THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Beyond the Command Line:

Broadening the scope of HPC Usage

Sarah Walters

Introduction

- I have been with the Research Computing Centre (RCC) for several years as part of the infrastructure team
- RCC has domain experts in a broad range of fields
- Extensive support for researchers through training, help desk, face-to-face drop-in "Hacky Hour" sessions, workflow platforms and more
- RCC provides computing and storage infrastructure for all UQ researchers and supports other QLD universities and NCRIS capabilities and their partners via QCIF

Bunya

UQ's "Evergreen" Supercomputer

- ~14000 CPU cores (most AMD)
- 143 nodes
- 750GB – 4TB RAM per system
- Rocky Linux 8.10
- >1 PB of 'scratch' storage
- Connected to long-term research storage (RDM)

- 6x AMD MI210x 64GB GPUs
- 11x NVIDIA A100 80GB GPUs
- 37x NVIDIA H100 80GB GPUs
- 18x NVIDIA L40 48GB GPUs
- 6x NVIDIA L40s 48GB GPUs
- 36x NVIDIA A16 12GB GPUs (12 boards with 4 GPUs per board)

Why a Visual Desktop on a Supercomputer?

- Workloads that require visualisation traditionally ran on workstations … now they need bigger compute
- Datasets are getting enormous
- Workstations and servers are getting hotter, noisier, and needing more power – they can't sit in your office anymore
- Hardware is increasingly expensive

The Need for Support

- Change in user profiles – lots of people who have never seen a command line
- Local (faculty/institute) teams don't always have the expertise or capacity to grow eResearch capability or securely administer servers

“Much of our current research would be impossible without resources such as the Bunya cluster, and we would be required to look for compute externally at a significant cost.

“If we had to look externally for equivalent compute, this would also likely be without the thorough support that we receive internally via the staff at the Research Computing Centre. Productivity and feasibility of our work would be markedly impacted by the absence of this expertise."

— Dr Dylan Anthony Black

What about … ?

- Our previous supercomputer Wiener, decommissioning soon, ran the Characterisation Virtual Laboratory (CVL)
- CVL was a huge success, but was designed for specific workloads and had some interface limitations
- A separate virtual desktop service would not provide integration with the Supercomputer
- Nectar is a great option for persistent workloads

Why Choose Open OnDemand?

- SC22 – a large community had formed around Open OnDemand
- Single Sign-on Authentication
- Dashboards run as the user – secure and ensures file permissions are enforced
- No custom client software – runs in a web browser
- Adaptive interface – works on a mobile phone
- Lots of customisation options

Deploying Open OnDemand

- Runs in a single VM
- 16 cores, 64 GB of RAM – only because of workshops!
- Installs easily with packages for EL or Debian/Ubuntu
- SSO with Apache modules – we use OpenID Connect to Keycloak
- Effectively a "login node"

Deploying Open OnDemand

- Choice of Linux desktops – XFCE, MATE, GNOME, KDE
- Need to install TurboVNC and websockify
- VirtualGL for hardware acceleration
- Lots of software! JupyterLab, CryoSparc, Relion, R Studio, Amira, Huygens, MATLAB ...

