

HIGH PERFORMANCE RESEARCH COMPUTING

Scientific Machine Learning Resources at
Texas A&M's High Performance Research Computing Facility

Scientific Machine Learning Workshop
October 27, 2020



TEXAS A&M UNIVERSITY

Division of Research



Texas A&M University

High Performance Research Computing

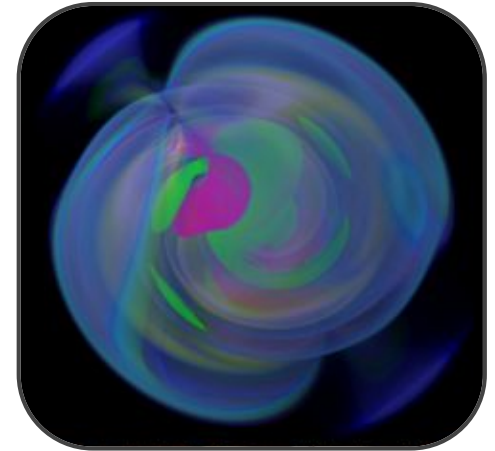
<https://hprc.tamu.edu>

High Performance Research Computing

An interdisciplinary research environment that advances computational and data-enabled sciences, engineering, and scholarship

Our Mission:

- Enable research and discoveries that advance science and technology.
- Enable computational and data-enabled research activities of students, faculty, and staff at Texas A&M University.
- Provide consulting, technical documentation, and training to support users of these resources.



HPRC Profile

- Operates three HPC clusters over 900 TFlops total computing capacity and over 12 PB storage for active users
 - ViDaL, Terra, and Ada
- Access is **open** and **free** of charge to ALL faculty, research staff, students and external collaborators
- Allocation by proposal is required. Support is available to help new users get started with the application process.
- Staff consists of 14 professionals and 9 student workers.
- Funding for staff, operations and maintenance from the Division of Research

HPRC Key Services

- Computing cycles, storage and networking for researchers
- Allocations and new user start-up assistance
- Individual consulting and help desk user services
- Advanced support for sponsored research projects
- Liaison program (HPRC Enablement) with Colleges of Science, Engineering, and Geosciences
- Access to state and national advanced computing resources
- Expertise in many science and engineering research domains
- Training classes and workshops for novices and experts
- Seminars and various outreach events

HPRC Systems - Terra & Ada

<https://hprc.tamu.edu/resources/>

Terra:

8,512-core hybrid system

304 28-core compute nodes equipped with the INTEL 14-core 2.4GHz Broadwell processor

48 GPU compute nodes with one **NVIDIA K80**

Interconnect fabric is Intel OmniPath Architecture (OPA)

Ada:

17,340-core hybrid system

845 20-core nodes equipped with the INTEL 10-core 2.5GHz IvyBridge processor.

15 nodes are 1TB and 2TB memory, 4-processor SMPs configured with the Intel 10-core 2.7GHz Westmere processors

30 GPU compute nodes with 2 **NVIDIA K20**

4 GPU compute nodes with 2 **NVIDIA V100**



Dr. Honggao Liu, Director
High Performance Research Computing



ViDaL: Secure & Compliant System

ViDaL: <https://vidal.tamu.edu>

- A 24-node **secure** and **compliant** computing environment that supports data intensive research using sensitive person-level data or proprietary licensed data to meet the myriad legal requirements of handling such data (e.g., HIPAA, Texas HB 300, NDA)
- 16 compute nodes with 192 GB Ram each and 4 large memory nodes with 1.5 TB Ram each, and 4 GPU nodes with 192 GB Ram and two NVIDIA V100 16GB GPUs each
- Both **Windows** and **Linux** environment.
- **2 PB** high performance DDN storage running IBM Spectrum Scale filesystem.
- Crowdstrike (antivirus/malware software) is used. Vidal systems are patched monthly. Splunk will be used for security information and event management
- Each Vidal server/VM is configured based on the CIS security benchmarks for Linux and Windows OSes and the TAMU Information Security Control Catalog.
- Housed in the secure West Campus Data Center



