Alphafold2(v2.2) on NUS HPC GPU Cluster

By Ku Wee Kiat Al Engineer/HPC Specialist Research Computing, NUS IT

Content

- Register for NUS HPC Account
- Accessing the HPC Environment
 - Upload/Downloading Data
- Resources
- Running Alphafold2 on NUS HPC GPU Cluster
- The PBS Job Scheduler
 - Submitting a Job
 - $\circ \qquad {\sf Checking Jobs}$
- From Start to End

Register for NUS HPC Account

Go to:

https://nusit.nus.edu.sg/services/hpc/getting-started-hpc/register-for-hpc/

and fill in the necessary information.

Guide: https://nusit.nus.edu.sg/services/getting-started/registration-guide/



Access

- Login via ssh to NUS HPC login nodes
 - o atlas9
- If you are connecting from outside NUS network, please connect to VPN first
 - http://webvpn.nus.edu.sg

Access

OS	Access Method	Command
Linux	ssh from terminal	<pre>ssh nusnet_id@atlas9-c01.nus.edu.sg</pre>
MacOS	ssh from terminal	ssh username@hostname
Windows	ssh using mobaxterm or putty or terminal (powershell)	ssh username@hostname



File Transfer

- 1. Mobaxterm built-in sftp client
- 2. Filezilla client
- 3. Linux/Mac OS/Windows Terminal Tools
 - a. scp
 - b. rsync
 - c. sftp

Filezilla

- Download the Filezilla client for your computer's operating system: <u>https://filezilla-project.org/download.php</u>
- Log in
 - Host: sftp://atlas9.nus.edu.sg
 - Username: Your NUSNET ID
 - Password: Your NUSNET Password
- When prompted to "Trust this host" -> Click OK



Uploading a File

- On the center-right panel, enter the path to your working directory in Remote Site box and hit Enter
- Drag and Drop files you want to upload here