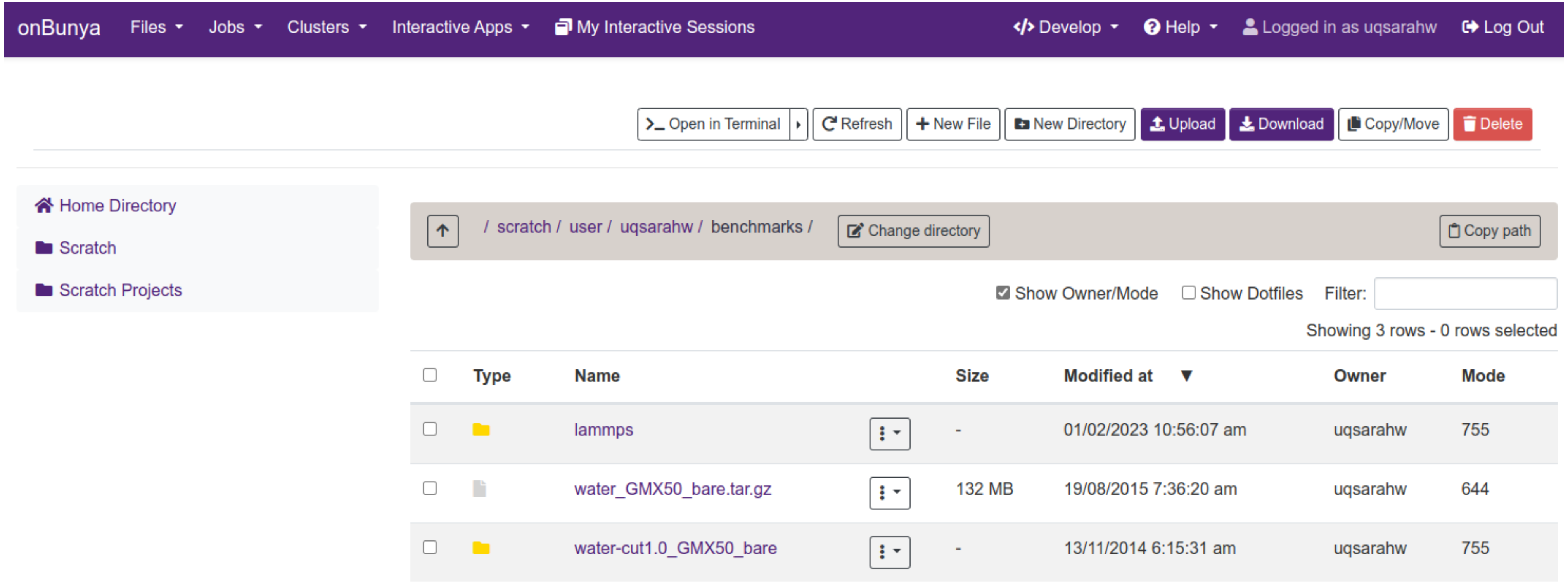
Making it our own

- Lots of configuration options
- We removed the default logo and changed colours to fit UQ branding.
- You can easily add directories to the file browser

```
# Add User scratch space directory
paths << FavoritePath.new("/scratch/user/#{User.new.name}", title: "Scratch")
```

onBunya

"On Demand" implies instant availability, so we renamed the service to "onBunya".



/ scratch / user / uqsarahw / benchmarks /

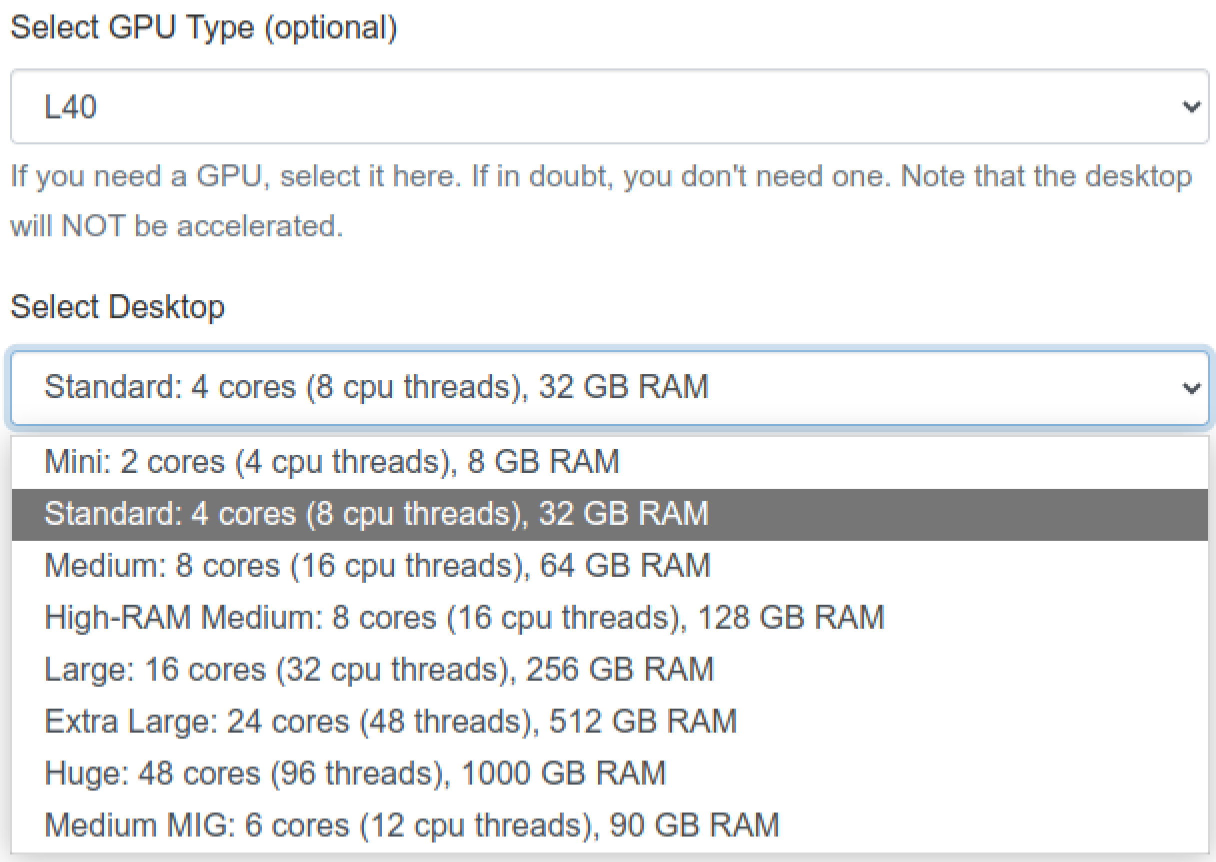
☑ Show Owner/Mode ☐ Show Dotfiles Filter:

Showing 3 rows - 0 rows selected

☐	Type	Name		Size	Modified at ▼	Owner	Mode
☐	📁	lammps	⋮ ▾	-	01/02/2023 10:56:07 am	uqsarahw	755
☐	📄	water_GMX50_bare.tar.gz	⋮ ▾	132 MB	19/08/2015 7:36:20 am	uqsarahw	644
☐	📁	water-cut1.0_GMX50_bare	⋮ ▾	-	13/11/2014 6:15:31 am	uqsarahw	755

Customising Forms

- Open OnDemand defaults allow researchers to choose any combination of memory, CPU, or GPU
- This creates a packing problem
- We tuned our forms to control the choices
- We did this by modifying the form code, but some changes can be done with configuration options



Select GPU Type (optional)

L40

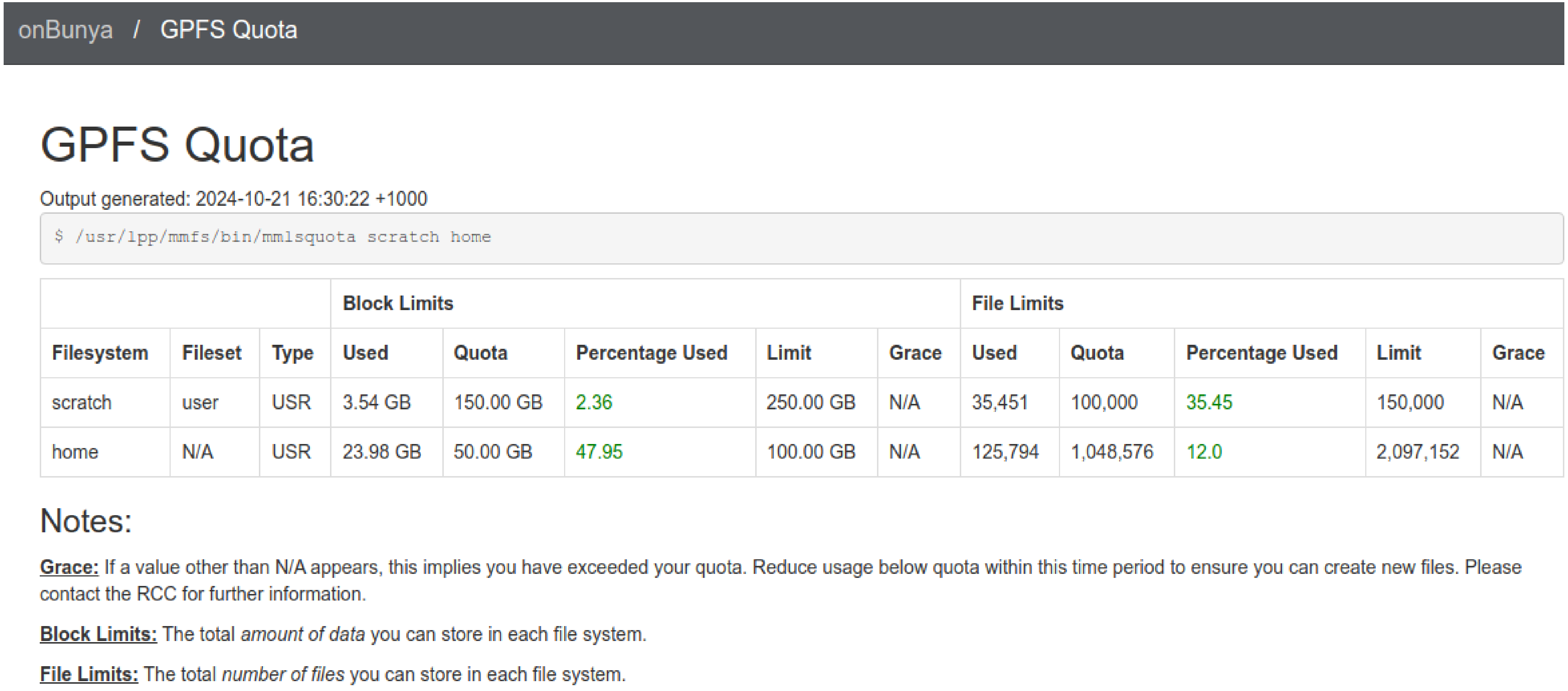
If you need a GPU, select it here. If in doubt, you don't need one. Note that the desktop will NOT be accelerated.