GRACE

Currently in Deployment

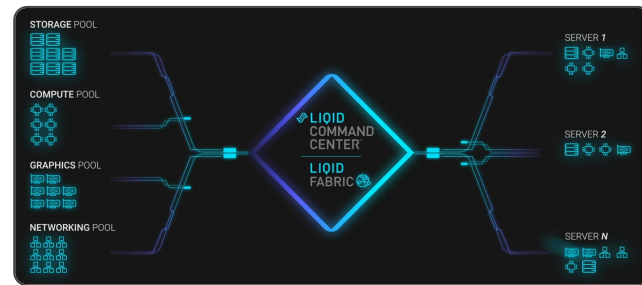


- The aggregate peak performance computing capacity is over 6 PFLOPS.
- 800 compute nodes equipped with 2 Intel 24-core 3.0 GHz processors and 384GB memory
- 100 GPU compute nodes have 2 **NVIDIA A100** 48GB GPUs with 2 Intel 24-core 3.0 GHz processors and 384GB memory
- 8 GPU compute nodes with 4 single precision **T4** 16GB
- 9 GPU compute nodes with 2 **RTX6000** 24GB GPUs, 2 Intel 24-core 3.0 GHz processors and 384GB memory
- 8 large memory compute nodes with 4 Intel 20-core 2.5GHz processors and 3.072 TB memory
- 5.12 PB of usable high-performance storage running Lustre parallel filesystem

FASTER

Fostering Accelerated Scientific Transformations, Education, and Research

To be deployed in Spring 2021



- Funded by NSF MRI grant (\$3.09M + \$1.32M TAMU match)
- Adopts the innovative LIQID **composable** software-hardware approach combined with cutting-edge technologies.
- Equipped with **NVIDIA A100**, and **T4/T4-Next** GPUs for AI/DL/ML workloads. Each node can access 16+ GPUs.
- Thirty percent of FASTER's computing resources will be allocated to researchers nationwide by the NSF XSEDE (Extreme Science and Engineering Discovery Environment) program.

FASTER project is supported by NSF award number 2019129



HPRC Training Short Courses

<https://hprc.tamu.edu/training>

Introduction to the Linux

Introduction to Using Ada & Terra

Introduction to NGS Data Analysis

RNA Sequencing

NGS Metagenomics

NGS RADSeq/GBS

NGS Assembly

Introduction to HPRC Galaxy

Introduction to R

Introduction to Perl

Introduction to Databases

Introduction to Python

Introduction to Scientific Python

Slurm Job Scheduling

Data Management Practices

Using the HPRC Portal: ParaView

Introduction to CUDA

Intermediate CUDA Programming

Introduction to the MATLAB Parallel Toolbox

Introduction to Deep Learning with PyTorch

Introduction to Deep Learning with TensorFlow

Introduction to Code Parallelization using OpenMP

Introduction to Code Parallelization using MPI

Introduction to Fortran

Introduction to MATLAB Programming

Python for MATLAB Users

Bring Your Own Code Workshops

and more...



Upcoming HPRC training

Technology Lab: Using AI Frameworks in Jupyter Notebook

Instructor: Jian Tao

Time: Friday, October 30, 1:30PM-4:00PM

Registration: <https://hprc.tamu.edu/training>

This short course contains a set of four sessions to help a new user start with his/her machine learning projects on HPRC systems

- Lab 1 - Jupyter Notebook - Set up a Python virtual environment and run JupyterLab on the HPRC Portal.
- Lab 2 - Data Exploration - Go through simple examples with two popular Python modules: Pandas and Matplotlib for simple data exploration.
- Lab 3 - Machine Learning - Learn to use scikit-learn for linear regression and classification applications.
- Lab 4 - Deep Learning - Learn to use Keras to create and train a simple image classification model with deep neural networks (DNN)