atus: Disconnected from s atus: Connecting to adjust	Username: ever russedung	Password: 1	Port Ourseconnect *			
atas: Connected to alles0 atas: Ratriening directory i atas: Listing directory hor atas: Directory listing of */	e01.nus.edu.sg inting noimu/coxkek tome/buu/cokkek* successful					
ocal site: /Users/wk/			Remote site: /horne/svu/coek	avk.		
			v 2)			_
h b Walanas			v 2 home			
bin			Y 🙁 643			
C6445			P Deciona			
der 📒						
🕨 📒 etz						
home						
ret						
P private						
in the	1 I I					
1 10				7		
lename A	LUCa	Last modified	Fierene A	кеn		Owner/Oreap
				(CII	iote	
.Trash		• 24:	.Temporaryiterr*		#**	ccekwk ex
and an address and a second se	an ot o		cache		· ·	ccekwk ex
Joint Carl						
hash_sees	recto	102200	.conda	rec	tories	ccekwk ex
tast, see DI	recto		conta conta D	rec	tories	ccekwk ex
Lasti user		1C51730 12:22 0809/21 1807544	conta config datas	rec	tories	ccekwk sx ccekwk sx ccekwk sx
Anatomic Antioned Ant		03/30/20 08:55: 03/30/20 08:55:	conda config deus encocid	Directory	CONCILES	ccekwk sx ccekwk sx ccekwk sx ccekwk sx
Anathana Anathanathana Anathana Anathana Anathana Anathana Anathana Anathan	Directory Directory Directory	1223 12222 URCH121 10201544 02(20)20 08:55 08(21)21 11:38:49 02(21)21 11:38:49 02(21)21 11:38:49	conda config deus erracsid tot accent	Directory Directory Directory	07/22/19 14: drax 01/22/21 17:1 drax 05114/21 10:1 drax	ccekwk sx ccekwk sx ccekwk sx ccekwk sx ccekwk sx
Anathiana conda config cups docker emacs.d	Directory Directory Directory Directory	00/01/21 11:00:00 00/01/21 11:00:00 00/01/21 11:00:00 00/02/20 10:07:10 00/02/20 10:07:10 00/02/20 10:07:10	conta conta deus arracsd tat goant goant	Directory Directory Directory Directory	07/29/18 14 drwx 01/29/21 17:1 drwx 05/14/21 16:1 drwx	ccekwk sx ccekwk sx ccekwk sx ccekwk sx ccekwk sx ccekwk sx
According conding c	Directory Directory Directory Directory Directory Directory	04204221 102010 04204221 102010 04204220 08355 0420421 11282-09 0722/20 1025710 0420421 1024109	conta conta dese arrest dest arrest doont goont goont	Directory Directory Directory Directory Directory	07/22/18 14: drace 01/26/21 17:1 drace 05/14/21 16:1 drace 10/77/18 16:0 drace 01/13/19 16:1 drace	ccekwk tx ccekwk tx ccekwk tx ccekwk tx ccekwk tx ccekwk tx ccekwk tx
Actives Loss Lossi, see Los conta	Directory Directory Directory Directory Directory Directory Directory	2522 452 45	conta conta dese erreced tot goont goont goont goont goont goont goont goont	Directory Directory Directory Directory Directory Directory	07/23/18 14: dvxc 01/22/21 17:1 dvxc 05/14/21 16:1 dvxc 13/17/18 16:0 dvxc 01/19/19 16:1 dvxc 01/19/21 16: dvxc	ccekwk sx ccekwk sx ccekwk sx ccekwk sx ccekwk sx ccekwk sx ccekwk sx ccekwk sx
conta conta conta docker eracod hopaw jocal matpictils autov	Directory Directory Directory Directory Directory Directory Directory Directory	04/07/27 14:38:34 03/27/20 03:55 03/27/27 14:38:34 07/27/27 14:38:34 07/27/27 14:38:34 07/27/27 14:38:34 03/27/27 14:38:34 03/27/27 14:38:34 03/27/27 14:38:34 03/27/27 14:38:34 03/27/27 14:38:34	conta otras d stras d stras d goont goont gauge garreener-0.10	Directory Directory Directory Directory Directory Directory Directory	07/23/18 14: druce 07/23/18 14: druce 05/14/21 16:1 druce 05/14/21 16:1 druce 01/11/18 16:0 druce 01/11/18 16:0 druce 01/12/22 11: druce 01/22/21 14: druce	coekwk tex coekwk tex coekwk tex coekwk tex coekwk tex coekwk tex coekwk tex coekwk tex
Londo and Londo Londo Config Copis doder emacud hopaw Jocal matpiotilio putty c	ECCLO Birectory Directory Directory Directory Directory Directory Directory Directory Directory	2522 252 2522 2	config dbss dbss dctr gcont gcont grupp gr	Directory Directory Directory Directory Directory Directory Directory	07/29/18 Mi drvc 01/29/21 171 drvc 05/14/21 161 drvc 05/14/21 161 drvc 01/11/18 HGD drvc 01/20/21 14 drvc 01/20/21 14 drvc 02/20/21 14 drvc	coelwk ex coelwk ex coelwk ex coelwk ex coelwk ex coelwk ex coelwk ex coelwk ex coelwk ex
Janty uses Lanty uses Londa confa doker erracs.d Jocal Jocal party / / seh	Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory	2222 URANIZ2 TROTAGE COLORDOR DESEL COLORDOR DESEL COLORDOR DESEL COLORDOR DESEL-1 COLORDOR DESEL-1 C	conda config dess enrace d fut goont goont goont goreenee-0.10 goreenee-0.10 goreenee-0.10 goreenee-0.10 goreenee-0.10 goreenee	Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory	07/22/19 Mi drue 01/22/21 17: drue 01/22/21 17: drue 01/17/91 19:0 drue 01/22/21 14: drue 01/22/21 14: drue 00/02/19 15:1 drue 00/02/19 15:1 drue	coelwk ex coelwk ex coelwk ex coelwk ex coelwk ex coelwk ex coelwk ex coelwk ex coelwk ex
Austrusen Dantyusen Coordia docker ereaci.d hopaw Jocal rrespicitio party r seh	Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory	2222 URCHT27 TROTA64 02(2)(2) 41522 02(2)(2) 41524 02(2)(2) 41524 02(2)(2) 41524 02(2)(2) 41524 02(2)(2) 41525 02(2)(2) 4155 02(2)(2) 4155 02(2)(2)(2)(2)(2)(2)(2)(2)(2)(2)(2)(2)(2)	conda dates avascut dates avascut datas da	Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory	COTI22/19 34 drve- 01/22/21 171 drve- 01/14/21 162 drve- 01/11/39 162 drve- 01/11/39 162 drve- 01/22/21 14 drve- 00/22/21 14 drve-	coelwek ex coelwek ex
Autoreau Danty uses Config Cups docker ertacs.cl Accal reation focal reation focal reation focal sch sch sch sch sch sch sch sch sch sch	Directory Directory	17330 16222 UQ47123 THC/TA4 UQ42122 01556 UQ42123 TH384.99 0772220 E105710 087A421 0955621 0400231 F105100 0400232 1106120 0400232 1106120 0400232 1106120 0400232 11062462 0400232 1106420 0400232 1106420 0400232 1106420 0400232 1106420 0400232 1106420 0400232 1106420 040023 1106420 040000000000	conda data data data data data data data d	Cecc Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory	COPIES 44 druc- 01/28/211771 druc- 01/28/211771 druc- 01/19/201718 druc- 01/19/201718 druc- 01/28/2118 druc- 01/28/2118 druc- 01/28/2118 druc- 01/28/2118 druc- 01/28/2118 druc-	coelwak ss coelwak ss
Abitives Diaty uses conda coorda cups docker enacs.d hogaw Jocal matpicella matpicella docker enacs.d hogaw Jocal matpicella docker enacs.d hogaw Jocal matpicella docker enacs.d hogaw Jocal matpicella docker enacs.d hogaw Jocal matpicella docker enacs.d hogaw Jocal matpicella docker enacs.d hogaw Jocal matpicella docker enacs.d hogaw Jocal matpicella docker enacs.d hogaw Jocal Matpicella docker enacs.d hogaw Jocal Matpicella docker enacs.d hogaw Jocal Matpicella docker enacs.d Jocal Matpicella docker enacs.d Jocal Matpicella docker enacs.d Jocal Matpicella docker enacs.d Jocal Seal Jocal Seal Jocal Seal Jocal Seal Jocal Jo	Linessy Directory	2222 02221 02222 0222	conda dates	Cecc Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory	COCIES 177 drac- 002623177 drac- 002623177 drac- 002623177 drac- 002623177 drac- 00171479145 drac- 00171479145 drac- 0026279145 drac- 0026279145 drac- 0026279145 drac- 0026279145 drac- 0026279145 drac- 0026279145 drac-	CONVAR SE. CONVAR SE. CONVAR SE. CONVAR SE. CONVAR SE. CONVAR SE. CONVAR SE. CONVAR SE. CONVAR SE. CONVAR SE.
Autoreau Data, sees Conda Cups docker eracs.d hoppaw Jocal ματηγ ratabiotis dramit othermit othermit othermit othermit othermit othermit 3 lites and 1 if erections. Tota errors/cond/fre	Linessy Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory	2222 Wington Wolf Ma Magazo Balk Nataria Walf Ma Nataria Wa	conda dasa dasa dasa dasa dasa dasa dasa d	Ceccory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory	COCI 1625 07/23/19 14:dwc 0(02/21177)dwc 0(02/21177)dwc 0(02/21177)dwc 0(02/2118)dwc 0(02/2118)dwc 0(02/2118)dwc 0(02/2118)dwc 0(02/218)dwc 0(02/218)dwc 0(02/218)dwc 0(02/218)dwc 0(02/218)dwc	CONVAR SE. CONVAR SE.
Autorean Databusen Coordig Cups docker Artasot.d Docal Party ratabiolitik Artasotiti Sabaretisin 13 files and 13 directiones. Toda Serverstool file	Linessay Diressay	2222 Wiley tar Wolf Ma Wolf 2022 Wiley tar Wolf Ma Wolf 2020 Wolf	conda datas	Cecc Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory	COCICE 44. dwe- 01/22/19 15:1. dwe- 01/22/19 15:1. dwe- 01/22/19 15:0. dwe- 19/7/19 16:0. dwe- 01/19 15:1. dwe- 01/19 15:1. dwe- 01/22/19 15:. dwe- 01/22/19 15:. dwe- 01/22/19 15:. dwe- 01/22/29 15:. dwe- 01/22/29 15:. dwe- 01/22/29 15:. dwe- 01/22/29 15:. dwe- 01/22/29 15:. dwe-	contract sec. contract sec.
Autoreal Aut	Linesay Diresay	2002 2013 2014 2	ansa data data data data data data data da	recory Directory Directory Directory Directory Directory Directory Directory Directory Directory	COCIES 101. dww 002621171. dww 002621171. dww 930708163. dww 001793191. dww 00262114. dww 00262114. dww 0026219197. dww 102620917. dww 10262091. dww	CONIAN ES CONIAN ES