Select Desktop

Standard: 4 cores (8 cpu threads), 32 GB RAM

Mini: 2 cores (4 cpu threads), 8 GB RAM
Standard: 4 cores (8 cpu threads), 32 GB RAM
Medium: 8 cores (16 cpu threads), 64 GB RAM
High-RAM Medium: 8 cores (16 cpu threads), 128 GB RAM
Large: 16 cores (32 cpu threads), 256 GB RAM
Extra Large: 24 cores (48 threads), 512 GB RAM
Huge: 48 cores (96 threads), 1000 GB RAM
Medium MIG: 6 cores (12 cpu threads), 90 GB RAM

Custom Apps



onBunya / GPFS Quota

GPFS Quota

Output generated: 2024-10-21 16:30:22 +1000

```
$ /usr/lpp/mmfs/bin/mmlsquota scratch home
```

			Block Limits					File Limits				
Filesystem	**Fileset**	**Type**	**Used**	**Quota**	**Percentage Used**	**Limit**	**Grace**	**Used**	**Quota**	**Percentage Used**	**Limit**	**Grace**
scratch	user	USR	3.54 GB	150.00 GB	2.36	250.00 GB	N/A	35,451	100,000	35.45	150,000	N/A
home	N/A	USR	23.98 GB	50.00 GB	47.95	100.00 GB	N/A	125,794	1,048,576	12.0	2,097,152	N/A

Notes:

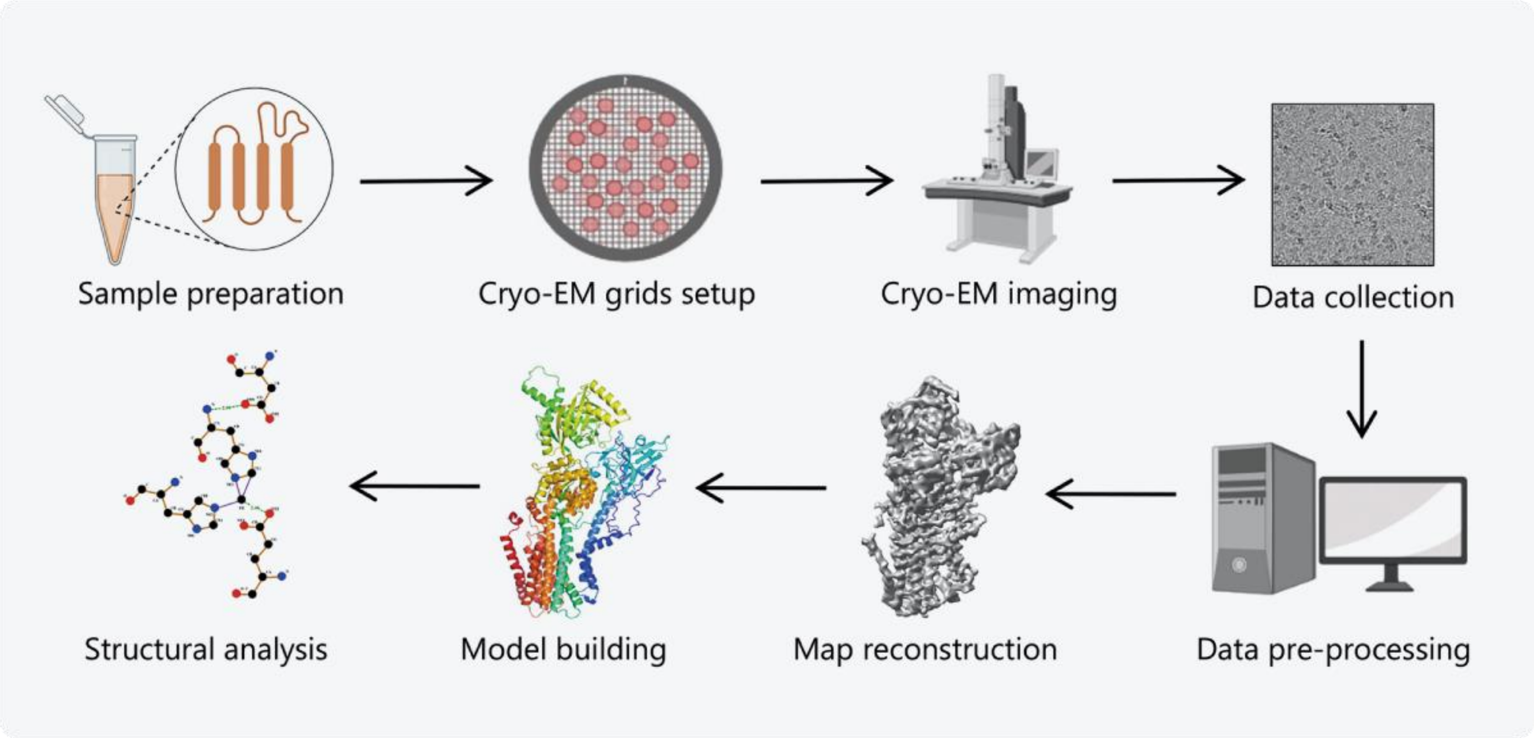
Grace: If a value other than N/A appears, this implies you have exceeded your quota. Reduce usage below quota within this time period to ensure you can create new files. Please contact the RCC for further information.

