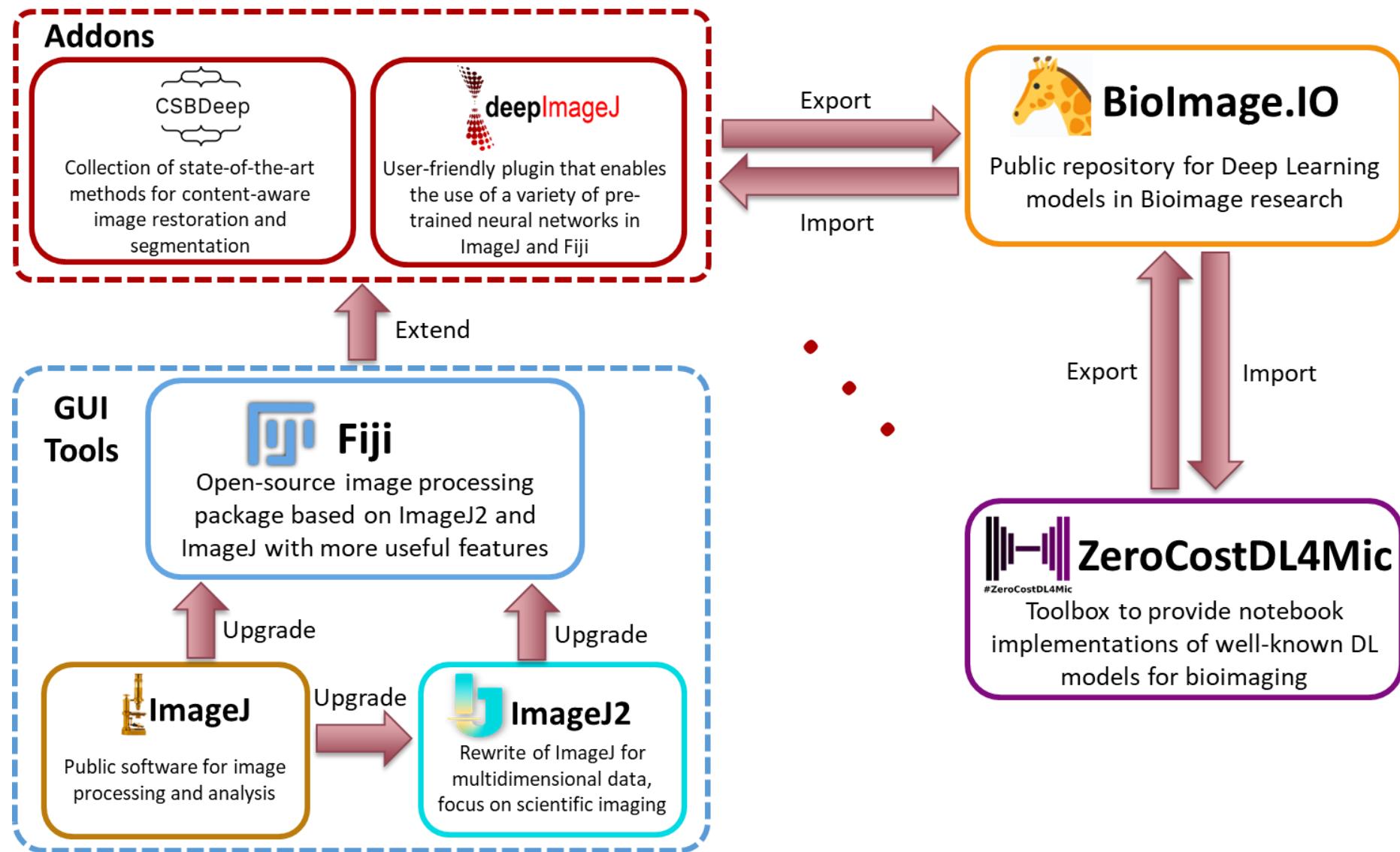


BIOIMAGING

- Overview of DL Algorithms in Bioimaging
 - Segmentation
 - Denoising and image restoration
 - Super-resolution microscopy
 - Object detection
 - Image-to-image translation
- Deep Learning Tools/Platforms in Bioimaging
 - Traditional ML tools: ImageJ, ImageJ2 and Fiji
 - DL tools: DeeplImageJ, CSBDeep, ZeroCostDL4Mic
 - Biolmage Model Zoo