Filename	File	size Filetyne	Last modified	Permissions	Owner/Grou
	T IIC	Size Thetype	Last mounica	T CITILISSIONS	- owner/orod
DL		Directory	10/27/20 13:	drwx	ccekwk sv.
Traffic		Directory	08/15/21 02:	drwxr-xr-x	ccekwk sv.
📒 alphafold		Directory	08/31/21 15:	drwxr-xr-x	ccekwk sv.
in bin		Directory	02/26/19 09:	drwx	ccekwk sv.
dataset		Directory	04/19/21 15:	drwx	ccekwk sv
📒 dcgan		Directory	12/20/19 12:0	drwx	ccekwk sv
📒 imagenet		Directory	03/26/19 13:	drwx	ccekwk sv
tiny-imagenet-200		Directory	02/09/15 08:	drwxrwxr-x	ccekwk sv
tiny_imagenet		Directory	12/25/19 17:3	drwx	ccekwk sv
objectives.txt	2	511 txt-file	08/31/21 16:	-rw	ccekwk sv

Remote site: /scratch2/ccekwk

ccekwk

home

Downloads





Resources: Hardware

GPU Clusters

• 9 nodes x 4 Nvidia Tesla V100-32GB

No internet access on Volta Servers

Resources: Hardware/Storage

Directories	Feature	Disk Quota	Backup	Description
/home/svu/\$USERID	Global	20 GB	Snapshot	Home Directory. U:drive on your PC.
/hpctmp/\$USERID	Local on All Atlas/Volta cluster	500 GB	No	Working Directory. Files older than 60 days are purged automatically
/scratch/\$USERID	Local to each Volta node	5 TB	No	For quick read/write access to datasets. Create a folder with your NUSNET ID. Routinely purged.
/scratch2/\$USERID	Available on Atlas 9 and Volta Cluster	1 TB	No	For quick read/write access to datasets. Create a folder with your NUSNET ID. Routinely purged.