Block Limits: The total *amount of data* you can store in each file system.

File Limits: The total *number of files* you can store in each file system.

Use Case 1: Protein Structure Analysis

- When we understand how proteins work, we can alter their function
- Researching protein structures helps us understand mutations that cause disease. This enables development of targeted drugs

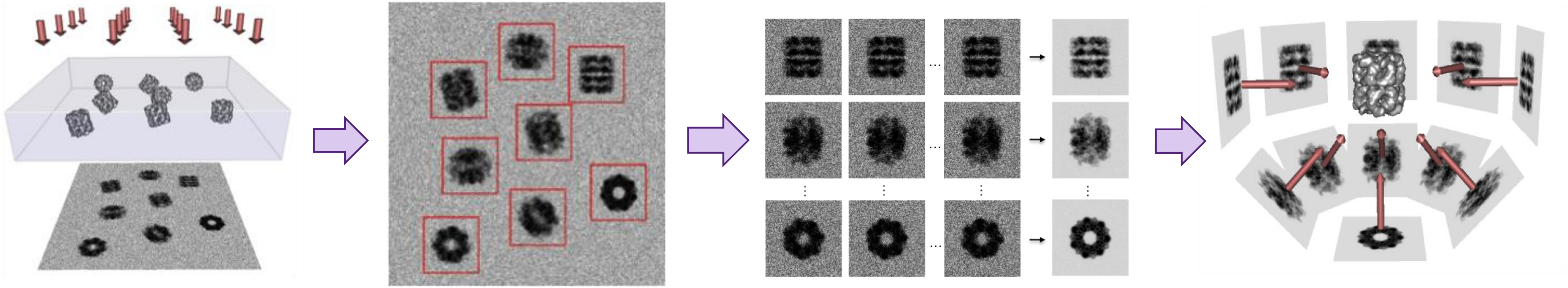


- One method to understand protein structures uses single particle Cryogenic Electron Microscopy (Cryo-EM)

Image Courtesy of Dr Farrah Blades
Adapted from: Zhu, K. F. et al., Military Med Res 10, 10 (2023)

Use Case 1: Protein Structure Analysis

- Compare pixel-by-pixel each individual image from the microscope, similar particle images get classified together
- Researchers carry "good" particle sets forward until they can eventually form a 3-D model. Requires visual intervention.
- Thousands of micrographs, millions of particles – GPU intensive work



Use Case 2: Deconvolution Pipelines

- Researchers in the Queensland Brain Institute (QBI) use Zebrafish to record and model neuron activity in the live brain
- Need to remove noise to get an accurate image
- Instruments are more sensitive: increasing dataset sizes caused software to lock up on their server.

`/scratch/user/uqramor/Out`

Workflow	Image file	Templates used		Status
4	sox6 6ohda25 tdtom h2a647 orf1p488 w5 60x	Eugenia_Diskov	Eugenia_Diskov	Finished
5	x6 6ohda25 tdtom h2a647 orf1p488 w5 60x p	Eugenia_Diskov	Eugenia_Diskov	Finished
6	x6 6ohda25 tdtom h2a647 orf1p488 w5 60x p	Eugenia_Diskov	Eugenia_Diskov	Running
7	x6 6ohda25 tdtom h2a647 orf1p488 w5 60x p	Eugenia_Diskov	Eugenia_Diskov	Queued

```
Workflow 6: Fast Classic MLE: starting...
Workflow 6: SNR:  12.0 (True mode)
Workflow 6: Using solid brick mode.
            Max iterations: 100
            Quality change threshold: 0.010
Workflow 6: Processing using GPU.
```