SOFTWARE TOOLS



IMAGEJ AND FIJI

- On local PC, users can go to Fiji's website and download the corresponding version of Fiji
- On AHPCC server, users can request an interactive desktop session and use ImageJ/Fiji as local PC

The link to Fiji website

Fiji is a distribution of ImageJ which includes many useful plugins contributed by the community.

~ Download Fiji for your OS ~

Windows 64-bit	imagej.net (USA) , micron.ox.ac.uk (European mirror)
Windows 32-bit	imagej.net (USA) , micron.ox.ac.uk (European mirror)
macOS (x86_64)	imagej.net (USA) , micron.ox.ac.uk (European mirror)
Linux (64-bit)	imagej.net (USA) , micron.ox.ac.uk (European mirror)
No JRE	imagej.net (USA) , micron.ox.ac.uk (European mirror)

Choose version that matches to your PC

IMAGEJ AND FIJI

- On AHPCC server, users can download and install ImageJ and Fiji through **wget** command

```
pinnacle-12:haovan:~$ wget https://downloads.imagej.net/fiji/latest/fiji-linux64.zip  
--2023-03-15 15:05:27-- https://downloads.imagej.net/fiji/latest/fiji-linux64.zip  
Resolving downloads.imagej.net (downloads.imagej.net)... 144.92.48.183  
Connecting to downloads.imagej.net (downloads.imagej.net)|144.92.48.183|:443... □
```

- Looks like the above command does not work ...
- Note:
 - AHPCC servers has 14 gateway nodes (pinnacle-l1, pinnacle-l2,...,pinnacle-l14)
 - Nodes pinnacle-l[5,6,7,12,13,14] can access to outside Internet
 - Others cannot do that

```
Last login: Wed Mar 15 14:53:23 2023 from ood-portal-dmz  
  
NOTE: pinnacle-12 cannot route outside of UAF campus network  
For git/wget/conda operations please use pinnacle-l[5,6,7,12,13,14]
```

IMAGEJ AND FIJI

- If users want to get data from outside campus, change to nodes
pinnacle-l[5,6,7,12,13,14]
- Then, users can use commands to transfer/download data as normal

