

Accelerated Genomics with NVIDIA Parabricks

NSF ACES

Wes Brashear
20 July 2024



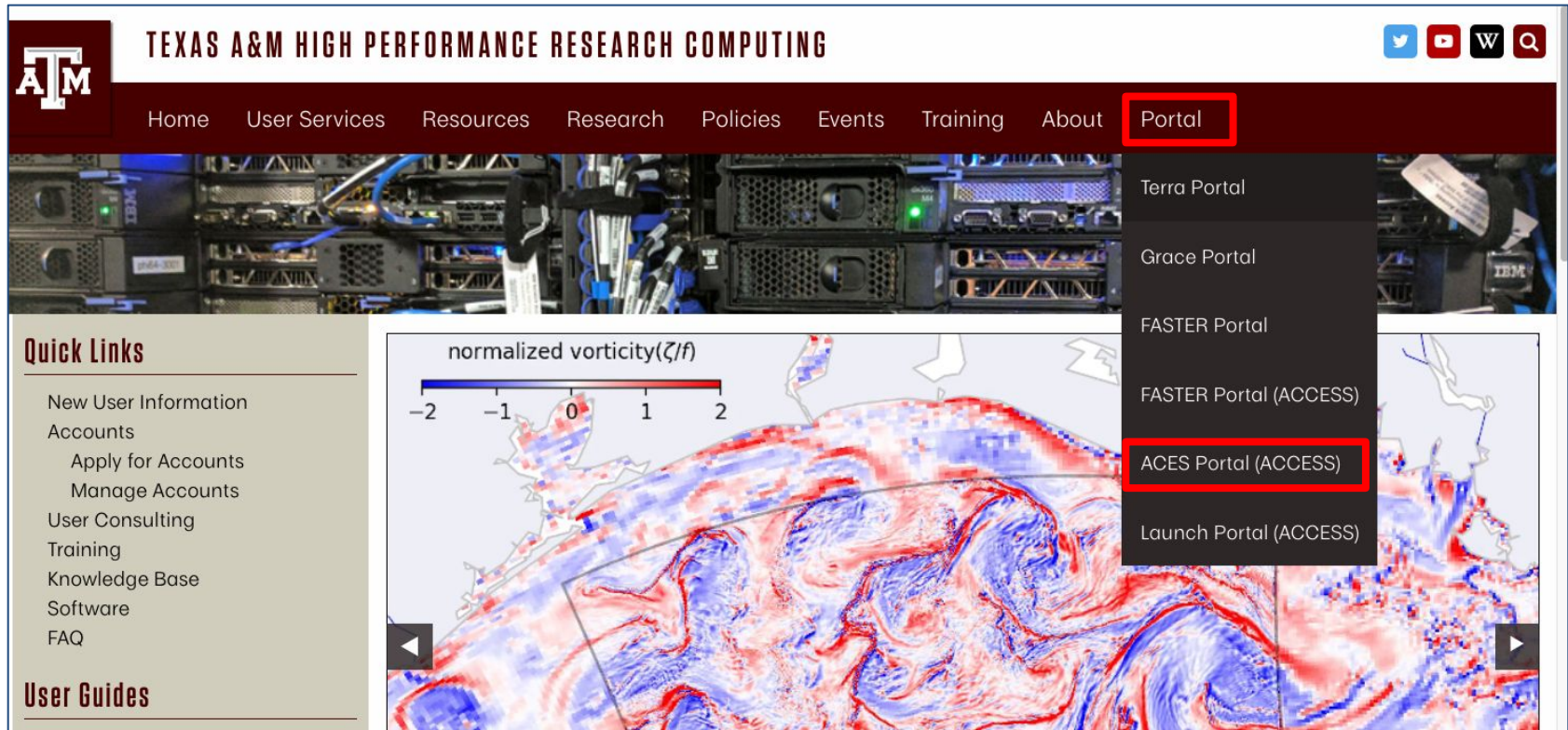
High Performance
Research Computing
DIVISION OF RESEARCH



UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN



Accessing the HPRC ACES Portal



The screenshot displays the Texas A&M High Performance Research Computing (HPRC) website. The header features the Texas A&M logo and the text "TEXAS A&M HIGH PERFORMANCE RESEARCH COMPUTING". A navigation bar includes links for Home, User Services, Resources, Research, Policies, Events, Training, About, and Portal. The "Portal" link is highlighted with a red box. A dropdown menu is open, showing options: Terra Portal, Grace Portal, FASTER Portal, FASTER Portal (ACCESS), ACES Portal (ACCESS) (highlighted with a red box), and Launch Portal (ACCESS). On the left, a "Quick Links" section lists various user services. The main content area shows a visualization of normalized vorticity (ζ/f) over a map of the Gulf of Mexico, with a color scale from -2 to 2.

