Open OnDemand, Open XDMoD, and ColdFront: An HPC center management toolset

WELCOME!

IF YOU HAVE NOT ALREADY DONE SO, PLEASE FOLLOW SETUP INSTRUCTIONS!

https://github.com/ubccr/hpc-toolset-tutorial

Join the Slack organization for the tutorial

https://tinyurl.com/pearc-slack
Open OnDemand, Open XDMoD, and ColdFront: An HPC center management toolset

Tutorial presented at PEARC22 by staff from:
Ohio Supercomputer Center
UB Center for Computational Research
Virginia Tech Advanced Research Computing
Tutorial Staff:

Andrew Bruno, UB            Matt Jones, UB
Gerald Byrket, OSC           Jeff Ohrstrom, OSC
Alan Chalker, OSC            Ryan Rathsam, UB
Andrew Collins, OSC          Travis Ravert, OSC
Robert DeLeon, UB            Dori Sajdak, UB
Trey Dockendorf, OSC         Bob Settlage, VT
David Hudak, OSC             Joseph White, UB
Agenda

• Getting Started & Tutorial Goals
• Brief intro on all three products
• Tutorial technology
• Part 1: ColdFront
• Break: 10-10:30am
• Part 2: Open XDMoD
• Lunch Break – 12-1:30pm
• Part 3: Open OnDemand
• Break: 3-3:30pm
• Part 4: Open OnDemand interactive app configuration
• Part 5: Dynamic Batch Connect Fields
• Post Workshop – breakout sessions & slack channel
Getting Started

• Join the Slack organization for the tutorial
  https://tinyurl.com/pearc-slack

• Clone the tutorial repo and follow instructions for starting containers
  https://github.com/ubccr/hpc-toolset-tutorial

• What to do if you’re having a technical problem – Slack us or raise your hand & we’ll do our best to help out
Tutorial Goals:

- Provide participants with an overview of each product & how they are installed/configured
- Point out a few “gotcha!”s to look out for
- Give participants a cluster in a container to practice using these products
- Supply participants access to the developers of these products as a resource for questions & help
- Show off the new features that allow the products to work together
ColdFront – Managing Access

- Used as the source of record in a HPC center to ensure security & continuity of the systems
- Provides center staff ability to manage center resources & who has access to them
- Portal for users to manage their access to center resources & report on their research
- Plug-ins for job scheduler, central authentication, job statistics (XDMoD), OnDemand, that enable automation of access to or removal from resources
- Reports for center management to demonstrate the center’s impact (publications, grants, research output)
Open OnDemand – Easy Access

• Web-based portal for accessing HPC services that removes the complexities of HPC system environments from the end-user

Includes:

• Files app for upload/download & editing of files
• Terminal app (no need for separate SSH client)
• Job app to create/edit/submit/monitor jobs
• Interactive apps to run GUI applications. Users can create and share apps. Centers can publish apps for all users
Open XDMoD – Usage & Performance Metrics

- Tool that aggregates job data & system performance metrics to inform system users, system staff & center decision makers
- Web portal providing job & system metrics, including utilization, quality of service metrics designed to proactively identify underperforming system hardware and software, and job level performance data for every job
- Role-based access to data with different levels of granularity, including job, user, or on a system-wide basis
- Ingest OnDemand logs into new OnDemand realm in XDMoD
Tutorial Container Architecture

Requirements: [https://github.com/ubccr/hpc-toolset-tutorial/edit/master/docs/requirements.md](https://github.com/ubccr/hpc-toolset-tutorial/edit/master/docs/requirements.md)

Clone the Github Repo:

git clone [https://github.com/ubccr/hpc-toolset-tutorial](https://github.com/ubccr/hpc-toolset-tutorial)
cd hpc-toolset-tutorial
./hpcts start

* The first time you do this, you'll be download ~20GB worth of containers from Docker Hub. This can take a long time depending on your network speeds. After downloaded, the containers are started and services launched.

WARNING!!! DO NOT run these containers on production systems. This project is for educational purposes only. The container images we publish for the tutorial are configured with hard coded insecure passwords and should be run locally in development for testing and learning only.
Tutorial Walk Through

https://github.com/ubccr/hpc-toolset-tutorial

Keep the applications page open for easy access to account usernames/passwords and portal URLs:

https://github.com/ubccr/hpc-toolset-tutorial/blob/master/docs/applications.md
OpenSource HPC resource allocation portal for users, system admins, & center staff

Tutorial presented at PEARC22 by:
Andrew Bruno, UB
Dori Sajdak, UB
Automate access to your HPC resources

Manage access to all your resources in one place
Why We Developed ColdFront:

• System Administrators wanted:
  • More automation, less manual error
  • One location for access management of all resources
  • Allow PIs to self-service access to resources