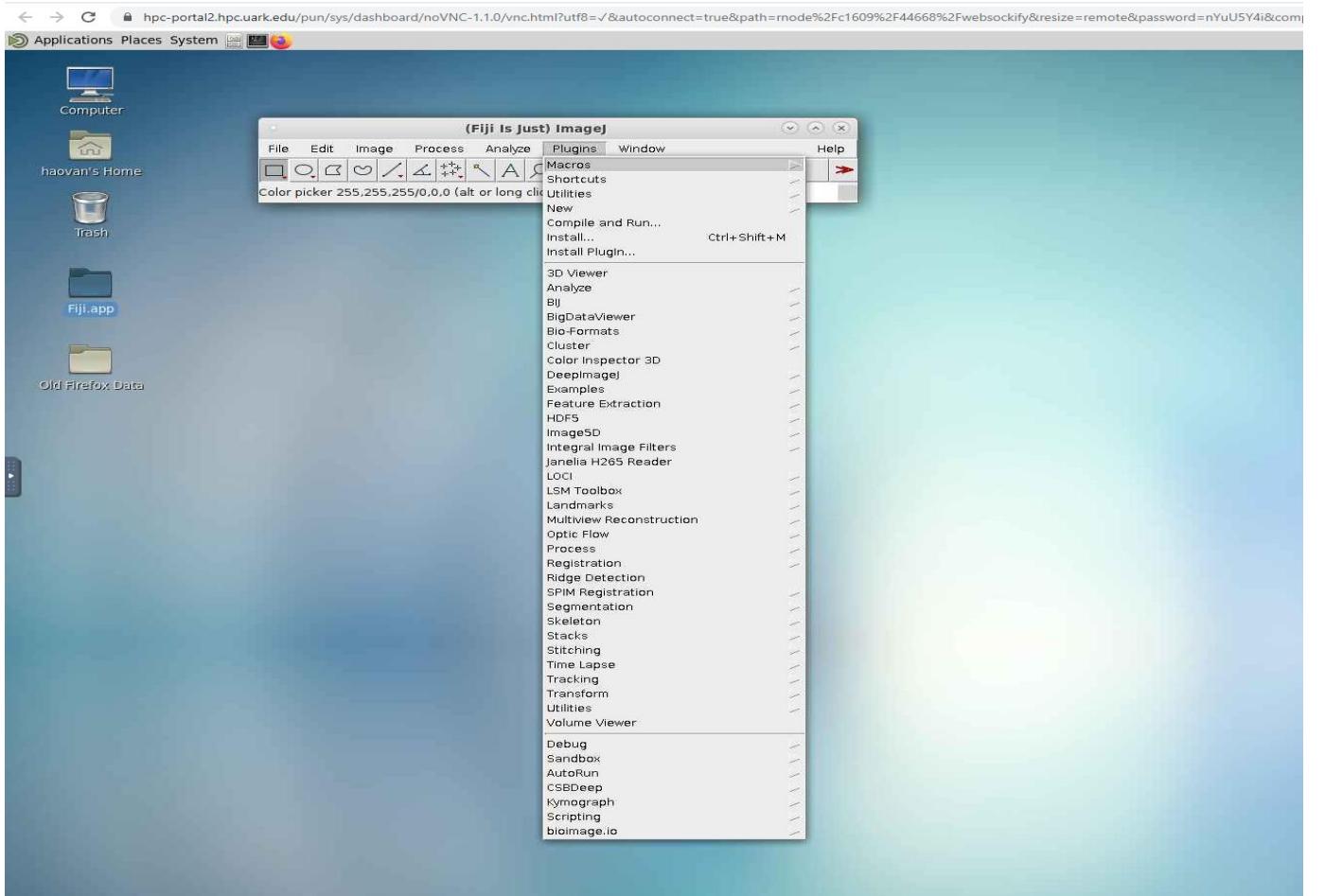
```
pinnacle-12:haovan:~$ ssh pinnacle-15
Warning: Permanently added 'pinnacle-15,172.16.16.55' (ECDSA) to the list of known hosts.

Currently Loaded Modules:
 1) os/el7

pinnacle-15:haovan:~$ wget https://downloads.imagej.net/fiji/latest/fiji-linux64.zip
--2023-03-15 15:13:53--  https://downloads.imagej.net/fiji/latest/fiji-linux64.zip
Resolving downloads.imagej.net (downloads.imagej.net)... 144.92.48.183
Connecting to downloads.imagej.net (downloads.imagej.net)|144.92.48.183|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 359332907 (343M) [application/zip]
15% [=----->
```

- Request an interactive desktop session on AHPCC and run the Fiji app
- Install plugins and run image processing or deep learning methods on server

IMAGEJ AND FIJI



EXERCISE 6 (5 MINUTES) FIJI



ZEROCOSTDL4MIC

- A toolbox to provide notebook implementations of well-known models and deploy them in Google Colab.
- Two use cases
 - Online: No need to handle Python environments or GPU setup when using Google Colab. However, the resources are limited with free subscription.
 - Offline: Using local machines or AHPCC server to run the notebook.

ZeroCostDL4Mic: exploiting Google Colab to develop a free and open-source toolbox for Deep-Learning in microscopy

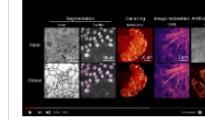
Tl;dr: this [wiki page](#) has everything you need to get started.

[DOI 10.5281/zenodo.5059813](#)

What is this?

ZeroCostDL4Mic is a collection of self-explanatory Jupyter Notebooks for [Google Colab](#) that features an easy-to-use graphical user interface. They are meant to quickly get you started on learning to use deep-learning for microscopy. Google Colab itself provides the computations resources needed at no-cost. ZeroCostDL4Mic is designed for researchers that have little or no coding expertise to quickly test, train and use popular Deep-Learning networks.