TEXAS A&M HIGH PERFORMANCE RESEARCH COMPUTING

Home User Services Resources Research Policies Events Training About **Portal**

Quick Links

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

User Guides

normalized vorticity(ζ/f)

-2 -1 0 1 2

Terra Portal

Grace Portal

FASTER Portal

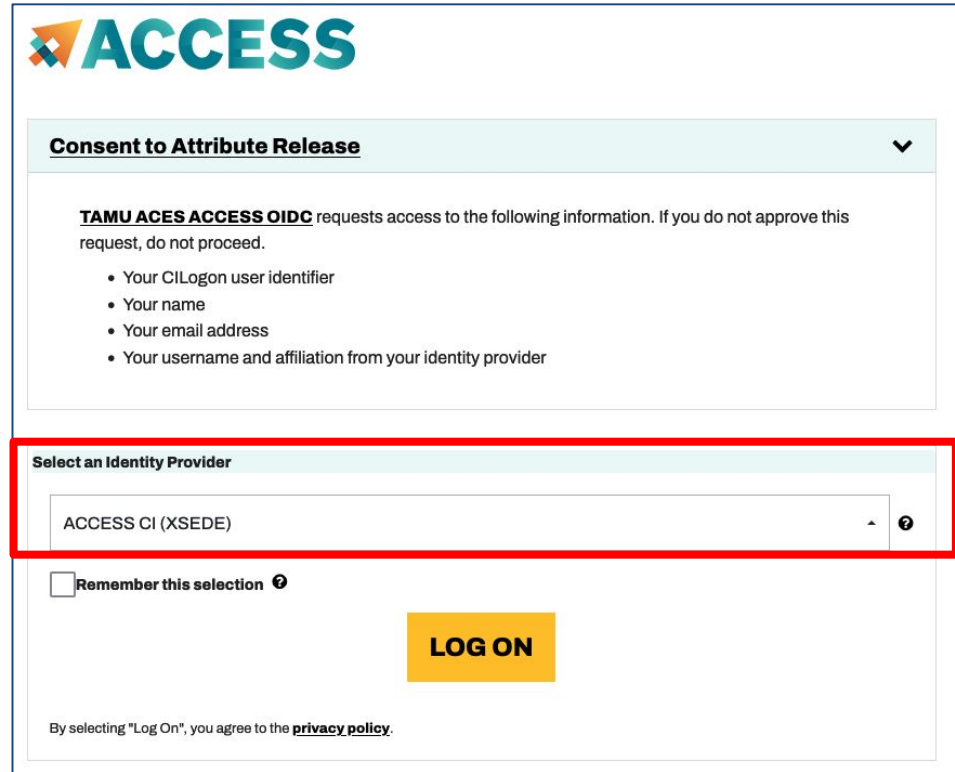
FASTER Portal (ACCESS)

ACES Portal (ACCESS)

Launch Portal (ACCESS)

Accessing the HPRC ACES Portal

Select the Identity Provider appropriate for your account



ACCESS

Consent to Attribute Release ▼

TAMU ACES ACCESS OIDC requests access to the following information. If you do not approve this request, do not proceed.

- Your CILogon user identifier
- Your name
- Your email address
- Your username and affiliation from your identity provider

Select an Identity Provider

ACCESS CI (XSEDE) ▲ ⓘ

☐ Remember this selection ⓘ

LOG ON

By selecting "Log On", you agree to the [privacy policy](#).



Accessing the ACES Shell

ACES OnDemand Portal Files ▾ Jobs ▾ Clusters ▾ Interactive Apps ▾ Affinity Groups ▾ Dashboard ▾ My Interactive Sessions

