ADACS

ASTRONOMY DATA AND COMPUTING SERVICES

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

Astronomy Australia Ltd.

SWINBURNE UNIVERSITY OF TECHNOLOGY

Curtin University

ADACS Introduction

ASTRONOMY DATA AND COMPUTING SERVICES

- **Vision:**
  - ***astronomy-focused*** training, support and expertise to maximise scientific return on investments in astronomical data & computing infrastructure
  - 3 service components:
    1. Training (face-to-face, webinars, internships)
    2. Astronomy software support for the OzSTAR supercomputer
    3. National Support
       - Professional software support
       - Astronomy Supercomputing Time Allocation committee (ASTAC)
       - Data management and collaboration platform (gDMCP)
- **Two nodes:**
  - Swinburne University (Melbourne) - principally responsible for software support
  - Curtin University (Perth) - principally responsible for training
- **Commenced operations March 2017**
- **Funded by Astronomy Australia Limited (AAL) through the astronomy National Collaborative Research Infrastructure Strategy (NCRIS) allocation**

GRAVITATIONAL WAVE DATA CENTRE

- Started operations July 2019
- Staffed @ 8-9 FTE through FY 21/22
- Operates alongside ADACS@Swinburne
- Guided by GWDC Science Advisory Panel
- Five core projects:

  GWCloud, GWLab, SPIIR, MeerTime, GWLandscape
- Provides guaranteed ADACS MAP time to GW researchers

see gwdc.org.au for more details

Merit Allocation Program

ADACS
ASTRONOMY DATA AND COMPUTING SERVICES

Merit-based allocation of professional *software development & training*

- Methodology:
  1. Researchers respond to calls for **Expressions of Interest** (EoIs); 1 page description of project
  2. **ADACS interviews applicant** to coax-out detailed technical specifications
  3. **ADACS develops an assessment** of required development time and skills required
  4. Users complete a **detailed application** and quote the ADACS assessment for their project
  5. An **independent time allocation committee** (TAC) selects projects to be supported, reconciling requested and available resources

Source: imgglobalinfotech.com

see adacs.org.au/software-services/applying-to-the-adacs-merit-allocation-program
for directions EoI submission directions

ADACS@Swinburne: What can we do?

ADACS
ASTRONOMY DATA AND COMPUTING SERVICES

- ~20 professional developers and research software engineers with expertise covering:
  - Scientific computing;
  - High performance computing;
  - Data science & machine learning;
  - Web development;
  - Large-scale scientific databases;
  - Cloud computing/micro-service architectures; and
  - Scientific visualisation.

Source: imgglobalinfotech.com

... augmented by a suitable mix of computationally skilled astronomers and product discovery & delivery professionals:

- *ensuring adequate domain & community knowledge;*
- *trained User eXperience (UX) expertise; and*
- *certified Agile product discovery & delivery*

Software Support Program

ADACS
ASTRONOMY DATA AND COMPUTING SERVICES

Semester	Project Title	Project type
2021A	Collaboration on Shared Codebases	training
2021A	Global Firebal Observatory operations - Ansible	training
2021A	ASVO UX Training	training
2021A	Global Fireball Observatory - Astrometric calibration container	dev ops
2021A	Universal Cosmic Ray Detection with CREDO	mobile app
2021A	Modernisation of Software for Epoch of Reionisation Science	refactor
2021A	3D Heat Transport, Burning, and Evolution	optimisation
2021A	Optimising the SAMI cubing code	optimisation
2021A	MPI Parallelisation of a Boltzmann Solver for Supernova Simulations	optimisation
2020B	GPU Acceleration of the DiFX Software Correlator	optimisation
2020B	Optimizing parallel Bilby (pBilby)	optimisation
2020B	Software Support for GASKAP Imaging	optimisation
2020B	A web based portal for COMPAS	web
2020B	Extending the Data Central Simple Spectra Viewer	web
2020B	Optimisation of the COMPAS rapid binary population synthesis code	optimisation
2020A	Global Cosmic Ray Detection	mobile app
2020A	Rapidly and Optimally Identifying Gravitational-Wave Optical Counterparts for GOTO	database
2020A	Fast becomes Faster: A Full OpenCL rewrite of Corrfunc	optimisation
2020A	Software support for the SMART pulsar survey	web
2019B	ProFound	optimisation
2019B	Final data release of the Australian Dark Energy Survey (OzDES)	web
2019B	Parallelisation of the SoFiA Source Finding Pipeline	optimisation
2019B	NBody and VR	optimisation
2019B	An implementation of the BFDMT in CUDA	optimisation
2019A	Bilby	web
2019A	Spectrum viewer	web
2019A	Building on Bilby-UI: bringing continuous gravitational wave science to the web	web
2019A	A Webapp for modeliing the Galalxy	web
2019A	DWF portal and database	web
2018B	Extended MWA Survey Progress and Monitoring	web
2018B	Bringing LIGO Science to the Masses	web
2018B	Multi-threading ASKAPSoft Synthesis Imaging	optimisation
2018B	Model dispersion with PRISM	optimisation
2018B	Corrfunc -- Blazing Fast Correlation Functions Now on the GPU	optimisation
2018A	An automated data reduction pipeline for AAO Data Central	data
2018A	GPU acceleration of gravitational-wave signal models	optimisation
2018A	Galaxy and black hole co evolution survey using active machine learning	web
2017B	Speeding-up Reionization with GPUs	optimisation
2017B	Building and Supporting User Communities with GPU-Accelerated Computing Services	web