Note: Type "hpc s" to check your disk quota for your home directory



Batch Jobs

Queue Resources

Max RAM = 142gb Max No. of CPU cores = 20 Max No. of GPUs = 2 Max Walltime = 72:00:00 Minimum No. of CPU cores = 5 Minimum No. of GPU = 1 Default Walltime = 04:00:00 Request CPU Core in increments of 1

Modes

- Monomer Full DBs
- Monomer Reduced DBs
- Multimer Full DBs
- Multimer Reduced DBs

Sample Job Script (Monomer)

For Alphafold2 Monomer Batch Jobs

Note:

- 1. **Do not copy and paste** the job script in the next slide directly into your text editor.
- 2. Please **type it out manually** to avoid hidden characters.
- 3. Ensure that the jobscript <u>conforms to</u> <u>BASH syntax.</u>

#!/bin/bash Orange is user configurable **#PBS** - P alphafold project name #PBS -i oe Green is updatable #PBS -N alphafold_job_name Black is fixed #PBS -g volta gpu #PBS -l select=1:ncpus=10:mem=100gb:ngpus=1 #PBS -l walltime=15:00:00 cd \$PBS_0_WORKDIR; np=\$(cat \${PBS NODEFILE} | wc -l); Path to your fasta file. No spaces allowed ##----- THE ONLY PART FOR YOU TO CHANGE -----## User settings INPUT_FASTA_FILE_PATH=wcrC_39.fasta; ## "my_abc123.fasta" is your input *.fasta file. ## "alphafold_output_1" defines output folder name. Path to your desired output OUTPUT DIR=`pwd`/alphafold22 output 1; folder. MAX TEMPLATE DATE='2022-03-30' - vvvv-mm-dd format No spaces allowed. MULTIMER_PREDICTIONS_PER_MODEL=5 DB PRESET=full dbs # db_presets: full_dbs, reduced_dbs MODEL PRESET='monomer' # model preset: monomer, monomer casp14, monomer ptm, multimer remove to not set a limit on PDB template date or just # Create output directory date limit as needed mkdir -p \${OUTPUT_DIR} ##----- END TO CHANGE -----## Might not need to change until there is a newer version ## ## Change this to a newer alphafold container when it is released IMAGE=/app1/common/singularity-img/3.0.0/alphafold/alphafold_v2.2.sif ## END ## ## DO NOT CHANGE ## # Params and DB paths ALPHAFOLD_DATA_PATH=/scratch2/biodata/alphafold/database/ ALPHAFOLD DATA PATH2=/scratch2/biodata/alphafold/database v2 2/ ALPHAFOLD MODELS=/scratch2/biodata/alphafold/database v2 2/params Continued on next slide ## Do not remove ## mkdir -p alphafold/common cp -n /scratch2/biodata/alphafold/setup_files/stereo_chemical_props.txt alphafold/common/

Orange is user configurable Green is updatable Black is fixed

##--- To start & run Alphafold2 in Singularity container. ---singularity run --nv \ -B \$ALPHAFOLD DATA PATH:/data \ -B \$ALPHAFOLD DATA PATH2:/data2 \ -B \$ALPHAFOLD_MODELS \ -B .:/etc \ --pwd `pwd` \$IMAGE \ --data dir=/data \ --output dir=\$OUTPUT DIR \ --fasta_paths=\$INPUT_FASTA_FILE_PATH \ --uniref90_database_path=/data/uniref90/uniref90.fasta \ --mgnify_database_path=/data/mgnify/mgy_clusters.fa --bfd database path=/data/bfd/bfd metaclust clu complete id30 c90 final seq.sorted opt \ --uniclust30 database path=/data/uniclust30/uniclust30 2018 08/uniclust30 2018 08 \ --template mmcif dir=/data2/pdb mmcif/mmcif files \ --pdb70_database_path=/data/pdb70/pdb70 \ --obsolete_pdbs_path=/data2/pdb_mmcif/obsolete.dat \ --num_multimer_predictions_per_model=\$MULTIMER_PREDICTIONS_PER_MODEL \ --model preset=\$MODEL PRESET \ --max_template_date=\$MAX_TEMPLATE_DATE \ --run relax=True \ --use gpu relax=True \ --db_preset=\$DB_PRESET > stdout.\$PBS_JOBID 2> stderr.\$PBS_JOBID

Sample Job Script (Monomer Reduced DBs)

For Alphafold2 Monomer Reduced DBs Batch Jobs

Note:

- 1. **Do not copy and paste** the job script in the next slide directly into your text editor.
- 2. Please **type it out manually** to avoid hidden characters.
- 3. Ensure that the jobscript <u>conforms to</u> <u>BASH syntax.</u>