>_aces Shell Access



ACES
ACCELERATING COMPUTING
FOR EMERGING SCIENCES



```

*****
*                               *
*      === IMPORTANT POLICY INFORMATION ===      *
* - Unauthorized use of HPRC resources is prohibited and subject to      *
*   criminal prosecution.                                                 *
* - Use of HPRC resources in violation of United States export control   *
*   laws and regulations is prohibited. Current HPRC staff members are  *
*   US citizens and legal residents.                                       *
* - Sharing HPRC account and password information is in violation of     *
*   Texas State Law. Any shared accounts will be DISABLED.              *
* - Authorized users must also adhere to ALL policies at:                *
*   https://hprc.tamu.edu/policies/ *
*****

**** ACES Status, June 28 ****

Most H100s are available again. We plan to return the remaining 4 H100s
to service early next week after a hardware component is replaced.

!! WARNING: THERE ARE ONLY NIGHTLY BACKUPS OF USER HOME DIRECTORIES. !!

Please restrict usage to 8_CORES across ALL login nodes.
Users found in violation of this policy will be SUSPENDED.

To see these messages again, run the motd command.
Your current disk quotas are:
Disk                               Disk Usage    Limit    File Usage    Limit
/home/u.eu100000                   2.7G       10.0G      8593         10000
/scratch/user/u.eu100000           1.6T       5.0T      80960        250000
/scratch/group/p.sta220004.000     4K         1.0T         1         500000
/scratch/group/p.tra230003.000     4K         1.0T         1         500000
/scratch/group/p.tra220029.000     4K         1.0T         1         500000
Type 'showquota' to view these quotas again.
[u.eu100000@aces.login1 ~]$

```



Hands-on Demonstration: Getting Started

1. Change your working directory to your SCRATCH space

```
1. [u.eu1000000@aces-login1] cd $SCRATCH
```

2. Copy the example data

```
2. [u.eu1000000@aces-login1] cp -r \  
/scratch/training/parabricks .
```

3. Examine the contents of the new directory

```
3. [u.eu1000000@aces-login1] cd parabricks && ls  
bwa-cpu.sh  HSA_NA12878_5x_1.fastq.gz  
HSA_NA12878_5x_2.fastq.gz  ref
```

4. Submit the bwa-cpu job

```
4. [u.eu1000000@aces-login1] sbatch bwa-cpu.sh
```



```
#!/bin/bash
#SBATCH --export=NONE           # do not export current env to the job
#SBATCH --job-name=bwa-cpu      # job name
#SBATCH --time=02:00:00         # max job run time dd-hh:mm:ss
#SBATCH --ntasks-per-node=1     # tasks (commands) per compute node
#SBATCH --cpus-per-task=48      # CPUs (threads) per command
#SBATCH --mem=240G              # total memory per node
#SBATCH --output=%x.%j.out      # save stdout to file
#SBATCH --error=%x.%j.err       # save stderr to file

# Load the necessary modules
module load GCCcore/12.2.0 GCC/12.2.0 BWA/0.7.17 SAMtools/1.17

# Mapping command
bwa mem -t $SLURM_CPUS_PER_TASK ref/Homo_sapiens_assembly38.fasta \
    HSA_NA12878_5x_1.fastq.gz HSA_NA12878_5x_2.fastq.gz | \
    samtools sort -o HSA_bwa.bam -m 4G -@ $SLURM_CPUS_PER_TASK
```



