subMIT Overview/Status

Mariarosaria D'Alfonso

Basic Computing Services (subMIT) Users Meeting

Nov. 19, 2024

Introduction

subMIT system provides an interactive login pool + scale-out to batch resources

- ○ Home directories
- ○ Convenient software environment (Alma Linux 9 native, docker/singularity images, conda)
- ○ SSH or Jupyterhub access
- ○ Local batch system with O(1000) cores, >50 GPU's
- ○ Additional storage for software installation/development, large datasets
- ○ Convenient access to larger external resources (OSG, CMS Tier-2 and Tier-3, LQCD Cluster, EAPS)
- User support is a key feature of the system
  - ○ Beyond basic troubleshooting
    - ■ Help users make optimal use of the available resources
    - ■ Expert advice on designing/improving workflows
    - ■ Customize and evolve system configuration to accommodate user needs as appropriate

subMIT: A MIT Physics Department Analysis Facility
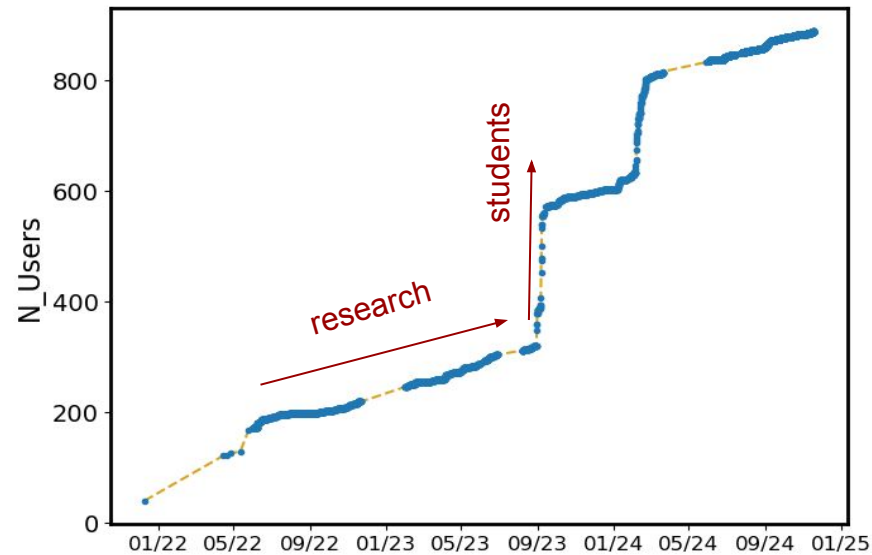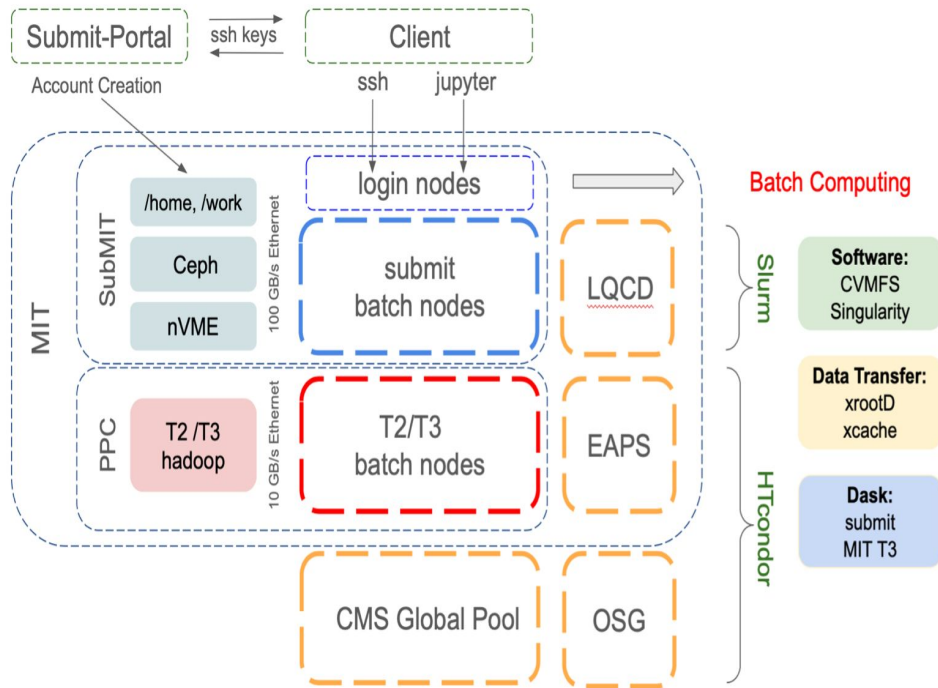
very flexible system, users utilize the sw/hw according to their requirements

Project Organization

- Formally the project is organized as ***Basic Computing Services*** in the Physics Department
  - **Project Team:** Implementation/Operations/Maintenance of the system
  - **Users Group:** Contact point between the user community and the project team, forum for user feedback, requests, information flow to and from users
  - **Steering Committee:** Faculty oversight, funding, etc
  - See https://submit.mit.edu/?page_id=6