• Center Director wanted:
  • To require PIs to update project info annually
  • Consistent reporting of publication & grant info
  • Easy displays of usage for annual reporting
Resources

• Resources are anything you want to control access to and/or monitor usage of
• Resources might include:
  • clusters, storage platforms, cloud, servers, scientific instruments, databases, software licenses
• Resources have attributes
  • Some might enforce limits:
    • storage (GB),
    • software (seats),
    • cloud (subscriptions)
  • Some might restrict access & may tie to plugins:
    • Is the resource private or public? Available only to certain users/groups?
    • Cluster config options – Slurm plugin
    • System access restricted by UNIX group? - FreeIPA plugin
    • Warranty expiration dates
• These attributes are customizable
• The attributes set on resources are inherited by allocations
Allocations

- Determines what resource an account has access to
- Allocations have start & end dates, creation & last modified dates, status, description, associated resource(s) & require a justification
- Allocations have attributes like resources that may set limits, restrict access, and/or tie into the ColdFront plugins. Examples include:
  - CPU/core hours
  - Scheduler account name
  - UNIX group
  - Storage quota
- Like Resource attributes, these are customizable
- Users emailed when expiration dates approach – configurable time spans
- Resource access can be removed when an allocation expires using ColdFront plugins
- Allocation Change requests
Projects

• Project = users, allocations for resources, reportable data (publications, grants)
• PIs (group leads) can request allocations for resources, add/remove users on their project & allocations, upload research info, complete annual project review, view group usage
• Role based logins allow for:
  • full project access for PIs
  • additional capabilities for managers assigned by PIs,
  • read-only views for users,
  • HPC center staff have access to tools for:
    – Allocation review, approval, & configuration
    – Annual project review approval
    – Other policy-driven tools
# Allocation Requests & Change Requests Can be Viewed by System Administrators

## Allocation Requests

For each allocation request below, there is the option to activate the allocation request and to view the allocation's detail page.

By default, activating an allocation will make it active for 365 days.

<table>
<thead>
<tr>
<th>#</th>
<th>Requested</th>
<th>Project Title</th>
<th>PI</th>
<th>Resource</th>
<th>Project Review Status</th>
<th>Status</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2449</td>
<td>Apr. 28, 2022</td>
<td>Computational and Data Science and Engineering</td>
<td>Eric Walker (ericwalk)</td>
<td>UB VPN Access (Software License)</td>
<td>Approved</td>
<td>Approved</td>
<td>Details</td>
</tr>
<tr>
<td>2438</td>
<td>Apr. 21, 2022</td>
<td>Sunstar</td>
<td>Patricia Díaz (padiazmo)</td>
<td>UB VPN Access (Software License)</td>
<td>New</td>
<td>Approved</td>
<td>Details</td>
</tr>
</tbody>
</table>

## Allocation Change Requests

For each allocation change request below, there is the option to activate the allocation request and to view the allocation change's detail page. If a change request is only for an extension to the allocation, they can be approved on this page. However, if the change request includes changes to the allocation's attributes, the request must be reviewed and acted upon in its detail page.

<table>
<thead>
<tr>
<th>#</th>
<th>Requested</th>
<th>Project Title</th>
<th>PI</th>
<th>Resource</th>
<th>Extension</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Apr. 06, 2022</td>
<td>Evolution of genomes with a focus on structural v...</td>
<td>Omer Gökçemen (omergoko)</td>
<td>ProjectStorage (Storage)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>May. 02, 2022</td>
<td>Nonadiabatic dynamics in solar energy materials: ...</td>
<td>Alexey Akimov (alexeyak)</td>
<td>ProjectStorage (Storage)</td>
<td>30 days</td>
<td></td>
</tr>
</tbody>
</table>
Annual Project Reviews

You cannot request a new allocation because you have to review your project first.

You need to review this project. Review Project

Test Project

Manage Project

Completed Annual Project Reviews Can be Viewed by Center Director and System Admins

Pending Project Reviews

<table>
<thead>
<tr>
<th>Project Title</th>
<th>Date Review Submitted</th>
<th>PI</th>
<th>Grants Last Updated</th>
<th>Publications Last Updated</th>
<th>Reason for not Updating Project</th>
<th>Project Review Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>My Test Project</td>
<td>May. 13, 2021</td>
<td>Dori Sajdak (djm29)</td>
<td>May. 13, 2021</td>
<td>May. 13, 2021</td>
<td></td>
<td>Mark Complete, Email</td>
</tr>
</tbody>
</table>
Center Directors are able to better demonstrate the center’s impact

- Report on resources & allocations
- Collect publication information
- Collect grant information