NVIDIA Clara Parabricks Pipelines



Applications

Alignment

Variant Calling

Variant Processing

Preprocess

Joint Genotyping

Quality Checking

CUDA-X

EGX



DGX



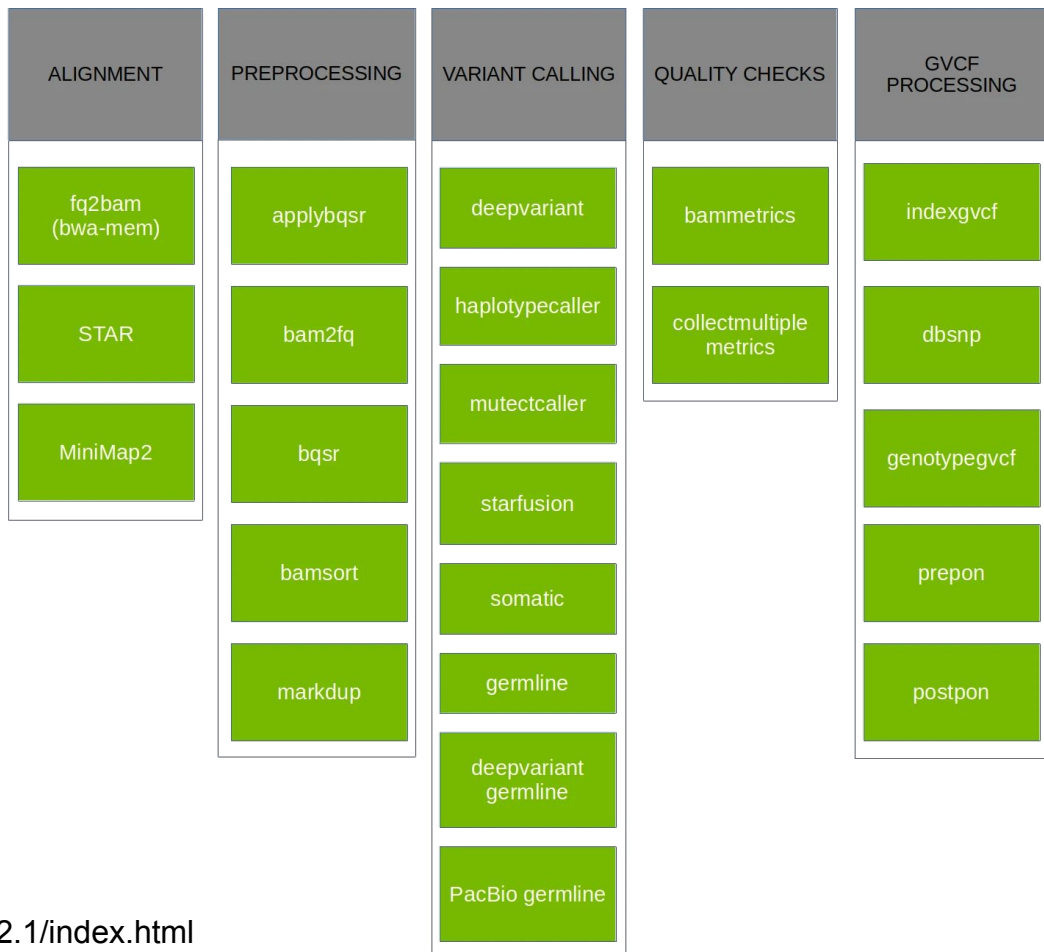
CLOUD



<https://developer.nvidia.com/blog/taking-gpu-based-ngs-data-analysis-to-another-level-with-clara-parabricks-pipelines-3-0/>



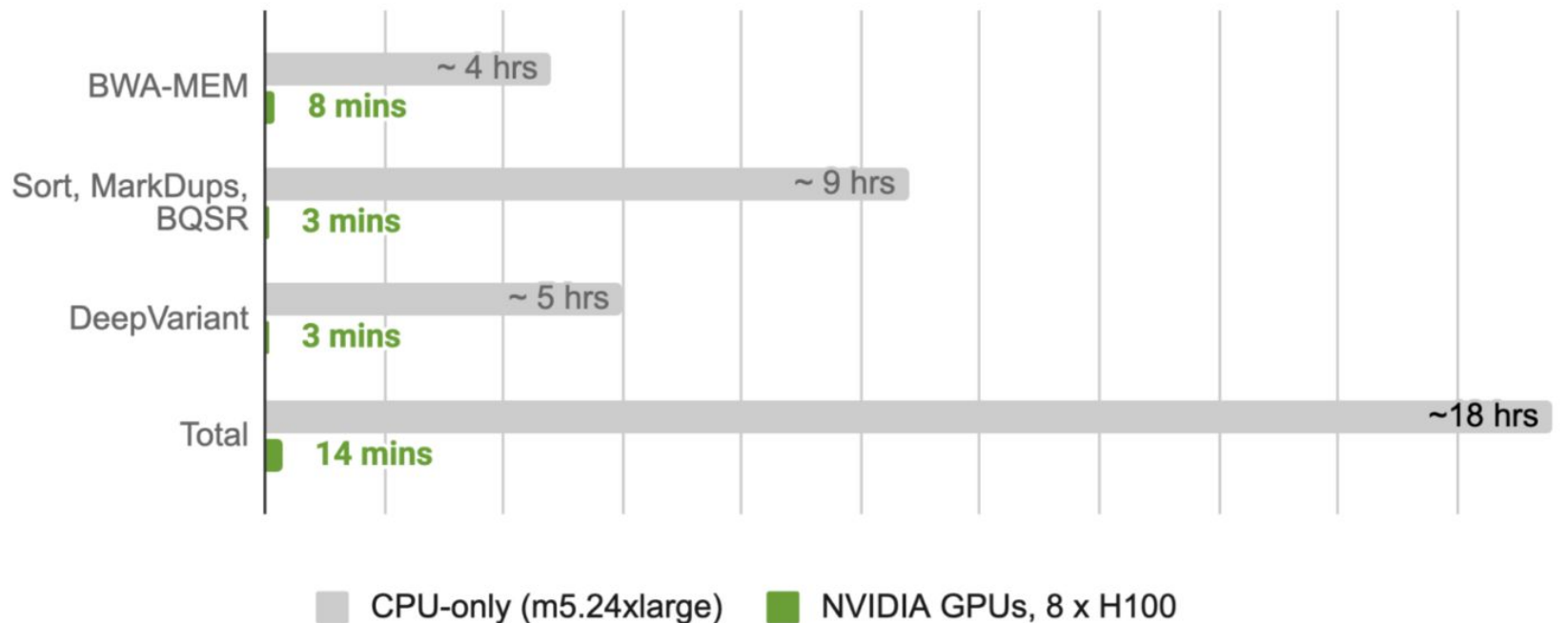
- GPU-enabled versions of commonly used programs
- Full pipelines in a single command
- Compatible with several pipeline frameworks (e.g. NextFlow)