Legend

- Web application
- CPU Optimisation
- GPU Optimisation
- Training
- Other

Training and Material Offered

ADACS
ASTRONOMY DATA AND COMPUTING SERVICES

Face to Face workshops

Over the past 2 years ADACS has delivered more than a dozen face-to-face events, ranging from outreach events, to half-day and multi-day workshops.

LMS

We have been developing a series of webinars to help the community master a variety of topics related to computing for astronomers. These webinars are a short series of informational videos compiled by theme on the ADACS LMS.

Youtube Channel

Our youtube channel (ADACS learning) makes a selection of our LMS webinars available to the public.

Github

Material we teach during these workshops is generally made available via our Github page

Internship

ADACS runs regular internships offering student the possibility to work on software development projects within astronomy

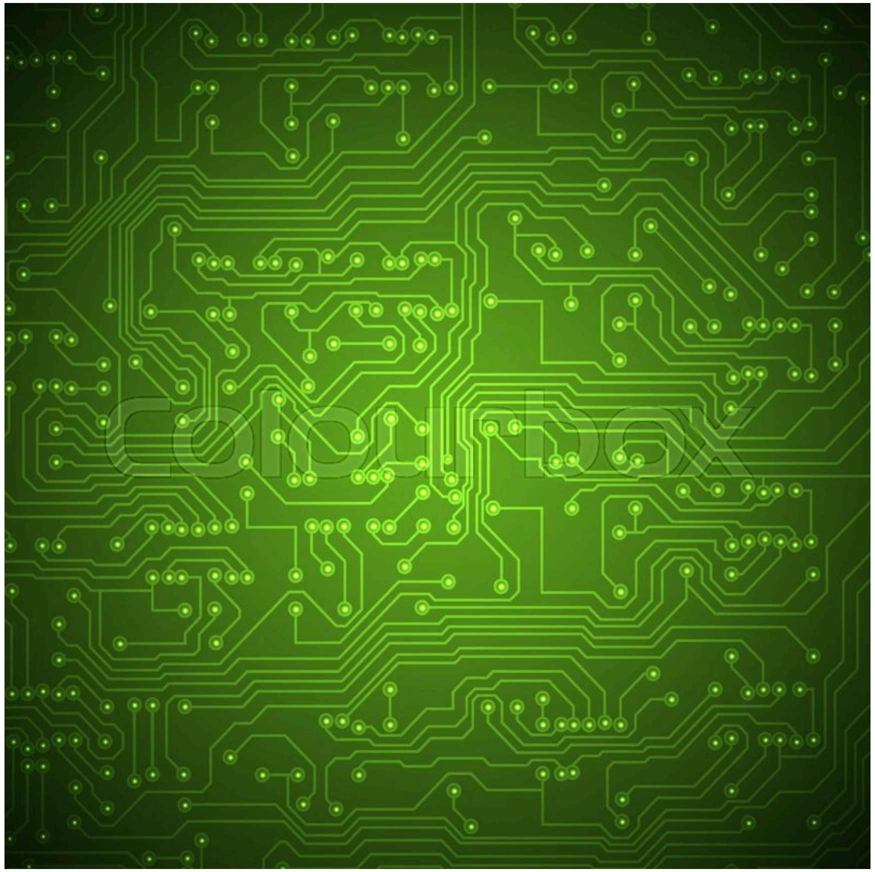
Online Resources

A curated list of tutorials,information and cheat sheets from around the web on key skills for astronomy researchers that are not explicitly covered in our webinars.