Want to see a short video demonstration?

Running a ZeroCostDL4Mic notebook	Example data in ZeroCostDL4Mic	Romain's talk @ Aurox conference	Talk @ SPAOM
			

Who is it for?

Any researcher interested in microscopy, independent of their background training. ZeroCostDL4Mic is designed for anyone with little or no coding expertise to quickly test, train and use popular Deep-Learning networks used to process microscopy data.

ZEROCOSTDL4MIC

Segmentation networks

Network	Paper(s)	Tasks	Status	Last test	Link to example training and test dataset	Direct link to the notebook in Colab
U-Net (2D)	here and here	Binary segmentation	Fully supported	⚠ broken (no GPU) (GJ)	here	Open in Colab
U-Net (3D)	here	Binary segmentation	Fully supported	⚠ broken (no GPU) (GJ)	EPFL dataset	Open in Colab
U-Net (2D) multilabel	here and here	Semantic segmentation	Under beta-testing	⚠ broken (no GPU) (GJ)	here	Open in Colab
DenoiSeg	here	Joint denoising and binary segmentation	Fully supported	⚠ broken (no GPU) (GJ)	Available soon	Open in Colab
StarDist (2D)	here and here	Instance segmentation	Fully supported	08/10/22 ✓ working (GJ)	here	Open in Colab
StarDist (3D)	here and here	Instance segmentation	Fully supported	07/10/22 ✓ working (GJ)	from Stardist github	Open in Colab
Celpose (2D and 3D)	here	Instance segmentation (Cells or Nuclei)	Fully supported	08/10/22 ✓ working (GJ)	Coming soon!	Open in Colab
SplineDist (2D)	here	Instance segmentation	Fully supported	07/10/22 ✓ working (GJ)	here	Open in Colab
EmbedSeg (2D)	here	Instance segmentation	Under beta-testing	⚠ broken by recent updates	here	Open in Colab
MaskRCNN (2D)	here	Instance segmentation	Under beta-testing		Coming soon!	Open in Colab
Interactive Segmentation - Kaibu (2D)	here	Interactive instance segmentation	Under beta-testing		Coming soon!	Open in Colab

Denoising and image restoration networks

Network	Paper(s)	Tasks	Status	Last test	Link to example training and test dataset	Direct link to the notebook in Colab
Noise2Void (2D)	here	Self-supervised denoising	Fully supported	06/10/22 ✓ working (EGM)	here or here	Open in Colab
Noise2Void (3D)	here	Self-supervised denoising	Fully supported	07/10/22 ✓ working (GJ)	here	Open in Colab
CARE (2D)	here	Supervised denoising	Fully supported	07/10/22 ✓ working (GJ)	here or here	Open in Colab
CARE (3D)	here	Supervised denoising	Fully supported	07/10/22 ✓ working (GJ)	here	Open in Colab
3D-RCAN	here	Supervised denoising	Under beta-testing	⚠ broken (no GPU)	here	Open in Colab
DecoNoise (2D)	here	Self-supervised denoising	Under beta-testing	07/10/22 ✓ working (GJ)	here or here	Open in Colab

Start notebook with Google Colab

Details about model and data

Super-resolution microscopy networks

Network	Paper(s)	Tasks	Status	Last test	Link to example training and test dataset	Direct link to the notebook in Colab
Deep-STORM	here	Single Molecule Localization Microscopy (SMLM) image reconstruction from high-density emitter data	Fully supported	08/10/22 ✓ working (G)	Training data simulated in the notebook or available from here	Open in Colab
DFCAN	here	image upsampling	Under beta-testing	08/10/22 ✓ working (G)	here	Open in Colab
WGAN	here	image upsampling	Under beta-testing	22/09/22 ✓ working (IvanHidalgo & EGM)	here	Open in Colab

ZEROCOSTDL4MIC

colab.research.google.com/github/HenriquesLab/ZeroCostDL4Mic/blob/master/Colab_notebooks/U-Net_2D_ZeroCostDL4Mic.ipynb#scrollTo=zCvebubeSaGY

U-Net_2D_ZeroCostDL4Mic.ipynb

File Edit View Insert Runtime Tools Help

View on GitHub Code + Text Copy to Drive

New notebook Open notebook Upload notebook

Rename Save a copy in Drive Save a copy as a GitHub Gist Save a copy in GitHub

Save Revision history

Download .ipynb Download .py

Download .ipynb

Notebook files can be downloaded here

Ignore the following message error message. Your Runtime has automatically restarted. This is normal.