<https://docs.nvidia.com/clara/parabricks/4.2.1/index.html>

PARABRICKS GERMLINE WORKFLOW ON H100

CPU-only vs 8xH100 GPUs



<https://developer.nvidia.com/blog/accelerate-genomic-analysis-for-any-sequencer-with-parabricks-v4-2/>



Running Parabricks on ACES

1. Start an srun session in ACES shell

```
[u.eu100000@aces-login1 parabricks] srun --nodes=1 --time=01:00:00 --mem=240G \  
--partition=GPU --gres=gpu:2 --cpus-per-task=48 --reservation=workshop_h100 --pty bash
```

2. Clear previously loaded modules and load Parabricks

```
[u.eu100000@aces-login1 parabricks] module purge && module load Parabricks/4.1.1
```

3. Run the analogous command we previously ran with BWA in Parabricks

```
[u.eu100000@aces-login1 parabricks] pbrun fq2bam --ref ref/Homo_sapiens_assembly38.fasta  
--in-fq HSA_NA12878_5x_1.fastq.gz HSA_NA12878_5x_2.fastq.gz --out-bam HSA-parabricks.bam  
--num-gpus 2
```



Texas A&M at PEARC24

Talk/Event	Date/Time	Room
Tutorial: Hands-on exercises on the Intel Data Center GPU Max 1100 (PVC-GPU) for AI/ML and Molecular	Mon, July 22, 2024 9:00 AM-12:30 PM ET	Room 553B
Seventh Workshop on Strategies for Enhancing HPC Education and Training (SEHET24)	Mon, July 22, 2024 9:00 AM-12:30 PM ET	Room 557
Workshop: Providing cutting-edge computing testbeds to the science and engineering community	Mon, July 22, 2024 1:30 PM-5:00 PM ET	Room 554A
Workshop: Engaging Secondary Students in Computing: K12 Outreach	Mon, July 22, 2024 1:30 PM-5:00 PM ET	Room 553A
Cultivating Cyberinfrastructure Careers through Student Engagement at Texas A&M University High Performance Research Computing	Tue, July 23, 2024 11:00 AM-11:25 AM ET	Junior Ballroom
Insight Gained from Migrating a Machine Learning Model to Intelligence Processing Units	Tue, July 23, 2024 11:00 AM-11:25 AM ET	Room 551 A&B
BOF 4: What's in it for me? How can we truly democratize the research computing and data community?	Tue, July 23, 2024 1:30 PM-2:30 PM ET	Room 551 A&B



Texas A&M at PEARC24

Talk/Event	Date/Time	Room
BRICCs: Building Pathways to Research Cyberinfrastructure at Under Resourced Institutions	Tue, July 23, 2024 3:25 PM-3:50 PM ET	Junior Ballroom
Memory Bandwidth Performance across Accelerators	Tue, July 23, 2024 3:25 PM-3:50 PM ET	Ballroom B
Container Adoption in Campus High Performance Computing	Wed, July 24, 2024 11:00 AM-11:25 AM ET	Ballroom B
Engaging Secondary Students in Computing and Cybersecurity	Wed, July 24, 2024 3:15 PM-3:30 PM ET	Room 557
Exploring the Viability of Composable Architectures to Overcome Memory Limitations in High Performance Computing Workflows	Wed, July 24, 2024 3:45 PM-4:00 PM ET	Room 553 A&B
Performance of Molecular Dynamics Acceleration Strategies on Composable Cyberinfrastructure	Wed, July 24, 2024 4:15 PM-4:30 PM ET	Room 551 A&B
BOF 17: Fantastic ACCESS Cyberinfrastructure Resources and Where to Find Them	Wed, July 24, 2024 4:45 PM-5:45 PM ET	Room 553 A&B
BOF 18: Recipes to build successful cross-institutional collaborative computing	Wed, July 24, 2024 4:45 PM-5:45 PM ET	Junior Ballroom





High Performance
Research Computing
DIVISION OF RESEARCH

Thank you

- We gratefully acknowledge support from National Science Foundation awards #2112356 (ACES), #2019129 (FASTER) and #19257614(SWEETER)
- Please visit our talks and BoFs at PEARC23
- Helpdesk: help@hprc.tamu.edu

