Getting Started with AHPCC

Dr. Recep Erol
HPC Admin & AI Facilitator
October 3, 2024

About Me

- HPC Admin and AI Facilitator
- AI/ML | Optimization | Inference Acceleration | Model Deployment
- Parallel Computing | Distributed Computing | Hardware Acceleration

rerol@uark.edu

erolrecep.github.io

@erolrecep

Agenda for Day 1

GettingStartedwithAHPCC Private

Unwatch 1

main · 1 Branch · 0 Tags · Go to file t · Add file · Code

erolrecep final version of the repository is ready — e012e78 · 3 hours ago · 14 Commits

Linux101	initial draft of day1	yesterday
SLURM_Hands-on	final version of the repository is ready	3 hours ago
SLURM_and_Module	final version of the repository is ready	3 hours ago
images	media were added	yesterday
.gitignore	mac cache files are removed	yesterday
README.md	schedule and plan is getting more and more available	8 hours ago

README

Getting Started with AHPCC

Arkansas High Performance Computing Center

Linux 101: Hands-On Training

In this training, we are going to learn the basics of how to use a linux based operating system (OS) operations using command environment, terminal in an HPC system. We will learn from creating a file to moving a folder to another folder, downloading a file from internet and more. Alright, let's get started.

In this hands-on session, you will learn what is terminal, why do we need it, and different from other operating systems, terminal is very important for the Linux-based OS. You can think of, terminal is a kind of interface between the user and the operating system's brain.

Many HPC users like to develop projects in HPC environment. To help new users to get used to this habit, we will build a project template together to help you how you can organize files/folder for your project in the HPC system.

If everyone is ready to start, let's get started!

Make sure to open the interactive GUI desktop application and then open a terminal session from the toolbar. Then follow the instructions below.

Since terminal shows us where we are, we like to see the full path with;

```
$ pwd
```

Change directory to where you want to create your project template

```
$ cd /home/rerol/Desktop
```

Create the project folder

```
$ mkdir python-project-template
```

List the directory to see if we actually created the folder ???

```
$ ls
```

Day 1 Learning Goals

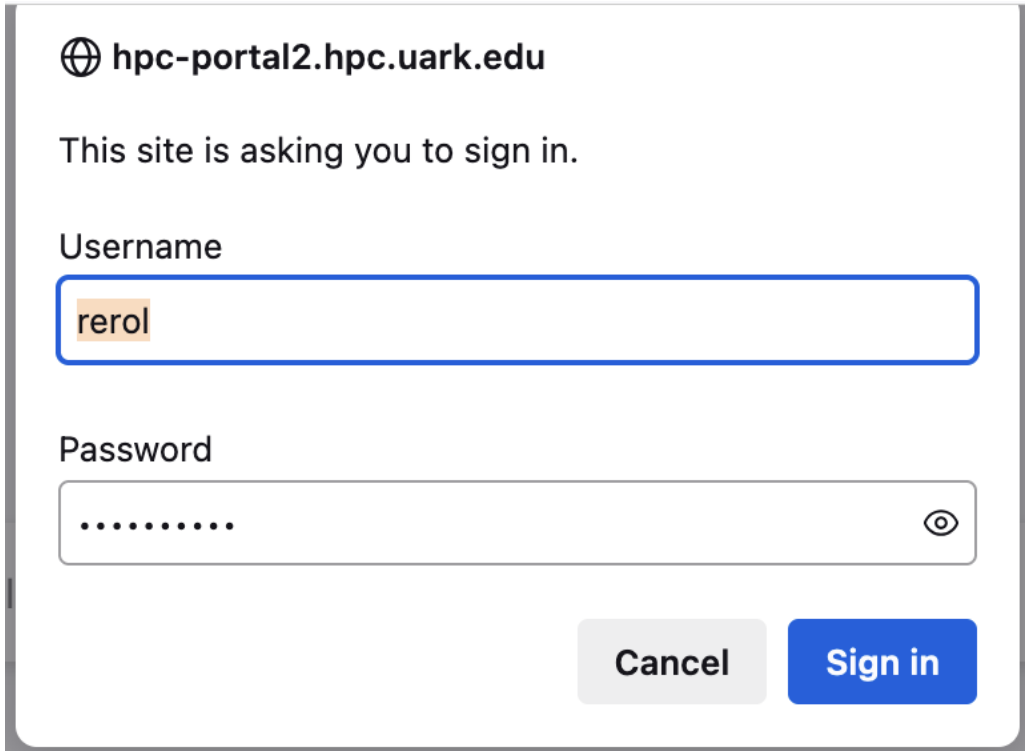
- Introduce HPC Portal
- Login system
- Login & Compute terminal contexts
- Hands-on Linux 101 training
- Clone a repository from GitHub and run in the terminal