#!/bin/bash Orange is user configurable **#PBS** - P alphafold project name #PBS -i oe Green is updatable #PBS -N alphafold_job_name Black is fixed #PBS -g volta gpu #PBS -l select=1:ncpus=10:mem=100gb:ngpus=1 #PBS -l walltime=15:00:00 cd \$PBS_0_WORKDIR; np=\$(cat \${PBS NODEFILE} | wc -l); Path to your fasta file. No spaces allowed ##----- THE ONLY PART FOR YOU TO CHANGE -----## User settings INPUT_FASTA_FILE_PATH=wcrC_39.fasta; ## "my_abc123.fasta" is your input *.fasta file. ## "alphafold_output_1" defines output folder name. Path to your desired output OUTPUT DIR=`pwd`/alphafold22 output 1; folder. MAX TEMPLATE DATE='2022-03-30' - vvvv-mm-dd format No spaces allowed. MULTIMER_PREDICTIONS_PER_MODEL=5 DB_PRESET=reduced_dbs # db_presets: full_dbs, reduced_dbs MODEL PRESET='monomer' # model preset: monomer, monomer casp14, monomer ptm, multimer remove to not set a limit on PDB template date or just # Create output directory date limit as needed mkdir -p \${OUTPUT_DIR} ##----- END TO CHANGE -----## Might not need to change until there is a newer version ## ## Change this to a newer alphafold container when it is released IMAGE=/app1/common/singularity-img/3.0.0/alphafold/alphafold_v2.2.sif ## END ## ## DO NOT CHANGE ## # Params and DB paths ALPHAFOLD_DATA_PATH=/scratch2/biodata/alphafold/database/ ALPHAFOLD DATA PATH2=/scratch2/biodata/alphafold/database v2 2/ ALPHAFOLD MODELS=/scratch2/biodata/alphafold/database v2 2/params Continued on next slide ## Do not remove ## mkdir -p alphafold/common cp -n /scratch2/biodata/alphafold/setup_files/stereo_chemical_props.txt alphafold/common/

Orange is user configurable Green is updatable Black is fixed

##--- To start & run Alphafold2 in Singularity container. ---singularity run --nv \ -B \$ALPHAFOLD DATA PATH:/data \ -B \$ALPHAFOLD DATA PATH2:/data2 \ -B \$ALPHAFOLD_MODELS \ -B .:/etc \ --pwd `pwd` \$IMAGE \ --data dir=/data \ --output dir=\$OUTPUT DIR \ --fasta_paths=\$INPUT_FASTA_FILE_PATH \ --uniref90_database_path=/data/uniref90/uniref90.fasta \ --mgnify_database_path=/data/mgnify/mgy_clusters.fa --small_bfd_database_path=/data2/small_bfd/bfd-first_non_consensus_sequences.fasta \ --template mmcif dir=/data2/pdb mmcif/mmcif files \ --pdb70 database path=/data/pdb70/pdb70 \ --obsolete_pdbs_path=/data2/pdb_mmcif/obsolete.dat \ --num_multimer_predictions_per_model=\$MULTIMER_PREDICTIONS_PER_MODEL \ --model preset=\$MODEL PRESET \ --max template date=\$MAX TEMPLATE DATE \ --run relax=True \ --use_gpu_relax=True \ --db preset=\$DB PRESET > stdout.\$PBS JOBID 2> stderr.\$PBS JOBID

Sample Job Script (Multimer)

For Alphafold2 **Multimer** Batch Jobs

Note:

- 1. **Do not copy and paste** the job script in the next slide directly into your text editor.
- 2. Please **type it out manually** to avoid hidden characters.
- 3. Ensure that the jobscript <u>conforms to</u> <u>BASH syntax.</u>

#!/bin/bash Orange is user configurable #PBS -P alphafold_project_name #PBS -i oe Green is updatable **#PBS** -N alphafold job name #PBS -q volta_gpu Black is fixed #PBS -l select=1:ncpus=10:mem=100gb:ngpus=1 #PBS -1 walltime=20:00:00 cd \$PBS 0 WORKDIR; np=\$(cat \${PBS_NODEFILE} | wc -l); Path to your fasta file. ##----- THE ONLY PART FOR YOU TO CHANGE -----No spaces allowed ## User settings INPUT_FASTA_FILE_PATH=P10_trimer.fasta; ### "my_abc123.fasta" is your input *.fasta file. OUTPUT DIR=`pwd`/alphafold22 output 1 m: - ## "alphafold output 1" defines output fold@atmtameour desired output folder. No spaces allowed. MAX_TEMPLATE_DATE='2022-03-30' # yyyy-mm-dd format MULTIMER PREDICTIONS PER MODEL=5 DB PRESET=full dbs # db presets: full dbs, reduced dbs MODEL PRESET='multimer' # model preset: monomer, monomer casp14, monomer ptm, multimer # Create output directory mkdir -p \${OUTPUT DIR} ##----- FND TO CHANGE ------## Might not need to change until there is a newer version ## ## Change this to a newer alphafold container when it is released IMAGE=/app1/common/singularity-img/3.0.0/alphafold/alphafold v2.2.sif ## END ## ## DO NOT CHANGE ## # Params and DB paths ALPHAFOLD DATA PATH=/scratch2/biodata/alphafold/database/ ALPHAFOLD_DATA_PATH2=/scratch2/biodata/alphafold/database_v2_2/ ALPHAFOLD_MODELS=/scratch2/biodata/alphafold/database_v2_2/params ## Do not remove ## Continued on next slide mkdir -p alphafold/common cp -n /scratch2/biodata/alphafold/setup_files/stereo_chemical_props.txt alphafold/common/