ADACS and Green Computing

I would like to explore three themes of how ADACS can be part of a green computing strategy for this community:

1. Professionally developed code:
   ➡ *Eliminate useless compute*
2. Data portals:
   ➡ *Eliminate redundant compute*
3. Software Optimisation:
   ➡ *New capabilities & new responsibilities*



Professionally developed code: Eliminate useless compute

Properly developed code just *works*

Professionally developed code has several ***green advantages***:

- It's tested, so it works more reliably
  ➡ *Less compute lost to flawed execution*
- It's easier to extend, refactor, etc
  ➡ *Less compute lost to wasteful debugging cycles*
- It's easier to build communities around
  ➡ *Less compute lost to the development of multiple redundant codebases*

ADACS

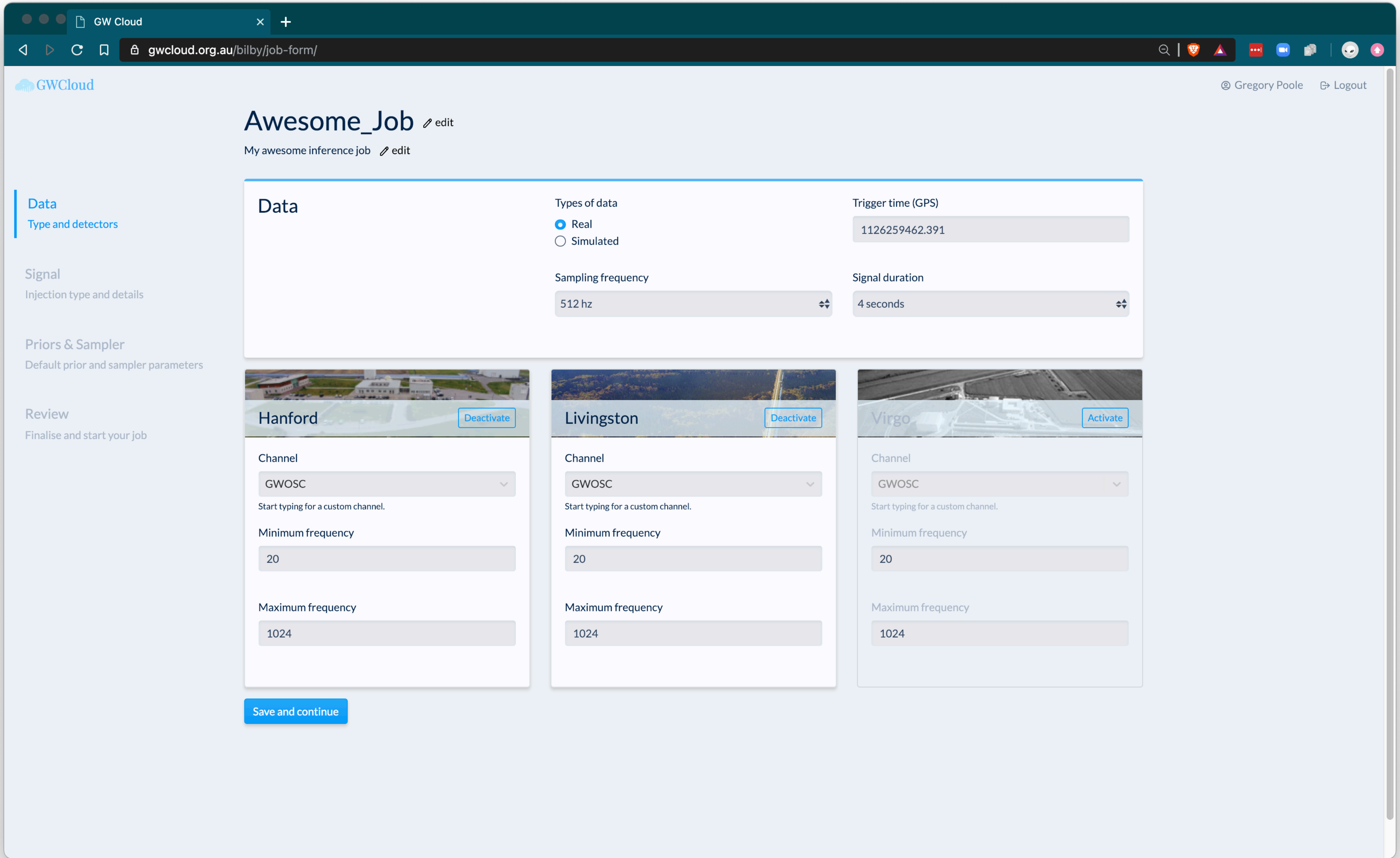
ASTRONOMY DATA AND COMPUTING SERVICES

Data portals:
Eliminate redundant compute

Let me start with an anecdote...