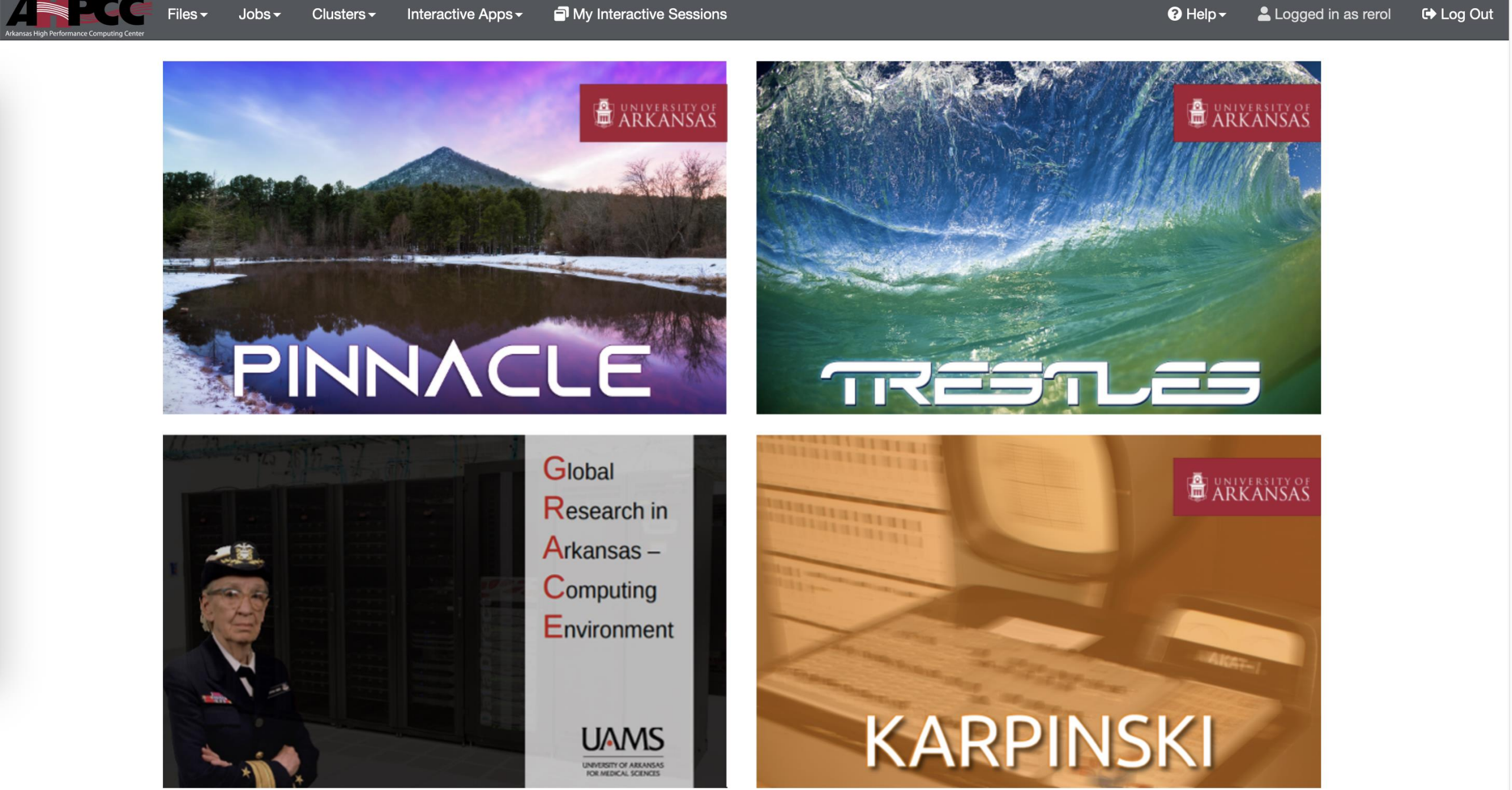
Check First

- You have a laptop?
- You are connected to ***UARK WIFI***, not ~~UARK Guest~~!
- Your browser is working?