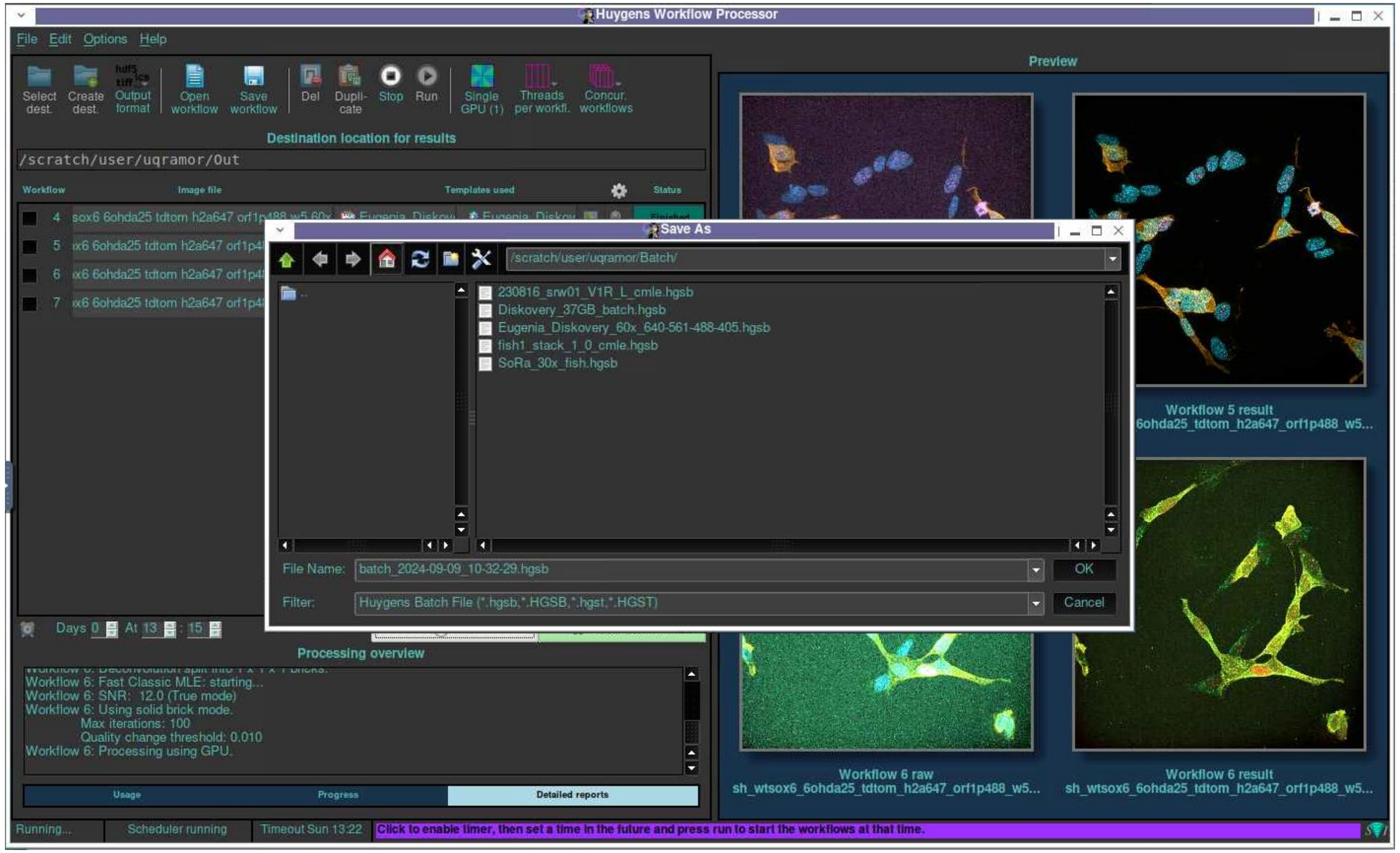
Running... | Scheduler running | Timeout Sun 13:22 | Click to enable timer, then set a time in the future and press run to start the workflows at that time.

Use Case 2: Deconvolution Pipelines

- Some scientific workflows use a lot of trial-and-error.
- Visual tools & feedback make setting up batch jobs easier & less time consuming.

"Although we submit Huygens Core jobs to Slurm, Huygens Professional (with the GUI) on OnDemand is also essential as we need to test the deconvolution parameters before applying everything into a big batch job."

Dr Rumelo Amor, Queensland Brain Institute



Destination location for results: `/scratch/user/uqramor/Out`

/scratch/user/uqramor/Batch/

- 230816_srw01_V1R_L_cmle.hgsb
- Diskovery_37GB_batch.hgsb
- Eugenia_Diskovery_60x_640-561-488-405.hgsb
- fish1_stack_1_0_cmle.hgsb
- SoRa_30x_fish.hgsb

File Name: batch_2024-09-09_10-32-29.hgsb

Filter: Huygens Batch File (*.hgsb,*.HGSB,*.hgst,*.HGST)

Days 0 At 13 : 15

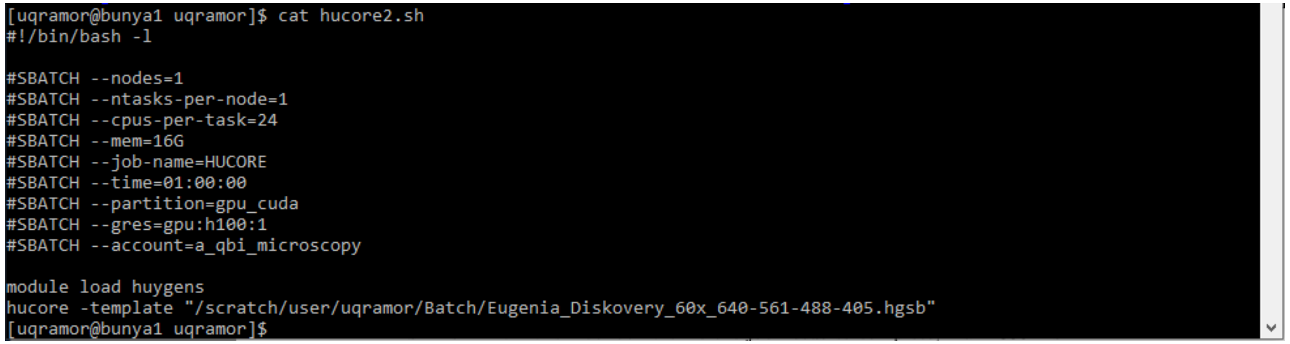
Processing overview

```
Workflow 6: Fast Classic MLE: starting...
Workflow 6: SNR:  12.0 (True mode)
Workflow 6: Using solid brick mode.
            Max iterations: 100
            Quality change threshold: 0.010
Workflow 6: Processing using GPU.
```

Running... | Scheduler running | Timeout Sun 13:22 | Click to enable timer, then set a time in the future and press run to start the workflows at that time.