High Performance
Research Computing
DIVISION OF RESEARCH

YouTube training videos



Texas A&M HPRC

157 subscribers

SUBSCRIBED



HOME

VIDEOS

PLAYLISTS

CHANNELS

DISCUSSION

ABOUT



Uploads ▶ PLAY ALL



HPRC Intro #11: Submitting a Job Using LSF

18 views · 6 days ago

CC



HPRC Primers: Introduction to Linux

19 views · 3 weeks ago



HPRC Intro #8: Submitting a Job Using SLURM

45 views · 3 weeks ago

CC



HPRC Intro: #6 The Modules System

71 views · 3 weeks ago

CC

FEATURED CHANNELS



Texas A&M University

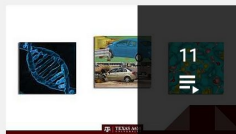
SUBSCRIBE



Research @ Texas A&M

SUBSCRIBE

All Playlists



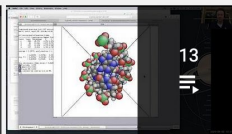
Getting Started with HPRC



HPRC Primers



OpenOnDemand Portal



Short Courses



Texas A&M University

High Performance Research Computing

<https://hprc.tamu.edu>

Documentation

<https://hprc.tamu.edu/wiki>

High Performance Research Computing

A Resource for Research and Discovery



TEXAS A&M
UNIVERSITY

Welcome to the TAMU HPRC Wiki

- [Ada Guide](#)
- [Software](#)

- [Terra Guide](#)
- [Usage Policies](#)

- [Contact Us](#)

Announcements

- **New GPU nodes in the Ada cluster:** Four new GPU nodes are now available in the Ada Cluster. Each GPU node has two Intel Skylake Xeon Gold 5118 20-core processors, 192 GB of memory and two NVIDIA 32GB V100 GPUs. To use these new GPU nodes, please submit jobs to the **v100** queue on Ada by including the following job directive in your job scripts:

```
#BSUB -q v100
```

Getting an Account

- **Understanding HPRC:** New to High Performance Computing (HPC)? This [HPC Introduction Page](#) explains the "why" and "how" of high performance computing. Also see the [Policies Page](#) to better understand the rules and etiquette of cluster usage.
- **Accessing the clusters:** All computer systems managed by the HPRC are available for use to TAMU faculty, staff, and students who require large-scale computing capabilities. The HPRC hosts the [Ada](#) and [Terra](#) clusters at TAMU. To apply for or renew an HPRC account, please visit the [Account Applications](#) page. For information on how to obtain an allocation to run jobs on one of our clusters, please visit the [Allocations Policy](#) page. *All accounts expire and must be renewed in September of each year.*

Using the Clusters

- **QuickStart Guides:** For just the "need-to-know" information on getting started with our clusters, visit our QuickStart pages. Topics discussed include cluster access, file management, the batch system, setting up a software environment using modules, creating your own job files, and project account management. [Ada QuickStart Guide](#), [Terra Quickstart Guide](#)
- **Batch Jobs:** As a shared resource between many users, each cluster must employ a batch system to schedule a time for each user's job to run. Without such a system, one user could use a disproportionate amount of resources, and cause other users' work to stall. Ada's batch system is called LSF, and Terra's batch system is called SLURM. While similar in function, they differ in their finer details, such as job file syntax. Information relevant to each system can be found below.

Ada / LSF Batch Pages

[Complete Ada Batch Page](#)

[Job Submission \(bsub\)](#)

[Ada Queue Structure](#)

Terra / SLURM Batch Pages

[Complete Terra Batch Page](#)

[Job Submission \(sbatch\)](#)

[Terra Queue Structure](#)

Available Software Modules

SOFTWARE MODULES ON THE TERRA CLUSTER

<https://hprc.tamu.edu/software/terra>
<https://hprc.tamu.edu/software/ada>

Last Updated: Mon Oct 12 00:00:02 CDT

The available software for the Terra cluster is listed in the table. Click on any software package name to get more information such as the available versions, additional documentation if available, etc.

Show entries

Name	Description
GPflow	GPflow is a package for building Gaussian process models in python
Horovod	Horovod is a distributed training framework for TensorFlow. URL: https://horovod.ai/
Keras	Keras is a minimalist, highly modular neural networks library, written on top of either TensorFlow or Theano. URL: https://keras.io/
segmentation-models	Python library with Neural Networks for Image Segmentation based on TensorFlow. URL: https://github.com/qubvel/segmentation_models
TensorFlow	An open-source software library for Machine Intelligence URL: https://www.tensorflow.org/

Showing 1 to 5 of 5 entries (filtered from 1,636 total entries)

```
[mouse@terra1 ~]$ ml spider TensorFlow
```

TensorFlow:

Description:

An open-source software library for Machine Intelligence

Versions:

```
TensorFlow/1.4.0-intel-2017A-Python-3.5.2
TensorFlow/1.5.0-foss-2017b-Python-3.6.3-03
TensorFlow/1.5.0-foss-2017b-Python-3.6.3
TensorFlow/1.5.0-fosscuda-2017b-Python-3.6.3
TensorFlow/1.5.0-goofc-2017b-Python-3.6.3
TensorFlow/1.6.0-foss-2018a-Python-3.6.4-CUDA-9.1.85
TensorFlow/1.6.0-foss-2018a-Python-3.6.4
TensorFlow/1.7.0-foss-2018a-Python-3.6.4-CUDA-9.1.85
TensorFlow/1.7.0-foss-2018a-Python-3.6.4
TensorFlow/1.8.0-foss-2017b-Python-2.7.14
TensorFlow/1.8.0-foss-2017b-Python-3.6.3
TensorFlow/1.8.0-foss-2018a-Python-3.6.4
TensorFlow/1.8.0-fosscuda-2017b-Python-2.7.14
TensorFlow/1.8.0-fosscuda-2017b-Python-3.6.3
TensorFlow/1.10.0-fosscuda-2018b-Python-2.7.15
TensorFlow/1.10.1-foss-2018b-Python-3.6.6
TensorFlow/1.10.1-fosscuda-2018b-Python-2.7.15
```

<https://hprc.tamu.edu/wiki/SW:Modules>



Advanced Support Program

HPRC offers collaborations in research projects with a large computational component. Under the **ASP**, one or more HPRC analysts will contribute expertise and experience in several areas of high performance computing in a sustained and focused way.

- Porting applications to our clusters
- Optimizing and analyzing serial code performance
- Developing parallel code from serial versions and analyzing performance
- Optimizing serial and parallel I/O code performance
- Assisting in the optimal use of mathematical libraries
- Assisting with code development and design
- Assisting with the improvement of workflow automation in scientific processes

If you are interested in a collaboration through our **ASP** program, please send us an e-mail at, help@hprc.tamu.edu.

ASP is supported in part by NSF award #1925764, CC* Team: SWEETER -- SouthWest Expertise in Expanding, Training, Education and Research.



Home / My Interactive Sessions

Interactive Apps

- BIO
- IGV
- Mauve
- Structure
- GUI
- ANSYS Workbench
- Abaqus/CAE
- LS-PREPOST
- MATLAB
- ParaView
- VNC
- Servers
- Jupyter Notebook
- JupyterLab
- RStudio Server with R 3.4.3 (Singularity)
- RStudio Server with R 3.6.1 (Singularity)
- Spark-Jupyter Notebook

You have no active sessions.

Open OnDemand (OOD) is open source software developed at the Ohio Supercomputing Center

Enables HPC centers to deploy advanced web and graphical interface for their users

HPRC Portal
YouTube tutorials

HPRC Portal



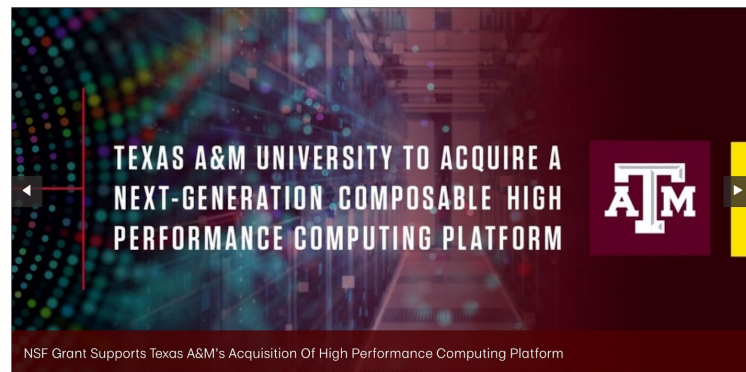
Quick Links

- New User Information
- Accounts
- Apply for Accounts
- Manage Accounts
- User Consulting
- Training
- Documentation
- Software
- FAQ