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

OnDemand Portal - Login



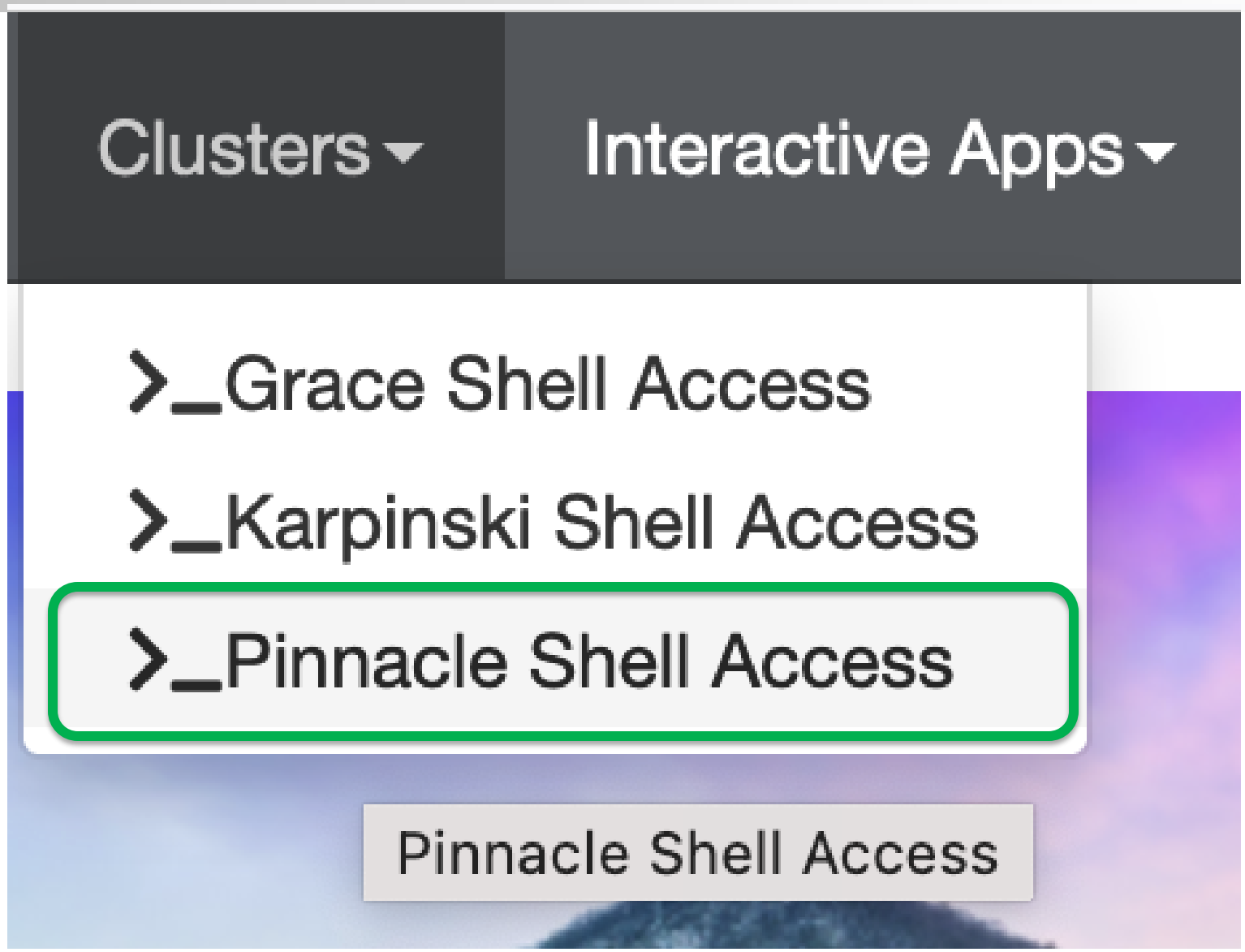


OnDemand provides an integrated, single access point for all of your HPC resources.

Let's Create Instances

- Pinnacle Desktop
- Jupyter Notebook

Portal Terminal



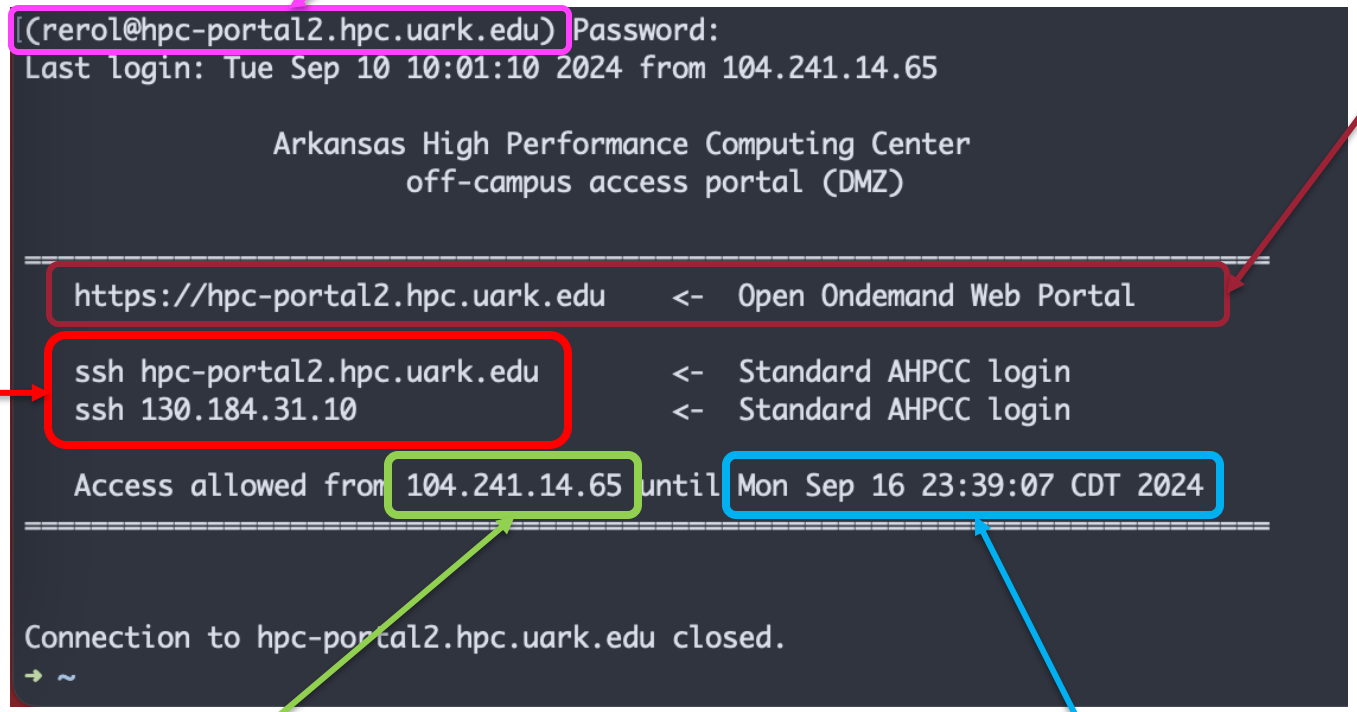
Portal Terminal

```
  INFO: this login node has **no access to external network**
        for wget, git, operations please use any one of pinnacle-l[3-5,8-10]
        or cloud compute nodes.

  Login nodes are for file operations and job submission.
  All login nodes run a different kernel than compute nodes, so:
    - Don't compile on login nodes. Resulting code won't work on compute nodes.
    - Don't run R or Python on login nodes.

  For a shared cloud compute node:
     srun -N1 -n1 -c1 -p cloud72 -q cloud -t 72:00:00 --pty /bin/bash


Last login: Wed Sep 25 11:30:33 2024 from 172.16.15.63
Quota   kilobytes       files
/scrfs  10,000,000,000  2,000,000
rerol   4,467,140       56,737

Currently Loaded Modules:
  1) null



pinnacle-l2:rerol:~$
```