Users Group In Practice

- Regular meetings (every few months)
  - Advertised and open to the broader community
  - Topical presentations from project team, Users Group representatives, or other users or community members
  - Forum for feedback and information flow between the user community and the project team
  - Regular timeslot: Tuesday 10:00-11:00 EST

- Users Group representatives
  - Identified representatives from research groups across the department
  - Attend the monthly meetings
  - Provide feedback from your groups/community
  - Distribute information/news from the project team

Communication Channels

- User support mailing list: submit-help@mit.edu
- Experimental large language model application under development for interactive user support and to augment support ticket handling
- Slack workspace: https://mit-submit.slack.com
  - "help-desk" channel
- Annual subMIT workshop
  - NEXT: January 2025 workshop

Call for subMIT User Group Representatives

Dear subMIT Users,

As a few of our User Group Representatives plan to move on to the next stages of their career beyond MIT, we thank them for their valuable service and are opening a call for new User Group Representatives.

User Group Representatives are critical, valuable members of subMIT with a special role in shaping the project. The Users Group serves to inform the Steering Committee about the performance and needs of the Basic Computing Service (subMIT), and we strive to have appropriate representation from across the Department. The Users Group is responsible for discussing and developing conceptual solutions for users' requests in communication with the Project Team. The members of the Users Group are also responsible to represent their local communities during meetings and report back to them. User Group Representatives meet approximately once a month.

If you would like to become a User Group Representative, please reply or email submit-help@mit.edu indicating your interest. As always, please reach out with any questions!

Thank you,
Josh, Matt, & Mariarosaria on behalf of subMIT

User Guide

https://submit.mit.edu/submit-users-guide/

User's Guide - subMIT

Contents:

- User's guide – subMIT login pool
- Getting started
- Access to subMIT
- Best practices
- Available software
- Batch computing
- User quota and storage at submit
- Monitoring at submit
- GPU resources
- Data backup
- Acknowledging subMIT

Next topic

User's guide – subMIT login pool

This Page

Show Source

Quick search

[   ] Go

Tutorials - subMIT

Tutorials:

- Tutorial 0: Introduction to the UNIX terminal
- Tutorial 1: Common software packages (python, Julia, MATLAB)
- Tutorial 2: Batch Job (HTCondor and Slurm)
- Tutorial 3: Containers (Podman and Singularity)
- Tutorial 4: Source Control (Git/Github) with Visual Studio Code (VSCode)
- Tutorial 5: Debugging Fortran code with Visual Studio Code (VSCode)
- Tutorial 6: Introduction to Pytorch Lightning
- Tutorial 7: Introduction to Snakemake

Today's agenda

Physics Basic Computing Services (subMIT) Users Meeting

📅 Tuesday Nov 19, 2024, 10:00 AM → 11:00 AM America/New_York

📍 Kolker Room (26-414) (MIT)

Description https://mit.zoom.us/j/96743699673?pwd=b3h2Q3c3cVQwYW12blhMUG5SWXZCZz09

10:00 AM → 10:20 AM **subMIT Overview, Status, & Updates**   ⏱20m

Speaker: Joshua Bendavid (Massachusetts Institute of Technology)

10:20 AM → 10:35 AM **What I'd like to use submit for**   ⏱15m

Benefits and drawbacks of trading my under-the-desk machine for using just subMIT, what the requires in terms of support, data volume, etc. to work

Speaker: Hans Günther (MIT)

📄 24_submit_UG.pdf

10:35 AM → 10:45 AM **Roundtable**   ⏱10m

Speakers: Amer Al-Hiyasat (MIT), Kaliroe Pappas (MIT laboratory for nuclear science), Molly Park (Massachusetts Institute of Technology), Prajwal Mohan Murthy (MIT LNS), Siddharth Mishra-Sharma (MIT), Yin Lin (Massachusetts Institute of Technology), Yitian Sun (Massachusetts Institute of Technology)

10:45 AM → 11:00 AM **Discussion**   ⏱15m

Backup

Storage breakdown

- Several different storage areas are available covering different use cases
  - /home/submit/<username>
    - Home directories (nfs server), redundant disk array with backups
    - 5GB quota
    - Use for software development and (small) critical data
  - /work/submit/<username>
    - Work directory (nfs server), no backups (but redundant disk array)
    - 50GB quota
    - Use for software installation (conda or docker/singularity images)
  - /ceph/submit/data/user/<username>
    - Large distributed disk system, no backups, but redundancy against disk failure ("erasure coding")
    - 1TB user quota, larger quotas available in dedicated group directories
    - Store large datasets here
  - /scratch/submit/<username>
    - Fast NVMe SSD array
    - Commissioned by several groups for high performance data analysis
  - /cvmfs/
    - Read-only distributed storage for distributing software, singularity images, etc
    - Several CERN-related repositories are available
    - Local repository /cvmfs/cvmfs.cmsaf.mit.edu where additional software or data can be added if needed
- Flexible tiered storage system, can accommodate a wide range of user needs
- Larger datasets encouraged to use shared group space, but quotas can be increased when needed
  - Group space in /ceph/submit/data/group/<groupname>