**Grants**

- National Institutes of Health (NIH): $78,599,277 (33)
- National Science Foundation (NSF): $52,283,068 (73)
- Other: $12,161,778 (49)
Extensible plug-in architecture allows for integration of nearly anything!
Integrations

Plug-ins (Django Apps)

• OnDemand
• XDMoD
• Slurm
• Mokey/Hydra OpenID Connect (Identity Management)
• FreeIPA (LDAP/AD)
• Other plug-in examples
• Other 3rd party APIs should be added as a new plug-in (Django app)

Community Plugins:
OpenStack
Keycloak
User Search
Starfish
Tutorial Steps:

- Create different user roles and access
- Create new cluster resource
- As PI user, create project and request allocation for cluster resource
- As sys admin user, activate allocation and sync with Slurm
- As PI user, run batch & OnDemand job
- As PI users, request allocation change & allocation renewal
- Configure user with center director access & check out Project Review process
- Enable OnDemand integration
Contact Info:

Andrew Bruno - aebruno2@buffalo.edu
Dori Sajdak - djm29@buffalo.edu

https://coldfront.io

More about UB CCR:
https://buffalo.edu/CCR
https://twitter.com/UBCCR

ColdFront BOF:
Thursday, 7/14 9-10am
Arlington
Important Info:

Tutorial Repo: https://github.com/ubccr/hpc-toolset-tutorial

Join us on Slack: https://tinyurl.com/pearc-slack

IF YOU HAVE NOT ALREADY DONE SO, PLEASE FOLLOW SETUP INSTRUCTIONS!

NOW: Break: 90 minutes
Startup again at 1:30pm

Coming Up!
Part 3: Open OnDemand install and configuration
PM Break: 3-3:30
Part 4: Open OnDemand interactive app configuration
Part 5: Dynamic Batch Connect Fields

Other Places You’ll Find us at PEARC22:

Performance Optimization of the Open XDMoD Datawarehouse - best full paper! Tues, 7/12 10:30-11am

Open OnDemand User Group Meeting: Tues, 7/12 1:30-2:30pm

Open XDMoD BOF: Wed, 7/13 1:30-2:30pm

Enhancing User-centric Workflows and Democratizing Access to Novel Advanced Research Computing BoF, Thur 7/14 8-9am

ColdFront BOF: Thur, 7/14 9-10am
Open, Interactive HPC via the Web

Alan Chalker, OSC
Travis Ravert, OSC
Trey Dockendorf, OSC
Jeff Ohrstrom, OSC
Bob Settlage, VT
Gerald Byrket, OSC

Open OnDemand is looking for contributors in the community. If interested, please speak with one of the OOD Team Members mentioned in this slide.
Introduction to Open OnDemand

Alan Chalker - OSC
Use our Discourse instance for help

Join our mailing list for updates

Join our Monthly Open Office Forum


Our webinars are roughly quarterly

This work is supported by the National Science Foundation of the United States under the awards NSF SI2-SSE-1534949 and CSSI-Software-Frameworks-1835725.
Approx Number of Institutions based on RPM logs

- 136 unique US locations
- 70 unique international locations
Example Current Engagements and Deployments
CLIENT

Zero Install
Requires only a modern web browser

SERVER FRONT END
(Runs as Apache User)

Functions
1. User Authentication
2. Reverse Proxy

SERVER BACK ENDS
(Each Runs as an Authenticated User)

Per-User NGINX (PUN)

Passenger

Interactive Jobs

Interactive HPC (iHPC)
VNC Server + Websockify
COMSOL Server
Jupyter Notebook Server
RStudio Server

HTTPS/WSS
HTTP/WS
IPC Sockets
Open OnDemand 2.0 Project Overview

Previous three year NSF SI2 award (#1534949) to develop OnDemand 1.x

Awarded follow on NSF CSSI award (#1835725) to develop OnDemand 2.x
    Project runs from Jan 2019 to Dec 2023
    Collaborators include SUNY Buffalo and Virginia Tech

Four areas

  **Visibility:** Enhancing resource utilization visibility by integrating the existing Open XDMoD platform
  **Scalability:** support more types of computing resources and software
  **Accessibility:** appeal to more scientists in more fields of science
  **Engagement:** establish community of departmental, campus and national HPC users and administrators
Check out the Project on GitHub

See all the projects at: https://github.com/OSC/ondemand/projects

Large features currently stated for release: https://github.com/OSC/ondemand/projects/10

Feel free to comment or react to tickets. Open feature requests or anything. We want to hear from you!

