

Workshop Summary & Follow-up

Matt Heine

<https://indico.mit.edu/event/714/>



subMIT

Getting physics things done at MIT

<https://submit.mit.edu/>

Overview

News

People

Contact

About

Users Guide

JupyterHub

[Events, News](#)

Another Successful Workshop on Basic Computing Services in the Physics Department

by **Matthew Heine** – January 25, 2023

Earlier this month, the new year was ushered in by the annual Basic Computing Services in the Physics Department workshop. [...]

Summary

Contributions

<https://indico.mit.edu/event/647/>

1. Introduction/subMIT Project Overview

 Joshua Bendavid (Massachusetts Insti...

 1/6/23, 9:00 AM

2. Available software and environments

 Matthew Heine (Massachusetts Insti...

 1/6/23, 9:30 AM

3. Discussion on Linux Distribution Upgrade

 Joshua Bendavid (Massachusetts Insti...

 1/6/23, 10:00 AM

4. Tutorial/Hands-on help session

 Chad Freer (Massachusetts Insti...

 1/6/23, 10:30 AM

7. Lunch Break

 1/6/23, 12:00 PM

8. Introduction to ORCD

 James Cuff (Massachusetts Insti...

 1/6/23, 1:30 PM

5. subMIT as an Analysis Facility

 Mariarosaria D'Alfonso (MIT)

 1/6/23, 2:00 PM

6. Feedback and Experience with Julia etc

 Washington Taylor (Massachusetts Insti...

 1/6/23, 2:30 PM

9. Analysis of ABRACADABRA Data

 Kaliroe Pappas (MIT laboratory for n...

 1/6/23, 2:50 PM

10. Break

 1/6/23, 3:10 PM

12. Plans/feedback/experience from Lattice QCD

 Yin Lin (Massachusetts Insti...

 1/6/23, 3:25 PM

11. Plans/Feedback/Experience from Kavli

 Joshua Borrow (Massachusetts Insti...

 1/6/23, 3:55 PM

13. Simulating active matter with subMIT

 Sunghan Ro (MIT)

 1/6/23, 4:25 PM

14. Containers as a Service

 Denis Boyda (Massachusetts Insti...

 1/6/23, 4:45 PM

Overview of subMIT

- Unify MIT Physics Department computing resources
- Design
 - Login pool for Physics computing resources
 - Internal cluster: Prepare & test workflow here, then farm out heavy workload to larger connected or external resources
 - Interactivity & flexibility.
- Internal subMIT cluster
 - ~1,000 CPU cores, >50 GPU's (more being integrated)
 - local storage: 1TB/user, 10's of TB group storage
- Connected resources
 - LatticeQCD Cluster (GPUs), Open Science Grid, CMS Tier 2, CMS Tier 3
- User groups (monthly meetings)
- Help desk with physics computing staff submit-help@mit.edu
- Can integrate your existing or new resources into subMIT (& assist selecting)

Software & Tools

- CentOS7. Native: Julia, Matlab, OpenMPI, Python, C++, Fortran, ...
- Jupyter Notebooks (machine learning)
- You can customize your software (install your own)

Tools to help you manage your software:

- Package & Environment Managers (Conda, Spack)
- Containers (Docker, Singularity)

Tools to help you parallelize your code

- dask, OpenMPI, OpenMP
- subMIT help desk

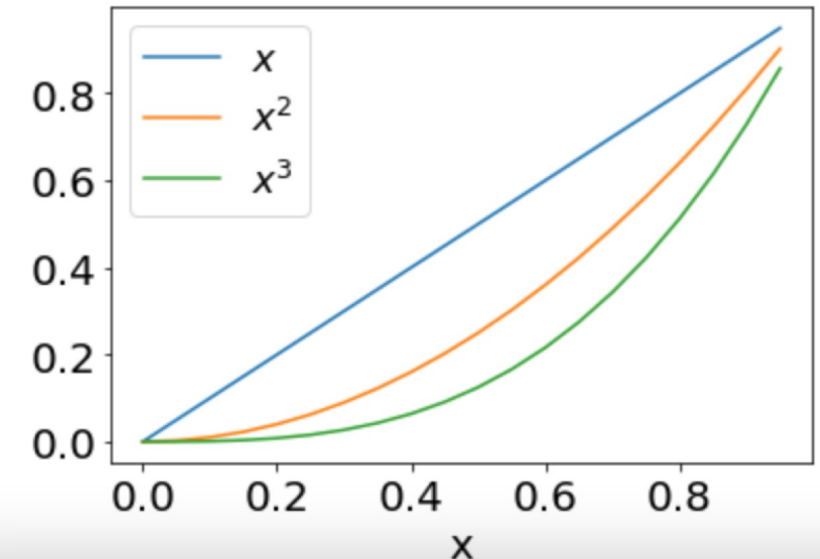
A Quick Example

```
In [1]: %%latex
Here we will plot  $x$ ,  $x^2$ , and  $x^3$ 
```