First attempt:
ssh rerol@portal2.hpc.uark.edu –p 2022

1. Login to the AHPCC via link:
   - ssh hpc-portal2.hpc.uark.edu
2. Login via IP address:
   - ssh 130.184.31.10

Web portal access link is also given here!



Your IP address

You're allowed to login to the system for the next 12 hours.

Now, we're in a "Login Node"

Create a shared cloud compute job

```
[rerol@hpc-portal2.hpc.uark.edu's password:

  Login nodes are for file operations and job submission.
  All login nodes run a different kernel than compute nodes, so:
    - Don't compile on login nodes. Resulting code won't work on compute nodes.
    - Don't run R or Python on login nodes.

  For a shared cloud compute node:
     srun -N1 -n1 -c1 -p cloud72 -q cloud -t 72:00:00 --pty /bin/bash

Last login: Fri Sep 13 11:01:48 2024 from 104.241.14.65
Quota   kilobytes       files
/scrfs  10,000,000,000  2,000,000
rerol   4,074,752       43,216

Currently Loaded Modules:
  1) null


pinnacle-l8:rerol:~$
```

You can submit a job from a login node!

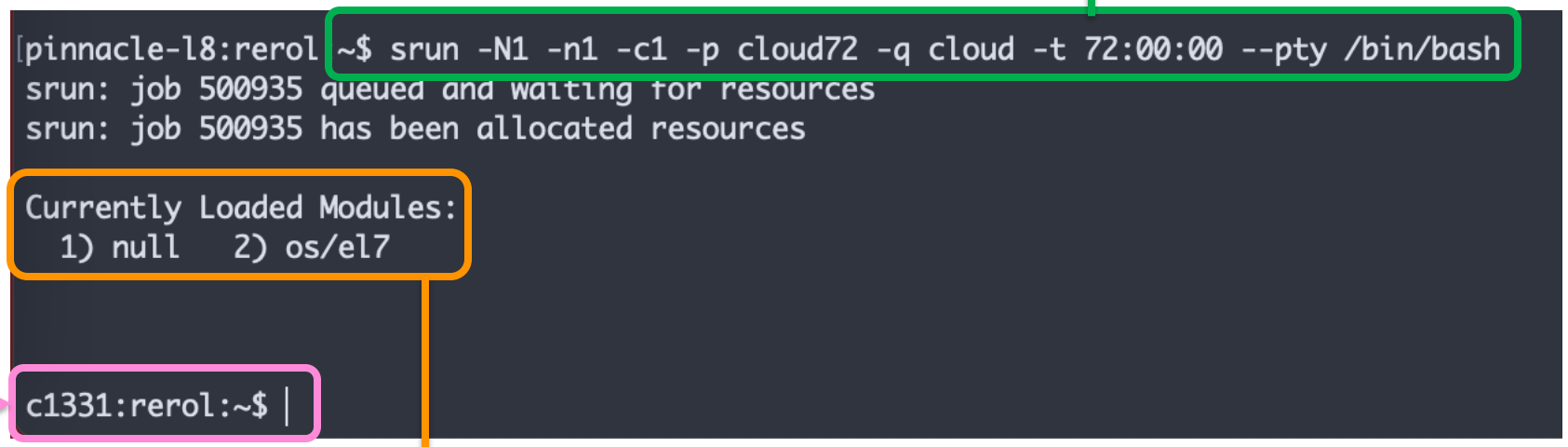
srun job.sh
or
Srun job.slurm

Account storage information

The system assigned me to the "pinnacle-18 login node"

SLURM assigned me to this compute node

Create a compute node job



Which software/packages are loaded?

Command	Explanation
Srun	Submit job or allocate resources in SLURM
-N1	Number of nodes in the HPC system
-n1	Number of tasks
-c1	Number of CPU core
-p	Partition name from SLURM resources
cloud72	Name of the partition in the SLURM
-q	Job queue type. In this case "cloud" type
cloud	Job queue type.
-t	Time parameter for "Maximum runtime"
72:00:00	72 hours
--pty	"psuedo-terminal" for interactive terminal session