1.1. Install key dependencies

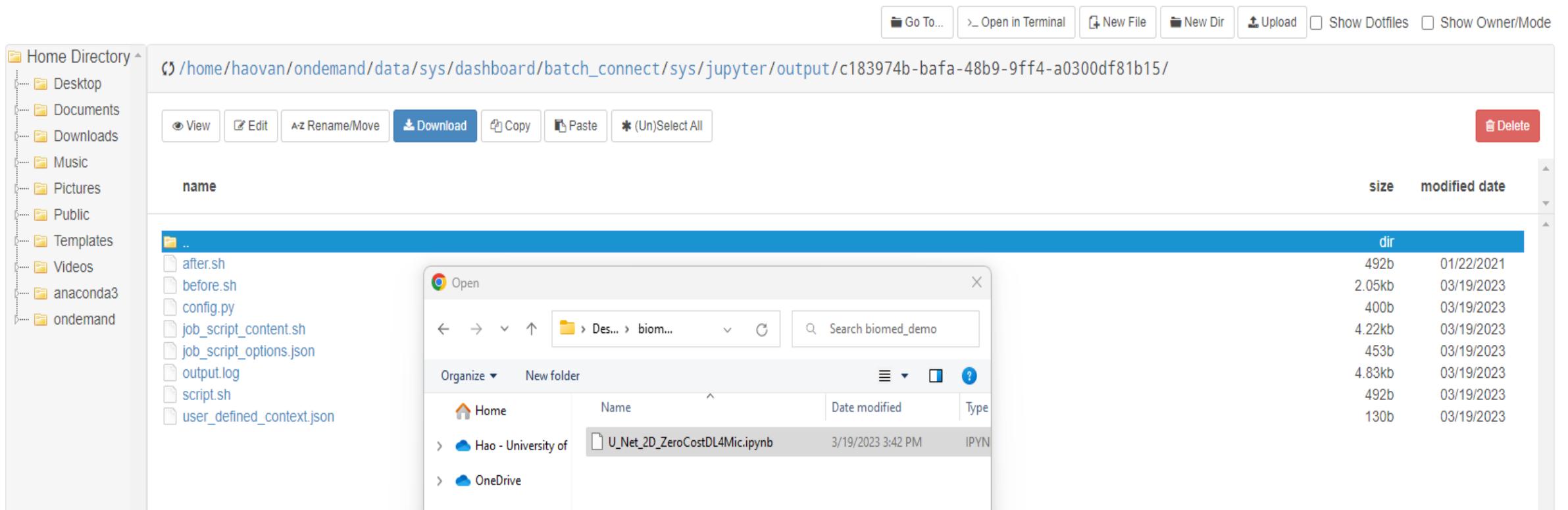
Play to install U-Net dependencies Show code

1.2. Mount your Google Drive

Your session crashed for an unknown reason. [View runtime logs](#)