Use Case 2: Deconvolution Pipelines

It's all about workflows

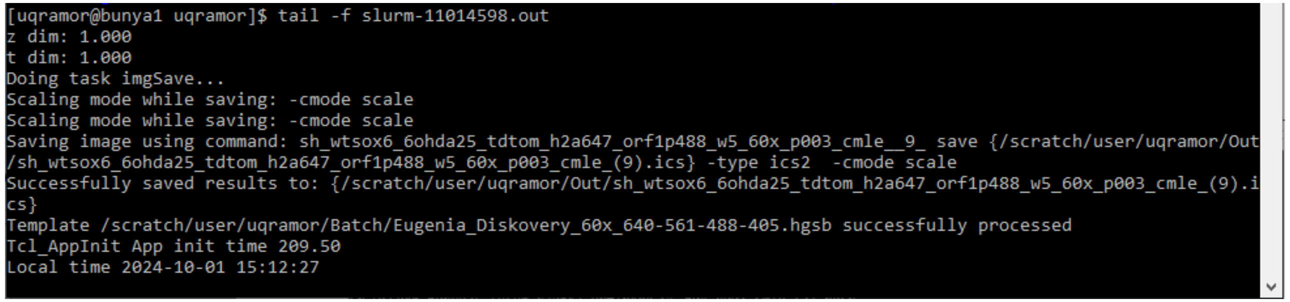


```
[uqramor@bunya1 uqramor]$ cat hucore2.sh
#!/bin/bash -l

#SBATCH --nodes=1
#SBATCH --ntasks-per-node=1
#SBATCH --cpus-per-task=24
#SBATCH --mem=16G
#SBATCH --job-name=HUCORE
#SBATCH --time=01:00:00
#SBATCH --partition=gpu_cuda
#SBATCH --gres=gpu:h100:1
#SBATCH --account=a_qbi_microscopy

module load huygens
hucore -template "/scratch/user/uqramor/Batch/Eugenia_Diskovery_60x_640-561-488-405.hgsb"
[uqramor@bunya1 uqramor]$
```

```
[uqramor@bunya1 uqramor]$ sbatch hucore2.sh
Submitted batch job 10002720
```

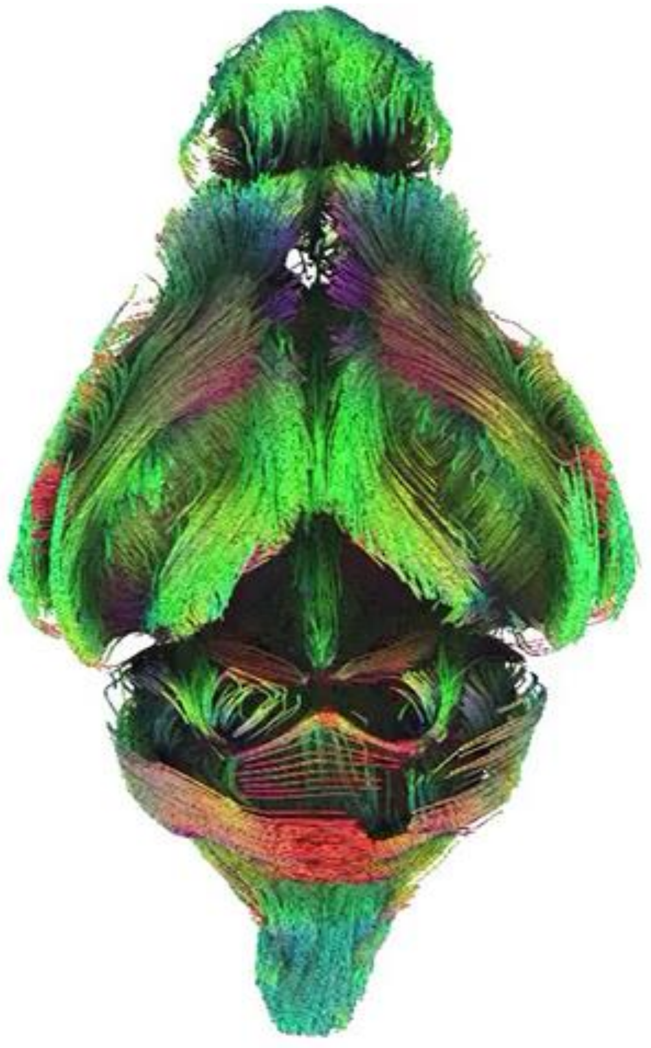


```
[uqramor@bunya1 uqramor]$ tail -f slurm-11014598.out
z dim: 1.000
t dim: 1.000
Doing task imgSave...
Scaling mode while saving: -cmode scale
Scaling mode while saving: -cmode scale
Saving image using command: sh_wtsox6_6ohda25_tdtom_h2a647_orf1p488_w5_60x_p003_cmle__9_ save {/scratch/user/uqramor/Out
/sh_wtsox6_6ohda25_tdtom_h2a647_orf1p488_w5_60x_p003_cmle_(9).ics} -type ics2  -cmode scale
Successfully saved results to: {/scratch/user/uqramor/Out/sh_wtsox6_6ohda25_tdtom_h2a647_orf1p488_w5_60x_p003_cmle_(9).i
cs}
Template /scratch/user/uqramor/Batch/Eugenia_Diskovery_60x_640-561-488-405.hgsb successfully processed
Tcl_AppInit App init time 209.50
Local time 2024-10-01 15:12:27
```