/bin/bash	Type of shell where the job will be submitted

```
[pinnacle-l8:rerol:~$ srun -N1 -n1 -c1 -p cloud72 -q cloud -t 72:00:00 --pty /bin/bash
```

Linux 101 – Hands-on Demo

- Create Python Project Template
  - Instructions are given in the GitHub repository
  - You can use browser-based terminal
  - Or Visual Studio Code --> Shown on the next slide

❖ Please ask any questions you may have

https://code.visualstudio.com/docs/remote/ssh

Version 1.93 is now available! Read about the new features and fixes from August.

Overview

SETUP

GET STARTED

USER GUIDE

SOURCE CONTROL

TERMINAL

GITHUB COPILOT

LANGUAGES

NODE.JS / JAVASCRIPT

TYPESCRIPT

PYTHON

JAVA

C++

C#

DOCKER

DATA SCIENCE

AZURE

REMOTE

- Overview
- SSH
- Dev Containers

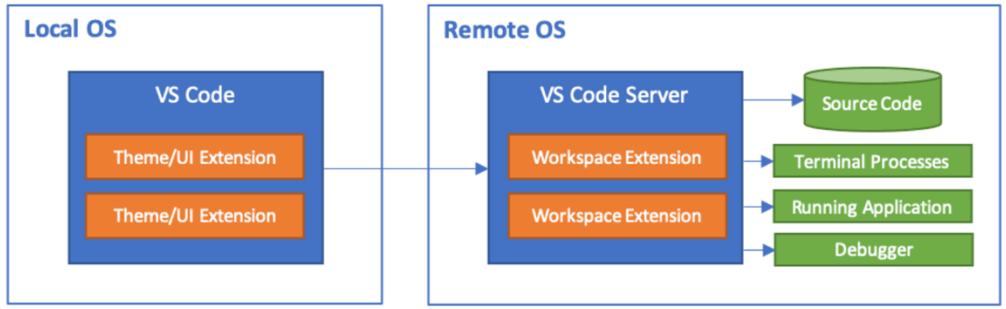
VS Code Remote Development

Edit

Visual Studio Code Remote Development allows you to use a container, remote machine, or the Windows Subsystem for Linux (WSL) as a full-featured development environment. You can:

- Develop on the **same operating system** you deploy to or use **larger or more specialized** hardware.
- **Separate** your development environment to avoid impacting your local **machine configuration**.
- Make it easy for new contributors to **get started** and keep everyone on a **consistent environment**.
- Use tools or runtimes **not available** on your local OS or manage **multiple versions** of them.
- Develop your Linux-deployed applications using the **Windows Subsystem for Linux**.
- Access an **existing** development environment from **multiple machines or locations**.
- Debug an **application running somewhere else** such as a customer site or in the cloud.