ZEROCOSTDL4MIC

- Upload downloaded notebook to AHPCC via OOD portal



ZEROCOSTDL4MIC

- Now, users can open the notebook on AHPCC and modify/run the code

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** hpc-portal2.hpc.uark.edu/node/c1713/31064/notebooks/U_Net_2D_ZeroCostDL4Mic.ipynb
- Header:** jupyter U_Net_2D_ZeroCostDL4Mic Last Checkpoint: a few seconds ago (autosaved)
- Toolbar:** File, Edit, View, Insert, Cell, Kernel, Widgets, Help, Trusted, Python 3 (ipykernel)
- Content Area:**
 - Section:** U-Net (2D)
 - Description:** U-Net is an encoder-decoder network architecture originally used for image segmentation, first published by [Ronneberger et al.](#). The first half of the U-Net architecture is a downsampling convolutional neural network which acts as a feature extractor from input images. The other half upsamples these results and restores an image by combining results from downsampling with the upsampled images.
 - Note:** This particular notebook enables image segmentation of 2D dataset. If you are interested in 3D dataset, you should use the 3D U-Net notebook instead.
 - Disclaimer:** This notebook is part of the Zero-Cost Deep-Learning to Enhance Microscopy project (https://github.com/HenriquesLab/DeepLearning_Collab/wiki). Jointly developed by the Jacquemet (link to <https://cellmig.org/>) and Henriques (<https://henriqueslab.github.io/>) laboratories. The BioImage Model Zoo export was jointly developed by [Estibaliz Gómez de Mariscal](#) (deepImageJ team).
 - Citation:** This notebook is largely based on the papers:
 - U-Net: Convolutional Networks for Biomedical Image Segmentation** by Ronneberger et al. published on arXiv in 2015 (<https://arxiv.org/abs/1505.04597>)
 - U-Net: deep learning for cell counting, detection, and morphometry** by Thorsten Falk et al. in Nature Methods 2019 (<https://www.nature.com/articles/s41592-018-0261-2>) And source code found in: <https://github.com/zhihuao/unet> by Zhihuao
 - Guidelines:** The guidelines to use the trained network in ImageJ with deepImageJ are given in the following paper:
 - DeepImageJ: a user-friendly environment to run deep learning models in ImageJ**, bioRxiv (2019) by [Estibaliz Gómez-de-Mariscal, Carlos García-López-de-Haro, Wei Ouyang, Laurène Donati, Emma Lundberg, Michael Unser, Arrate Muñoz-Barrutia and Daniel Sage](#) (<https://doi.org/10.1101/799270>)
 - Footnote:** Please also cite this original paper when using or developing this notebook.

ZEROCOSTDL4MIC

For U-Net to train, it needs to have access to a paired training dataset corresponding to images and their corresponding masks. Information on how to generate a training dataset is available in our Wiki page:

<https://github.com/HenriquesLab/ZeroCostDL4Mic/wiki>

We strongly recommend that you generate extra paired images. These images can be used to assess the quality of your trained model (**Quality control dataset**). The quality control assessment can be done directly in this notebook.

Additionally, the corresponding Training_source and Training_target files need to have **the same name**.

Here's a common data structure that can work:

- Experiment A
 - **Training dataset**
 - Training_source
 - img_1.tif, img_2.tif, ...
 - Training_target
 - img_1.tif, img_2.tif, ...
 - **Quality control dataset**
 - Training_source
 - img_1.tif, img_2.tif
 - Training_target
 - img_1.tif, img_2.tif
 - **Data to be predicted**
 - **Results**

Important note

- If you wish to **Train a network from scratch** using your own dataset (and we encourage everyone to do that), you will need to run **sections 1 - 4**, then use **section 5** to assess the quality of your model and **section 6** to run predictions using the model that you trained.

- If you wish to **Evaluate your model** using a model previously generated and saved on your Google Drive, you will only need to run **sections 1 and 2** to set up the notebook, then use **section 5** to assess the quality of your model.

- If you only wish to run predictions using a model previously generated and saved on your Google Drive, you will only need to run **sections 1 and 2** to set up the notebook, then use **section 6** to run the predictions on the desired model.

- Users should pay attention to the instruction in each notebook
- Each notebook needs different packages in Python to execute
- Moreover, the data directory needs to be consistent to the one shown in the introduction

EXERCISE 7 (15+ MINUTES)

MACHINE LEARNING

!CHANGE KERNEL TO “PYTORCH-CUDA-3.14”



THANK YOU!

CHECK MORE INFORMATION AT ARKANSAS HPCC USER SUPPORT WIKI
[HTTPS://HPCWIKI.UARK.EDU/DOKU.PHP](https://hpcwiki.uark.edu/doku.php)

FOR QUESTIONS, EMAIL TO (PRATEEK, XINTAOWU) @ UARK.EDU