- A prepared template can be submitted to the batch system

- Once processing is complete, the outputs can be viewed, copied into long-term research storage, and automatically available on the researcher's desktop.

Use Case 3: How Brains Develop

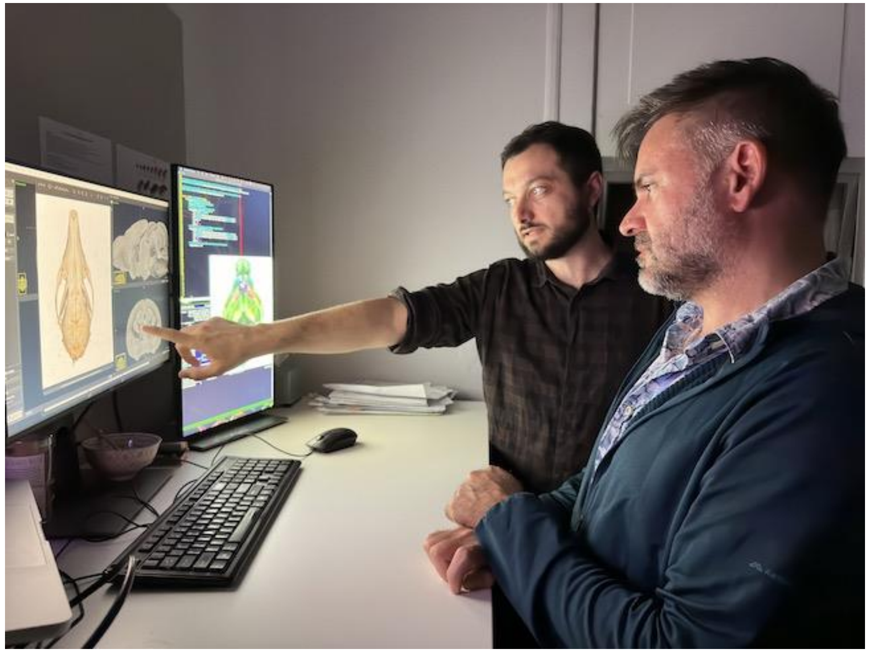


- ***Image:*** *Axial rotations of reconstructed white matter tracts following diffusion-weighted magnetic resonance imaging of an adult Australian marsupial, the fat-tailed dunnart.*
- The Suárez group uses brain scans of marsupials as part of its work to develop a model of mammalian brain development
- Optical and magnetic resonance imaging (MRI) methods are inherently visual

Use Case 3: How Brains Develop

Complex and demanding processing for "human-in-the-loop" workflows

“For our experienced users, onBunya allows for critical visual inspection of data, and integration of GUI-dependent steps into our workflows.



“Importantly, onBunya is also allowing our lab members with less experience to access high-performance computing systems, lowering the barrier to entry, and democratising access to these resources.”

Dr Dylan Anthony Black (left)
with Associate Professor Rodrigo Suarez

Implications for Cluster Management

- People expect to be able to instantly fire up a desktop on demand
- Need tuning to ensure prioritisation without starving batch jobs
- All the usual problems with interactive jobs apply here

Busy Systems

Full queues

Wait Times

Interrupted workflows

Bad Habits

Frustrated users
"reserving" nodes

Busier Systems

Wasted resources
Longer queues

Implications for Hardware

- A100/H100 has no framebuffer – encourage batch on these
- MIG is an option for flexibility (esp. for trial-and-error)
- Mix of L40(s) and A16 nodes for framebuffer
- Visual workloads tend to need more CPU cores and RAM than GPU compute batch jobs
- Demand will grow quickly!

Questions

Sarah Walters

Research Systems Projects & Delivery Manager

sarah.walters@uq.edu.au