Orange is user configurable Green is updatable Black is fixed

##--- To start & run Alphafold2 in Singularity container. ---singularity run --nv \ -B \$ALPHAFOLD DATA PATH:/data \ -B \$ALPHAFOLD DATA PATH2:/data2 \ -B \$ALPHAFOLD MODELS \ -B .:/etc \ --pwd `pwd` \$IMAGE \ --data dir=/data \ --output dir=\$OUTPUT DIR \ --fasta paths=\$INPUT FASTA FILE PATH \ --uniref90 database path=/data/uniref90/uniref90.fasta \ --mgnify_database_path=/data/mgnify/mgy_clusters.fa --bfd database path=/data/bfd/bfd_metaclust_clu_complete_id30_c90_final_seq.sorted_opt \ --uniclust30 database path=/data/uniclust30/uniclust30 2018 08/uniclust30 2018 08 \ --template mmcif dir=/data2/pdb mmcif/mmcif files \ --obsolete pdbs path=/data2/pdb mmcif/obsolete.dat \ --pdb_segres_database_path=/data2/pdb_segres/pdb_segres.txt \ --uniprot_database_path=/data2/uniprot/uniprot.fasta \ --num_multimer_predictions_per_model=\$MULTIMER_PREDICTIONS_PER_MODEL \ --model preset=\$MODEL PRESET \ --max_template_date=\$MAX_TEMPLATE_DATE \ --run relax=True \ --use_gpu_relax=True \ --db preset=\$DB PRESET > stdout.\$PBS JOBID 2> stderr.\$PBS JOBID

Sample Job Script (Multimer Reduced DBs)

For Alphafold2 Multimer Reduced DBs Batch Jobs

Note:

- 1. **Do not copy and paste** the job script in the next slide directly into your text editor.
- 2. Please **type it out manually** to avoid hidden characters.
- 3. Ensure that the jobscript <u>conforms to</u> <u>BASH syntax.</u>

#!/bin/bash Orange is user configurable #PBS -P alphafold_project_name #PBS -i oe Green is updatable **#PBS** -N alphafold job name #PBS -q volta_gpu Black is fixed #PBS -l select=1:ncpus=10:mem=100gb:ngpus=1 #PBS -1 walltime=20:00:00 cd \$PBS 0 WORKDIR; np=\$(cat \${PBS_NODEFILE} | wc -l); Path to your fasta file. ##----- THE ONLY PART FOR YOU TO CHANGE -----No spaces allowed ## User settings INPUT_FASTA_FILE_PATH=P10_trimer.fasta; ### "my_abc123.fasta" is your input *.fasta file. OUTPUT DIR=`pwd`/alphafold22 output 1 m: - ## "alphafold output 1" defines output fold@atmtameour desired output folder. No spaces allowed. MAX_TEMPLATE_DATE='2022-03-30' # yyyy-mm-dd format MULTIMER PREDICTIONS PER MODEL=5 DB PRESET=full dbs # db presets: full dbs, reduced dbs MODEL PRESET='multimer' # model preset: monomer, monomer casp14, monomer ptm, multimer # Create output directory mkdir -p \${OUTPUT DIR} ##----- FND TO CHANGE ------## Might not need to change until there is a newer version ## ## Change this to a newer alphafold container when it is released IMAGE=/app1/common/singularity-img/3.0.0/alphafold/alphafold v2.2.sif ## END ## ## DO NOT CHANGE ## # Params and DB paths ALPHAFOLD DATA PATH=/scratch2/biodata/alphafold/database/ ALPHAFOLD_DATA_PATH2=/scratch2/biodata/alphafold/database_v2_2/ ALPHAFOLD_MODELS=/scratch2/biodata/alphafold/database_v2_2/params ## Do not remove ## Continued on next slide mkdir -p alphafold/common cp -n /scratch2/biodata/alphafold/setup_files/stereo_chemical_props.txt alphafold/common/

Orange is user configurable Green is updatable Black is fixed

##--- To start & run Alphafold2 in Singularity container. ---singularity run --nv \ -B \$ALPHAFOLD DATA PATH:/data \ -B \$ALPHAFOLD DATA PATH2:/data2 \ -B \$ALPHAFOLD MODELS \ -B .:/etc \ --pwd `pwd` \$IMAGE \ --data dir=/data \ --output dir=\$OUTPUT DIR \ --fasta paths=\$INPUT FASTA FILE PATH \ --uniref90 database path=/data/uniref90/uniref90.fasta \ --mgnify_database_path=/data/mgnify/mgy_clusters.fa --small_bfd_database_path=/data2/small_bfd/bfd-first_non_consensus_sequences.fasta \ --template mmcif dir=/data2/pdb mmcif/mmcif files \ --obsolete pdbs path=/data2/pdb mmcif/obsolete.dat \ --pdb segres database path=/data2/pdb segres/pdb segres.txt \ --uniprot_database_path=/data2/uniprot/uniprot.fasta \ --num_multimer_predictions_per_model=\$MULTIMER_PREDICTIONS_PER_MODEL \ --model_preset=\$MODEL PRESET \ --max template date=\$MAX TEMPLATE DATE \ --run relax=True \ --use gpu relax=True \ --db_preset=\$DB_PRESET > stdout.\$PBS_JOBID 2> stderr.\$PBS_JOBID

Differing Flags

Additional DB paths/flags to include if using multimer --pdb_seqres_database_path=/data2/pdb_seqres/pdb_seqres.txt \ --uniprot_database_path=/data2/uniprot/uniprot.fasta \

Additional DB path/flags to include is using monomer/not multimer --pdb70_database_path=/data/pdb70/pdb70 \

Reduced DBs:

--small_bfd_database_path=/data2/small_bfd/bfd-first_non_consensus_sequences.fasta \

Configure MAX_TEMPLATE_DATE in 'yyyy-mm-dd' format to set the cutoff time point prior to the release date of structures.

