Intel-Optimized AlphaFold on ACES

July 15, 2023
ACES Workshop
Michael Dickens

High Performance
Research Computing
DIVISION OF RESEARCH

ACES
ACCELERATING COMPUTING
FOR EMERGING SCIENCES

Intel-Optimized AlphaFold (iaf2)

- Using the ACES Open OnDemand (OOD) Portal
  - use GCATemplates for a template job script
  - launch iaf2 job script
- Intel AlphaFold background
- Job resource monitoring (cores, memory)
- Viewing results

Intel AlphaFold using the Command Line

- https://portal-aces.hprc.tamu.edu

Finding job template scripts using GCATemplates

```
mkdir $SCRATCH/workshop
```

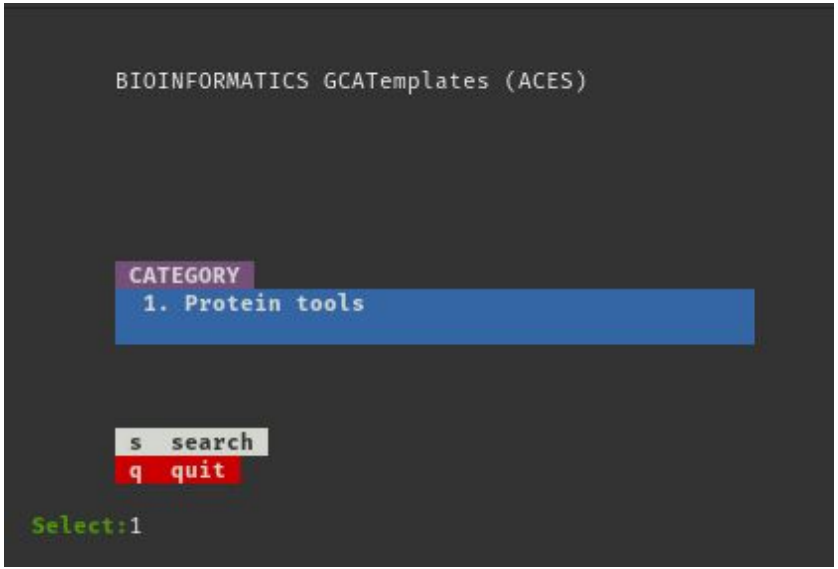
```
cd $SCRATCH/workshop
```

```
gcatemplates
```

We will copy a template file and submit it

- Enter 1 then continue through the menus to find the template that contains iaf2
  - or use the search to find iaf2
- Final step will save a template job script file to your current working directory
- Submit job script to Slurm scheduler

Genomic Computational Analysis Templates




```
sbatch run_alphafold_iaf2-2.9.0_monomer_aces.sh
```

Intel-Optimized AlphaFold Background

Intel-Optimized AlphaFold Pipeline

- Designed and Optimized for Intel Xeon Processors
- Does not use GPUs or FPGAs
- Still under development
  - only model 1 is available
  - does not rank models and select best scoring model
- Run on ACES compute nodes
  - Intel(R) Xeon(R) Platinum 8468 (2.10 GHz)
  - a.k.a. Intel Sapphire Rapids
  - must reserve entire compute node
    - 96 cores; 480GB memory

Intel-Optimized AlphaFold Pipeline

1. Preprocessing
    a. retrieve structure templates from similar known protein sequences
2. Model Inference
    a. Predict and construct 3D representation of structure based on preprocessing step
3. Post-Processing
    a. Refine the 3D structure

Intel-Optimized AlphaFold Pipeline

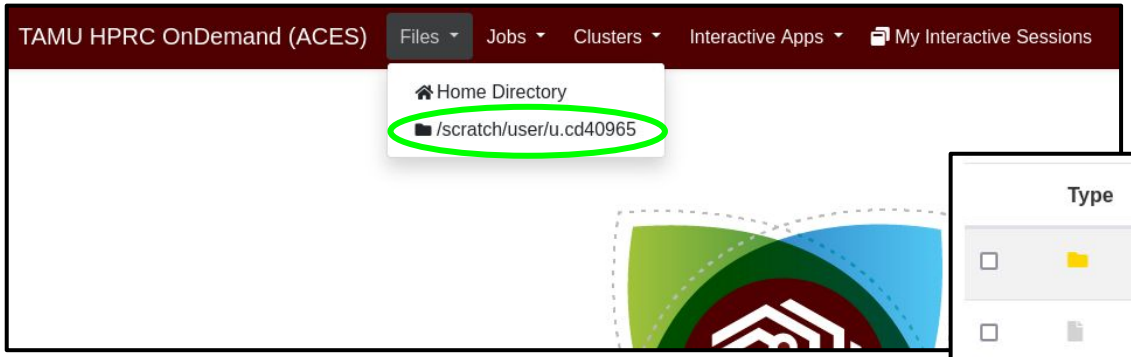
1. Monomer
    a. Single Instance preprocessing
    b. Single Instance model inference
2. Multimer
    a. Multi-instance preprocessing
    b. Multi-instance model inference