Here we will plot x , x^2 , and x^3

```
In [7]: import matplotlib.pyplot as plt
import numpy as np

x = np.arange(0, 1, 0.05);
plt.plot(x, x, label='$x$');
plt.plot(x, x**2, label='$x^2$');
plt.plot(x, x**3, label='$x^3$');
plt.xlabel('x');
plt.rc('font', size=18);
plt.legend();
```



Linux Upgrade

- CentOS 7 (Current): end of life (maintenance updates) June 2024
- Dec. 2020 decision of Red Hat to reorganize CentOS disrupted the otherwise natural CentOS 7 -> CentOS 8 transition
- Carefully weigh next distribution
 - Functionality
 - Support lifetime / update cycles
 - Ease of transition
- Work currently underway by subMIT Project Team:
 - Alma Linux 9 test installation underway
 - creating & documenting CentOS 7 containers
 - ease the transition
 - similar to existing environment
 - similar software installed
 - easily accessible

Introduction to ORCD – James Cuff (Office of Research Computing & Data)

- Statements from James Cuff:
- “Computing at MIT is complex, and highly decentralized.”
- “there are unlimited opportunities for collaboration and coordination of research computing services, systems and talent across the university.”
- “eager to engage with our community as ORCD finds their feet to be able to carry out their mission of providing quality and equitable access to research computing systems, services and skilled professional support staff.”
- leave existing systems in place (“do no harm”)
- investigations / discovery
- staffing, budget, org charts, capital planning

Contributions

1. Introduction/subMIT Project Overview

 Joshua Bendavid (Massachusetts Insti...

 1/6/23, 9:00 AM

2. Available software and environments

 Matthew Heine (Massachusetts Insti...

 1/6/23, 9:30 AM

3. Discussion on Linux Distribution Upgrade

 Joshua Bendavid (Massachusetts Insti...

 1/6/23, 10:00 AM

4. Tutorial/Hands-on help session

 Chad Freer (Massachusetts Insti...

 1/6/23, 10:30 AM

7. Lunch Break

 1/6/23, 12:00 PM

8. Introduction to ORCD

 James Cuff (Massachusetts Insti...

 1/6/23, 1:30 PM

5. subMIT as an Analysis Facility

 Mariarosaria D'Alfonso (MIT)

 1/6/23, 2:00 PM

6. Feedback and Experience with Julia etc

 Washington Taylor (Massachusetts Insti...

 1/6/23, 2:30 PM

9. Analysis of ABRACADABRA Data

 Kaliroe Pappas (MIT laboratory for n...

 1/6/23, 2:50 PM

10. Break

 1/6/23, 3:10 PM

12. Plans/feedback/experience from Lattice QCD

 Yin Lin (Massachusetts Insti...

 1/6/23, 3:25 PM

11. Plans/Feedback/Experience from Kavli

 Joshua Borrow (Massachusetts Insti...

 1/6/23, 3:55 PM

13. Simulating active matter with subMIT

 Sunghan Ro (MIT)

 1/6/23, 4:25 PM

14. Containers as a Service

 Denis Boyda (Massachusetts Insti...

 1/6/23, 4:45 PM

Feedback & Follow-Up

Workshop Feedback & Follow-up: More options for Jupyter Hub

- More flexibility / options for reservation of compute resources via Jupyter Hub
- Request compute nodes via Slurm from Jupyter Hub page
- Yin Lin: Plans/feedback/experience from Lattice QCD



jupyterhub

Select a job profile:

Slurm - Submit - 1 CPU, 500 MB



Start



jupyterhub

Select a job profile:

- ✓ Slurm - Submit - 1 CPU, 500 MB
- Slurm - Submit - 2 CPUs, 1000 MB
- Slurm - Submit-GPU - 1 GPU
- Slurm - Submit-GPU1080 - 1 GPU
- Local server - Submit01 - 1 CPU, 500 MB - /home/submit/{username}/
- Local server - Submit01 - 1 CPU, 500 MB - /work/submit/{username}/