Wrong: image = /path/to/container/ INPUT_FASTA_FILE_PATH = abc.fasta INPUT_FASTA_FILE_PATH = my abc.fasta

Correct
image=/path/to/container
INPUT_FASTA_FILE_PATH=abc.fasta
INPUT_FASTA_FILE_PATH=my_abc.fasta

PBS Job Scheduler

Submitting a Job

Steps

You have to run:

- 1. Prepare your fasta in your working directory
- 2. Create a PBS job script and save it in your working directory
 - a. Example job scripts are in the following 2 slides
- 3. Submit PBS job script to PBS Job Scheduler

Server will run:

- 1. Job is in PBS Job Scheduler queue
- 2. Job Scheduler waits for server resources to be available
- 3. If available, Job Scheduler runs your script on remote gpu server

Submitting a Job

Save your job script (previous slides for examples) in a text file (e.g. train.pbs) then run the following commands

shell\$ qsub train.pbs
675674.venus01

Job Status

shell\$ qstat -xfn

venus01:

Req'd Req'd Elap Job ID Jobname SessID NDS TSK Memory Time S Time Username Queue 669468.venus01 ccekwk volta cifar_noco 1 20gb 24:00 F 1 --_ _ 1 674404.venus01 ccekwk volta cifar_noco 1 20gb 24:00 F - -- -TestVM/0 675674.venus01 ccekwk volta cifar noco 1 1 20gb 24:00 Q ----_ _

Statuses: Q(ueue), F(inish), R(unning), E(nding), H(old)

Submitting a Job

[ccekwk@atlas8-0 697978.venus01 [ccekwk@atlas8-0	c01 class: c01 class:	ification ification]\$ qsub tra:]\$ qstat -x	in.pbs fn						
venus01:								2 33		-1
1-L TD		0	7-1-2	C TD	NIDO	TOK	Req'd	Req'd	~	Elap
JOD ID	Username	Queue	Jobname	Sessin	NDS	ISK	memory	Time	ъ	lime
					222	222	11221	11111	5	
695126.venus01	ccekwk	azgpu	citar_noco	111	1	4	40gb	24:00	F	7.7
697978.venus01 TestVM/0*4	ccekwk	azgpu	cifar_noco		1	4	40gb	24:00	R	
[ccekwk@atlas8-0	c01 class:	ification]\$ 📙							

Statuses: Q(ueue), F(inish), R(unning), E(nding), H(old)

Job Chaining and Dependencies

Execute jobs in sequence

- qsub -W depend=afterok:<Job-ID> <JOB SCRIPT>
 - o qsub -W depend=afterany:836578.venus01 volta_benchmark.pbs
- Job script <QSUB SCRIPT> will be submitted after the Job, <Job-ID> is successfully completed. Useful options to "depend=..." are:
 - **afterok**:<Job-ID> Job is scheduled if the Job <Job-ID> exits without errors or is successfully completed.
 - **afternotok**:<Job-ID> Job is scheduled if the Job <Job-ID> exited with errors.
 - **afterany**:<Job-ID> Job is scheduled if the Job <Job-ID> exits with or without errors

ccekwk@atlas9-c0 836578_venus01	1 /hpctmp)/ccekwk/1	tf_benchmar	ks/conta	aine	r\$ q	sub vol	ta_bend	chma	ark.pt	5	
ccekwk@atlas9-c0 ccekwk@atlas9-c0 .pbs	1 /hpctmp 1 /hpctmp)/ccekwk/1)/ccekwk/1	tf_benchmar tf_benchmar	ks/conta ks/conta	aine aine	r\$ v . r\$ q	im volt sub -W	a_benct depend=	ma af	rk.pbs terany	s y:836578.venus01	volta_benchmark
836582.venus01 ccekwk@atlas9-c0	1 /hpctmp	o/cce <mark>kwk</mark> /1	tf_benchmar	ks/conta	aine	r\$ q	stat -fi	ns1				
venus01:							Pogld	Pogld		Elan		
Job ID	Username	Queue	Jobname	SessID	NDS	TSK	Memory	Time	S	Time		
836578.venus01 Job run at Th	ccekwk u Apr 18	volta_lo at 09:00	<pre>gpu_tf_ben on (volta0)</pre>	392016 1:ncpus	1 =20:n	20 nem=	80gb 8388608	04:00 0kb:ngp	R (00:01 	volta01/0*20	
836582.venus01	ccekwk	volta_lo	gpu_tf_ben		1	20	80gb	04:00	H			