If you have an idea or want to work on any of the features or bug fixes, please feel free to reach out to us and we will get you started. We always want to include our community in the Open OnDemand efforts.
Customizing the OnDemand Dashboard

Gerald Byrket - OSC
Hands on Tutorial: Dashboard in Development Mode

It Covers:

• Setting up the dashboard in development mode
• Changing the navigation bar color
• Pinning Apps to the dashboard
• Changing the layout of the dashboard
• Adding custom widgets to the dashboard
Dashboard Tutorial: Pinning Apps to the dashboard

- Pinning Apps and then grouping them
Dashboard Tutorial: Changing the layout

• Change the layout so that Message of the Day is on the left
Dashboard Tutorial: Adding a new widget

- Add a new custom widget
Create a Jupyter “Interactive App Plugin”

Travis Ravert - OSC
Hands on Tutorial: Create a Jupyter “Interactive App Plugin”

It Covers:

• Getting the app to work.
• Checking logfiles to debugging failures.
• Changing the types of form items
  • From text input to select widgets
• Adding new form options
• Using Native scheduler options
• Explanations of files
• Promoting the app to production
Jupyter Tutorial: Get the App working

- Jupyter example application doesn’t work out of the box
  - Configure it to use this cluster
  - Configure it to use the correct Jupyter installation

- The card is shown when a successful Jupyter job is launched
Jupyter Tutorial: Modify the Partition

- Change the partition element to be a select dropdown instead of a text field
Jupyter Tutorial: Deploy to production

- Deploy the app to production for other users
Jupyter Tutorial: Set the memory request for the job

- Use the `script.native` attributes to set the `--mem` SLURM argument

Memory (MB)

600

RSS Memory

Launch

* The HPC Tutorial Jupyter session data for this session can be accessed under the `data` root directory.
Jupyter Tutorial: Limit the number of nodes

- Put an upper limit on the number of nodes allowed

Number of nodes

17

Please select a value that is no more than 2.

* The HPC Tutorial Jupyter session data for this session can be accessed under the data root directory.
Jupyter Tutorial: Add a radio button to start JupyterLab

- Add radio buttons so users can boot JupyterLab or Jupyter Notebook

Mode
- Jupyter Lab
- Jupyter Notebook
Jupyter Tutorial: Delete unused fields

- Delete unused fields to clean up the form

**Partition**
- Compute

**Number of hours**
- 1

**Number of nodes**
- 1

**Memory (MB)**
- 600

___

Use JupyterLab instead of Jupyter Notebook?

JupyterLab is the next generation of Jupyter, and is completely compatible with existing Jupyter Notebooks.

Launch

* The HPC Tutorial Jupyter session data for this session can be accessed under the data root directory.
If you have an idea or want to work on any of the features or bug fixes in OOD, please feel free to reach out to us and we will get you started. We always want to include our community in the Open OnDemand efforts.
Dynamic Batch Connect Fields

Travis Ravert - OSC
Future plans

1. Completed Jobs App
2. Server side integration to address Single Sign On problems by moving communication to the server
3. Provide OnDemand usage metrics through XDMoD

Find more ways to help users optimize their jobs!
Funding and other acknowledgements:

- OnDemand is supported by the National Science Foundation – award numbers NSF#1534949 and NSF#1935725
- Open XDMoD is supported by the National Science Foundation – award numbers ACI 1025159 and ACI 1445806
- We gratefully acknowledge the partnership with Virginia Tech on our current joint NSF project
Thank you…

OSC, VT, and UB staff and students for helping with the tutorial today!
NSF Proposal Collaboration

Looking for smaller centers interested in working with us to get assistance on installing & managing these products

Contact Joseph White at UB
jpwhite4@buffalo.edu
How to reach us:

Center for Computational Research – [https://buffalo.edu/ccr](https://buffalo.edu/ccr)

Open XDMoD - [https://open.xdmod.org/](https://open.xdmod.org/)

ColdFront - [https://github.com/ubccr/coldfront](https://github.com/ubccr/coldfront)

Ohio Supercomputer Center - [https://www.osc.edu/](https://www.osc.edu/)

OnDemand - [https://openondemand.org/](https://openondemand.org/)

Virginia Tech – Advanced Research Computing - [https://arc.vt.edu/](https://arc.vt.edu/)
Other places you’ll find us at PEARC22:

Performance Optimization of the Open XDMoD Datawarehouse - **best full paper!** Tues, 7/12 10:30-11am

Open OnDemand User Group Meeting: Tues, 7/12 1:30-2:30pm

Open XDMoD BOF: Wed, 7/13 1:30-2:30pm

Enhancing User-centric Workflows and Democratizing Access to Novel Advanced Research Computing BoF, Thur 7/14 8-9am

ColdFront BOF: Thur, 7/14 9-10am

**Staff may be available after workshop concludes for specific questions. Please also join us on Slack!**
Thank you for attending!

Please fill out the post-tutorial survey

We value your opinions!