No source code needs to be on your local machine to get these benefits. Each extension in the Remote Development extension pack can run commands and other extensions directly inside a container, in WSL, or on a remote machine so that everything feels like it does when you run locally.



IN THIS ARTICLE

- Getting started
- Remote tutorials
- GitHub Codespaces
- Questions or feedback

Subscribe

Ask questions

Follow @code

Request features

Report issues

Watch videos

List of commands we use in the hands-on session

- pwd
- cd
- mkdir
- ls
- touch

Getting Started with AHPCC

Dr. Recep Erol
HPC Admin & AI Facilitator
October 4, 2024

Day 2

Welcome Back!

What We Learned Yesterday

- Interactive Portal
  - Created a remote desktop environment
  - Login to the system through terminal
  - Learned about file system and other information
- Hands-on experience on Linux 101
- Cloned GitHub repository and run the project

Agenda for Day 2

SLURM and Module

In this part of the workshop, we will introduce SLURM and Module software to you.

SLURM and Module are the two very important software we use every day to manage our compute environment and jobs to submit in an HPC system. Once you grasp these software, the rest of the workflow will be a lot easier for you in an HPC systems.

This README file is a kind of guideline for the hands-on exercises for this part of the workshop.

What is Module? and Why we use it?

High Performance Computing (HPC) or Supercomputing systems are shared environments. In other words, multiple users use these systems together. Everybody has their own setup to work on different tasks. To be able to manage these demands, there should be a mechanism where user should be able to manage their software requirements and dynamic modifications. Module allows these changes and modifications on the software.

There are number of software and almost each software has multiple versions. It's almost impossible to install all these software at the same time to a computer or compute environment. Very possible to have conflicts between software and some will prevent others to run. These and a lot more issues would easily be solved if we use **Module** for our software management.

What is SLURM? and Why we use it?

Simple Linux Resource Management or Slurm Workload Management is a software that is designed and provided to schedule your jobs, to manage your resources including CPU's, GPU's, RAM, or storage, or monitor submitted jobs to an HPC system. As we can guess from the name of the software, SLURM is a very useful software for multiple purposes. There are many commands that you can use and these commands are introduced in the hands-on exercises.

By completing the hands-on exercises in this part of the workshop, you're going to have a solid understanding of how you can manage your compute environment and manage your jobs in an HPC.

Hands-on Exercises

- Module Hands-on Exercise
- SLURM Hands-on Exercise

Good Luck!

Day 2 Learning Goals

- Introduce HPC Software
  - SLURM & Module
- Resource management and job submission with SLURM
- Software management with Module
- Hands-on training with,
  - Module
  - SLURM

Check First

- You have a laptop?
- You are connected to ***UARK WIFI***, not ~~UARK Guest~~!
- Your browser is working?

SLURM & Module

ENVIRONMENT
MODULES

What is Module?

- HPC is a shared environment
- There are a lot of software and difficult to manage all at once

Hands-on Module Training

- Follow instructions in GitHub repository!

What is SLURM?

- The Slurm Workload Management or Simple Linux Utility Resource Management (SLURM)
- Job scheduler
- Create interactive compute job with “srun”
- Job monitoring

SLURM Hands-on Training

- Follow instructions in GitHub repository!

Recap for Today

- Introduce HPC Software
  - SLURM & Module
- Resource management and job submission with SLURM
- Software management with Module
- Hands-on training with,
  - Module
  - SLURM

Getting Started with AHPCC

Dr. Recep Erol
HPC Admin & AI Facilitator
October 3 & 4, 2024