Slurm - Submit - 1 CPU, 500 MB

Start

Quick introduction:

- **Spawn server menu:**
 - Slurm - Submit - 1 CPU, 500 MB: spawns a server on submit slurm partition.
 - Slurm - Submit - 2 CPUs, 1000MB: similar as above, with more resources allocated.
 - Slurm - SubmitGPU - 1 GPU: spawns a server on submit-gpu slurm partition, requesting 1 GPU.
 - Slurm - SubmitGPU1080 - 1 GPU: spawns a server on submit-gpu1080 slurm partition, requesting 1 GPU.
 - Local server - Submit01 - 1 CPU, 500 MB, /home/submit/{username}/: spawns on submit01, in your /home/submit/{username}/ directory.
 - Local server - Submit01 - 1 CPU, 500 MB, /work/submit/{username}/: spawns on submit01, in your /work/submit/{username}/ directory.
- **GPUs:** you can use GPU resources in your notebooks or Jupyterhub's terminal if you spawn a server on submit-gpu or submit-gpu1080, supported through Slurm.
- **Conda:** your conda environments should be automatically loaded as kernels by Jupyterhub, and can be used in notebooks. See User Guide for more info.
- **Singularity:** you can manually set up a kernel based on a singularity environment's python. See User Guide for more info.

For more information about Submit, conda, GPUs, Jupyterhub, etc., see:

User Guide

For any questions, comments, or feedback, please send an email to [submit-jupyter](#).

Workshop Feedback & Follow-up: Add Globus Endpoint

- tool for transferring & sharing data
- Non-profit but requires subscription
- Yin Lin: Plans/feedback/experience from Lattice QCD
- Joshua Borrow: Plans/Feedback/Experience from Kavli
- Action: currently looking into availability of subscription
- Backup plan: ensure sufficient documentation for xrootd

Workshop Feedback & Follow-up: mount Engaging NFS (/pool001)

- more easily prep, post-process/analyze, & transfer large engaging datasets on subMIT.
- Joshua Borrow: Plans/Feedback/Experience from Kavli
- Action: working towards implementing this

Workshop Feedback & Follow-up: increased storage

- need for expanded storage and archival
- Kaliroe Pappas: Analysis of ABRACADABRA Data
- Joshua Borrow: Plans/Feedback/Experience from Kavli
- Action: investigating MIT-provided options for tape archival

Workshop Feedback & Follow-up: more CPU nodes

- LQCD cluster is a GPU cluster
- Yin Lin: Plans/feedback/experience from Lattice QCD
- Joshua Borrow: Plans/Feedback/Experience from Kavli
- Action: purchase & integration of several large core / high-memory nodes

Workshop Feedback & Follow-up: add container application servers

- set up application server (container server) on subMIT
- Yin Lin: Plans/feedback/experience from Lattice QCD
- Denis Boyda: Containers as a Service
- Action: Denis will try setting this up & alert subMIT project team of any needs

Workshop Feedback & Follow-up: need for user training & “outreach”

- Joshua Borrow: Plans/Feedback/Experience from Kavli

Workshop Feedback & Follow-up: need for user training & “outreach”

excerpt from announcement to Physics community

Thursday, February 2

Taking advantage of Physics computing resources at MIT

10:00-12noon in Marlar Lounge (37-252/37-272)

9:45-10:00 coffee and light refreshments

10:00-10:30 An introduction to the computing resources available at MIT Speaker: Josh Borrow

Computing resources at MIT can be difficult to understand due to their fragmented nature. [...]

10:30-10:35 Q&A

10:35-11:05 From Laptops to High Performance Computing: Low-Hanging Fruit in Parallelization Speaker: Matt Heine

This talk is especially geared towards researchers who currently run calculations on their laptop or desktop computers, and will highlight common Physics computing workflows that lend themselves naturally to parallelization with little to no change in software or code. [...].

11:05-11:15 Q&A

Following the talks, Josh and Matt will be available for one-on-one conversations to discuss which computing system may best suit your needs and/or help spot low-lying fruit for parallelization in your workflow to take advantage of available computing resources.

Workshop Feedback & Follow-up: need for user training & “outreach”

- Further action: Survey to be sent out to MKI
- Further action: increased documentation & examples

Feedback Overview

- Spawning Jupyter to compute nodes w/ options
- add Globus endpoint
- mount engaging /pool001
- increased storage including archival
- more nodes
- application (container) server
- increased “outreach” & education (particularly in MKI)