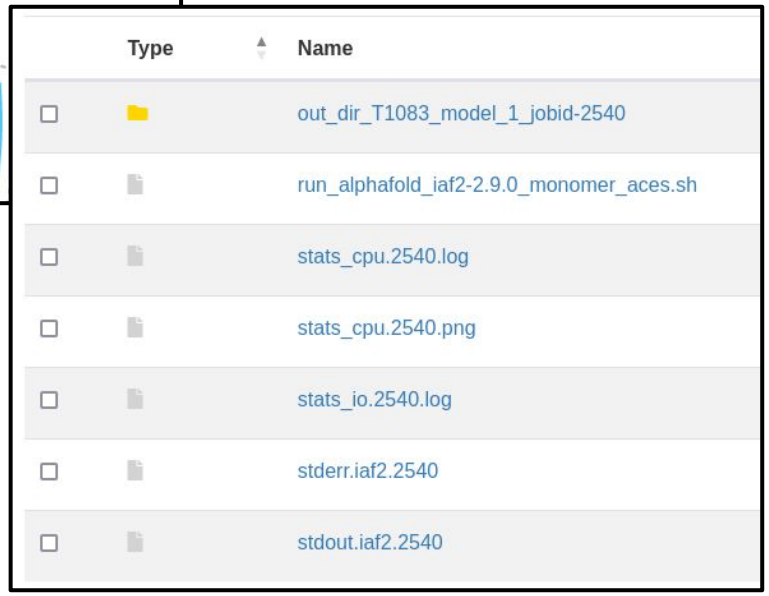
Intel-Optimized AlphaFold on ACES

- Intel's development version of Intel-Optimized AlphaFold
  - source code available on Github
  - install instructions for installing a Conda environment in the run directory
  - two scripts each for monomer and multimer sequence processing
    - preprocessing step
    - inference step
- TAMU HPRC's configuration of Intel-Optimized Alphafold (iaf2)
  - Pipeline steps simplified
    - Intel's pipeline of two scripts per structure prediction run is wrapped with the iaf2 script to run both steps in one command
    - Shared conda environment included
  - Load iaf2 module
    - module load iaf2/.2.0.9
  - Activate conda environment
    - conda activate /sw/hprc/sw/Anaconda3/2022.10/envs/iaf2-2.9.0