User Guides

- Ada
- Terra
- Portal
- Galaxy

Cluster Status



<https://portal-ada.hprc.tamu.edu>
<https://portal-terra.hprc.tamu.edu>



HPRC Open OnDemand Portal

<https://portal.hprc.tamu.edu>

MOLDEN

```
mouse@gpu256-3008 [1021]: molden cnt3_cnt2_s
Allocating memory for orbitals !
First standard orientation encountered usec
for density calculations

Use keyword gfinput in gaussian inputfile
error while reading Basis set!

WARNING: composite optimisation/frequency ;
The frequency job should be a separate file
With First line:

Entering Gaussian System
mouse@gpu256-3008 [1022]: molden cnt_cnt2.gas
Allocating memory for orbitals !
First standard orientation encountered usec
for density calculations

Use keyword gfinput in gaussian inputfile
error while reading Basis set!
```

TAMU HPRC OnDemand (Terra) Files Jobs Clusters Interactive Apps My Interactive Sessions

OnDemand provides an interactive environment for all of your research needs.

Message of the Day

IMPORTANT POLICY INFORMATION

- Unauthorized use of HPRC resources
- Use of HPRC resources in violation of laws and regulations
- Sharing HPRC account and password
- Authorized users must adhere to the University of Texas System Information Security Policy

!! WARNING: There are NO active backups !!

Servers

- Jupyter Notebook
- JupyterLab
- RStudio R 3.6.1
- Spark-Jupyter Notebook

_TESTING

- IGV
- TESTING
- TESTING_TV

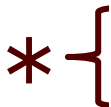
powered by
OPEN OnDemand
<https://portal-terra.hprc.tamu.edu/portal/sys/dashboard#>





<https://hprc.tamu.edu>

Quick Links



- New User Information
- Accounts
 - Apply for Accounts
 - Manage Accounts
- User Consulting
- Training
- Documentation
- Software
- FAQ

User Guides

- Ada
- Terra
- Curie
- Portal
- Galaxy

Cluster Status

Terra

Nodes	293/316 (93%)
Cores	7133/9352 (76%)

TEXAS A&M UNIVERSITY TO ACQUIRE A NEXT-GENERATION COMPOSABLE HIGH PERFORMANCE COMPUTING PLATFORM

NSF Grant Supports Texas A&M's Acquisition Of High Performance Computing Platform

News

AUG 7 [NSF Grant Supports Texas A&M's Acquisition Of High Performance](#)

Events

AUG 7 [Short Course: NGS Assembly](#)



HPRC Account: PI Eligibility

For the purpose of HPRC allocations, only **active faculty** members and **permanent research staff** (subject to HPRC-RAC Chair review and approval) of Texas A&M System Members headquartered in Brazos County can serve as a PI.

Adjunct and Visiting professors do not qualify themselves, but can use HPRC resources as part of a sponsoring PI's group.

Note that:

- A PI can have more than one allocation.
- A user can work on more than one project and with more than one PI

<https://hprc.tamu.edu/policies/allocations.html>



HPRC Account Allocations

<https://hprc.tamu.edu/policies/allocations.html>

Allocation Type	Who can apply?	Minimum SUs per Allocation per Machine	Maximum SUs per Allocation per Machine	Maximum Total SUs per Machine	Maximum Number of Allocations per Machine	Allowed to spend more than allocation?	Reviewed and approved by
Basic	Faculty, Post-Docs*, Research Associates, Research Scientists, Qualified Staff, Students*, Visiting Scholars/Students*	5,000	5,000	5,000	1	No	HPRC Staff
Startup	Faculty, Research Associates, Research Scientists, Qualified Staff	5,000	200,000	400,000	2	No	HPRC Director
Research (Ada)	Faculty, Research Scientists, Qualified Staff	300,000	8,000,000	8,000,000	Determined by HPRC-RAC	No	<u>HPRC-RAC</u>
Research (Terra)	Faculty, Research Scientists, Qualified Staff	300,000	5,000,000	5,000,000	Determined by HPRC-RAC	No	<u>HPRC-RAC</u>

Note: '*' - requires a PI





**HIGH PERFORMANCE
RESEARCH COMPUTING**
TEXAS A&M UNIVERSITY

More information is available at:

<https://hprc.tamu.edu>

For support and questions,
contact the HPRC Helpdesk

help@hprc.tamu.edu

(979) 845-0219