Useful PBS Commands

Action	Command
Job submission	qsub my_job_script.txt
Job deletion	qdel my_job_id
Job listing (Simple)	qstat
Job listing (Detailed)	qstat -ans1
Queue listing	qstat -q
Completed Job listing	qstat -H
Completed and Current Job listing	qstat -x
Full info of a job	qstat -f job_id

Checking Jobs

Log Files

• Output (stdout)

• stdout.\$PBS_JOBID

- Error (stderr)
 - stderr.\$PBS_JOBID
- Job Summary
 - o job_name.o\$PBS_JOBID

[ccekwk@atlas	s8-c01 (lassi	fication]	s ls	-1	- Z	ă ă
total 16604							
-rw 1	ccekwk	admin	15325	Nov	12	12:24	cifar10_resnet.py
- rw 1	ccekwk	admin	865	Nov	12	12:26	cifar_nocont.o697978
drwx 2	ccekwk	admin	348	Nov	12	12:28	logs
-rw 1	ccekwk	admin	13589605	0ct	19	11:04	logs.tar.gz
drwx 3	ccekwk	admin	456	Sep	21	16:04	mnist
drwxr-xr-x 2	ccekwk	admin	98	Nov	12	12:26	saved_models
-rw 1	ccekwk	admin	1209	Nov	12	12:26	stderr.697978.venus01
-rw 1	ccekwk	admin	62224	Nov	12	12:26	stdout.697978.venus01
- rw 1	ccekwk	admin	832	0ct	3	13:29	tf_gcpu24.pbs
-rw 1	ccekwk	admin	849	Sep	28	16:39	tf.pbs
-rw 1	ccekwk	admin	612	0ct	1	08:55	train_gpu_container.pb
-rw 1	ccekwk	admin	300	Nov	8	13:21	train.pbs
[ccekwk@atlas	s8-c01 (lassi	fication]	ş 📕			

From Start to End

Setting up

ssh nusnet_id@atlas9

mkdir /scratch2/`whoami`

cd /scratch2/`whoami`

mkdir alphafold_workdir

cd alphafold_wokrdir

nano jobscript.txt # opens a text editor

Paste in sample job script available in /app1/common/alphafold/samples_jobscript

Crtl+x -> y -> Enter | Save your jobscript



Submitting your Alphafold2 Job

Back to the terminal

qsub jobscript.txt

Your job is now submitted

Check job status

qstat -xfn

Job Complete, Retrieve Results

Remember your output directory set in the job script?

OUTPUT_DIR=`pwd`/alphafold_output_5

You can find it in: /scratch2/your_nusnet_id/alphafold_workdir/alphafold_output_5

You can now download the output folder using filezilla

*OUTPUT_DIR might differ, please refer to the actual job script used

	emote site: /scratch2	2/ccekwk/alp	hafold/alphaf	old_output_5/wci	rC_39		
_	📫 alpha	fold_output					
	<mark>?</mark> alpha	fold_output_1					
	<mark>?</mark> alpha	fold_output_2					
	? alpha	fold_output_3					
	<mark>?</mark> alpha	fold_output_4					
	🔻 📒 alpha	fold_output_5					
	🕨 🔚 wo						
	<mark>?</mark> alpha	fold_output_6					
	<mark>?</mark> alpha	fold_output_7_l	ogin				
	<mark>?</mark> alpha	fold_output_8_I	ogin				
	? alpha	fold_output_9_I	ogin				
	? logs						
File	ename 🔨	Filesize	Filetype	Last modified	Permissions	Owner/Group	
	msas		Directory	08/26/21 18:	drwxr-xr-x	ccekwk sv	
	features.pkl	1501425	pkl-file	08/26/21 18:	-rw-rr	ccekwk sv	
	ranked_0.pdb	480338	pdb-file	08/26/21 18:	-rw-rr	ccekwk sv	
8	ranking_debug.json	102	Azure Dat	08/26/21 18:	-rw-rr	ccekwk sv	
	relaxed_model_1.p	480338	pdb-file	08/26/21 18:	-rw-rr	ccekwk sv	
	result_model_1.pkl	51043067	pkl-file	08/26/21 18:	-rw-rr	ccekwk sv	
		100	Aruro Dot	00/26/21 10.	-rw-rr	ccekwk sv	
	timings.json	182	Azure Dat	00/20/21 10	1.66 1.1		

7 files and 1 directory. Total size: 202383075 bytes

Acknowledgement of Usage of NUS HPC Resources

Our primary mission is to provide the best of class, high-performance computing resources to support your computational research needs free of charge. To continuously improve the service, anticipate future demands, and keep track of our HPC facility's impact on the NUS research community in general, we request you to cite the REC team in your published research works.

Below is an example of a citation that may work for you:

"We would like to acknowledge that computational work involved in this research work is partially / fully supported by NUS IT's Research Computing group"

We would appreciate if you could send us a copy of your publication as well.

General Support nTouch https://ntouch.nus.edu.sg/ux/myitapp/#/catalog/home

Project/Research Collaboration or Long Term Engagement Email dataengineering@nus.edu.sg