Viewing Results

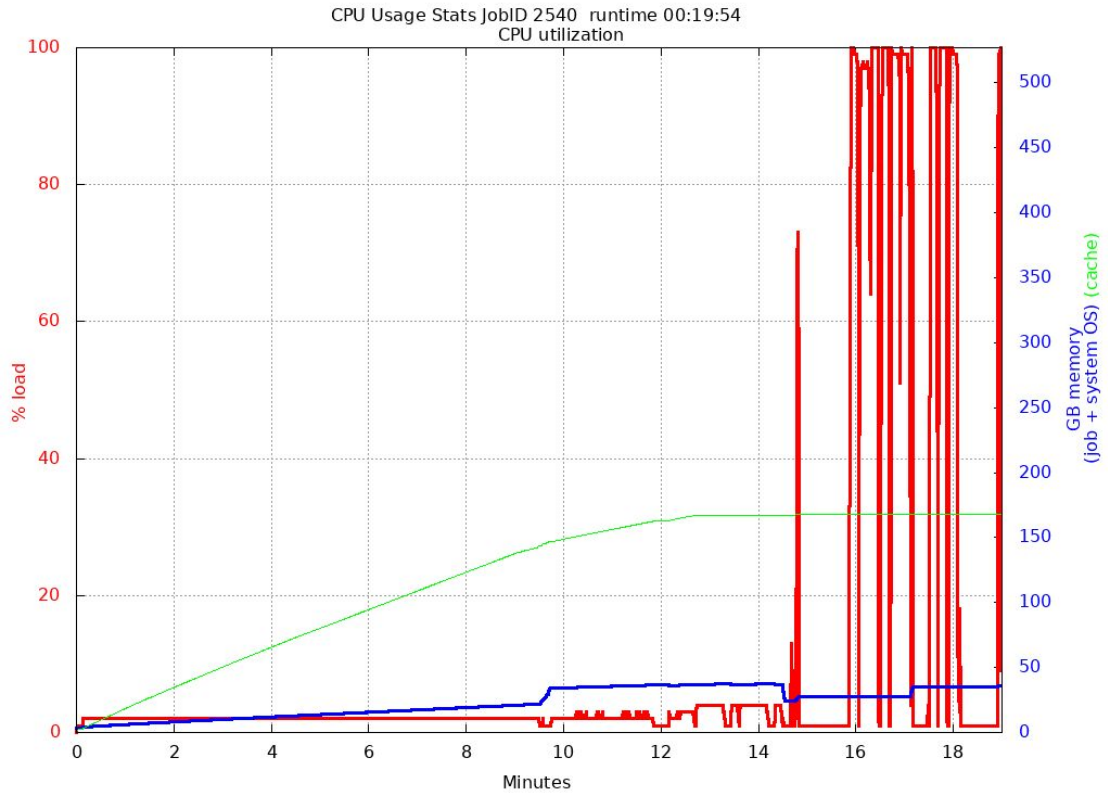
Viewing Results Files





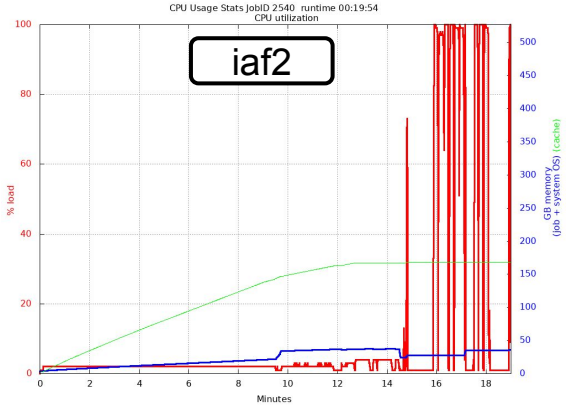
- Navigate to your $SCRATCH/workshop directory
- click .png file to view resource usage plot
- predicted model file found in output directory
  - **`IL2Y/unrelaxed_model_1.pdb`**
- can view .pdb files using ChimeraX

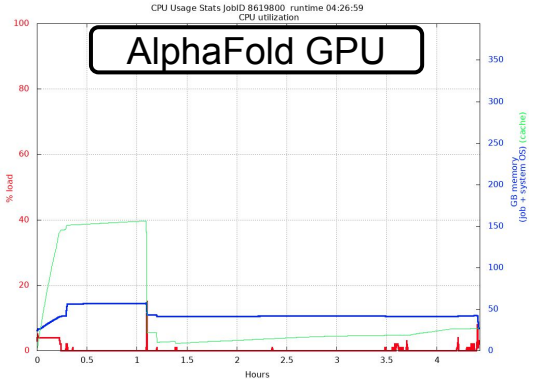
Intel-Optimized AlphaFold CPU Resource Usage

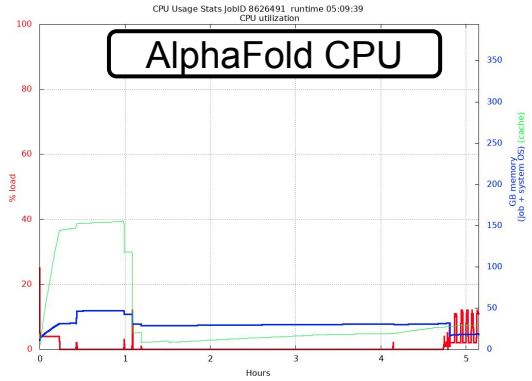


Monomer (20 aa) Run Times

	iaf2	AlphaFold 2.2.0 Grace 1 x A100 GPU	AlphaFold 2.2.0 Grace CPU-only
cluster:cores	ACES:96	Grace:48	Grace:48
model	model_1	models 1-5	models 1-5
runtime	19 min 54 sec	4 hr 27 min	5 hr 9 min







High Performance
Research Computing
DIVISION OF RESEARCH

https://hprc.tamu.edu

HPRC Helpdesk:

help@hprc.tamu.edu
Phone: 979-845-0219

Help us help you. Please include details in your request for support, such as, Cluster (ACES, FASTER, Grace, Terra, ViDaL), NetID (UserID), Job information (JobID(s), Location of your jobfile, input/output files, Application, Module(s) loaded, Error messages, etc), and Steps you have taken, so we can reproduce the problem.