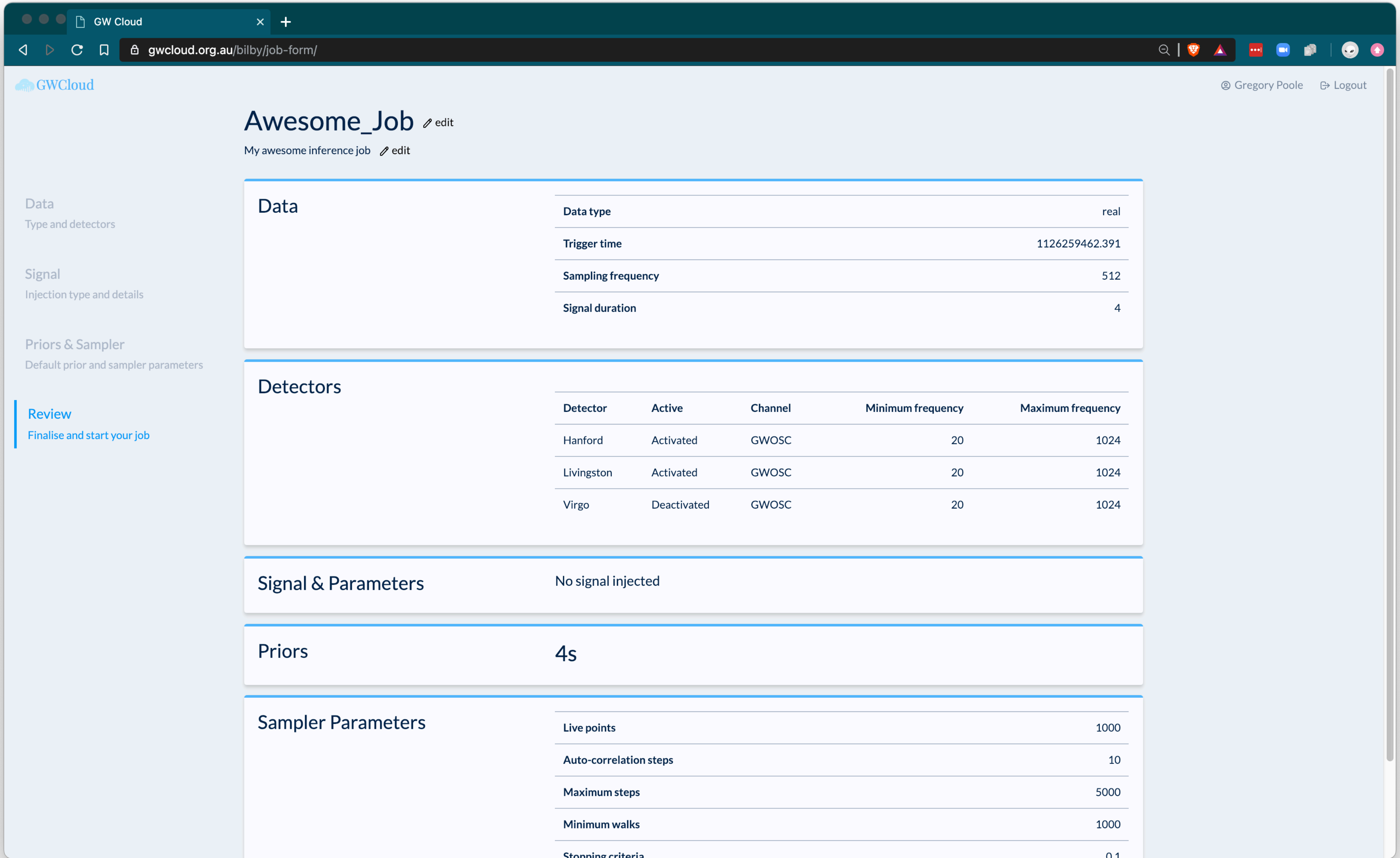
- 200-300 GW astronomers in the LIGO Parameter Estimation (PE) Group have been running and rerunning the **same analyses on the same set of events repeatedly**
- For example, from a study being conducted by Avi Vajpeyi (Monash):
  - Avi has run large suites of inference jobs on 100's-to-1000's of LIGO triggers
  - He estimates **167 tonnes of carbon** (from 1.3M cpu-hrs of compute) were produced during the course of the analysis
  - Of this, **80-90% of that had was repetition of compute** that had been conducted somewhere else already

GWCloud


GW Cloud
gwcloud.org.au/bilby/job-form/
GWCloud
Gregory Poole
Logout
Awesome_Job
edit
My awesome inference job
edit
Data
Type and detectors
Signal
Injection type and details
Priors & Sampler
Default prior and sampler parameters
Review
Finalise and start your job
Data
Types of data
Real
Simulated
Trigger time (GPS)
1126259462.391
Sampling frequency
512 hz
Signal duration
4 seconds
Hanford
Deactivate
Channel
GWOSC
Start typing for a custom channel.
Minimum frequency
20
Maximum frequency
1024
Livingston
Deactivate
Channel
GWOSC
Start typing for a custom channel.
Minimum frequency
20
Maximum frequency
1024
Virgo
Activate
Channel
GWOSC
Start typing for a custom channel.
Minimum frequency
20
Maximum frequency
1024
Save and continue

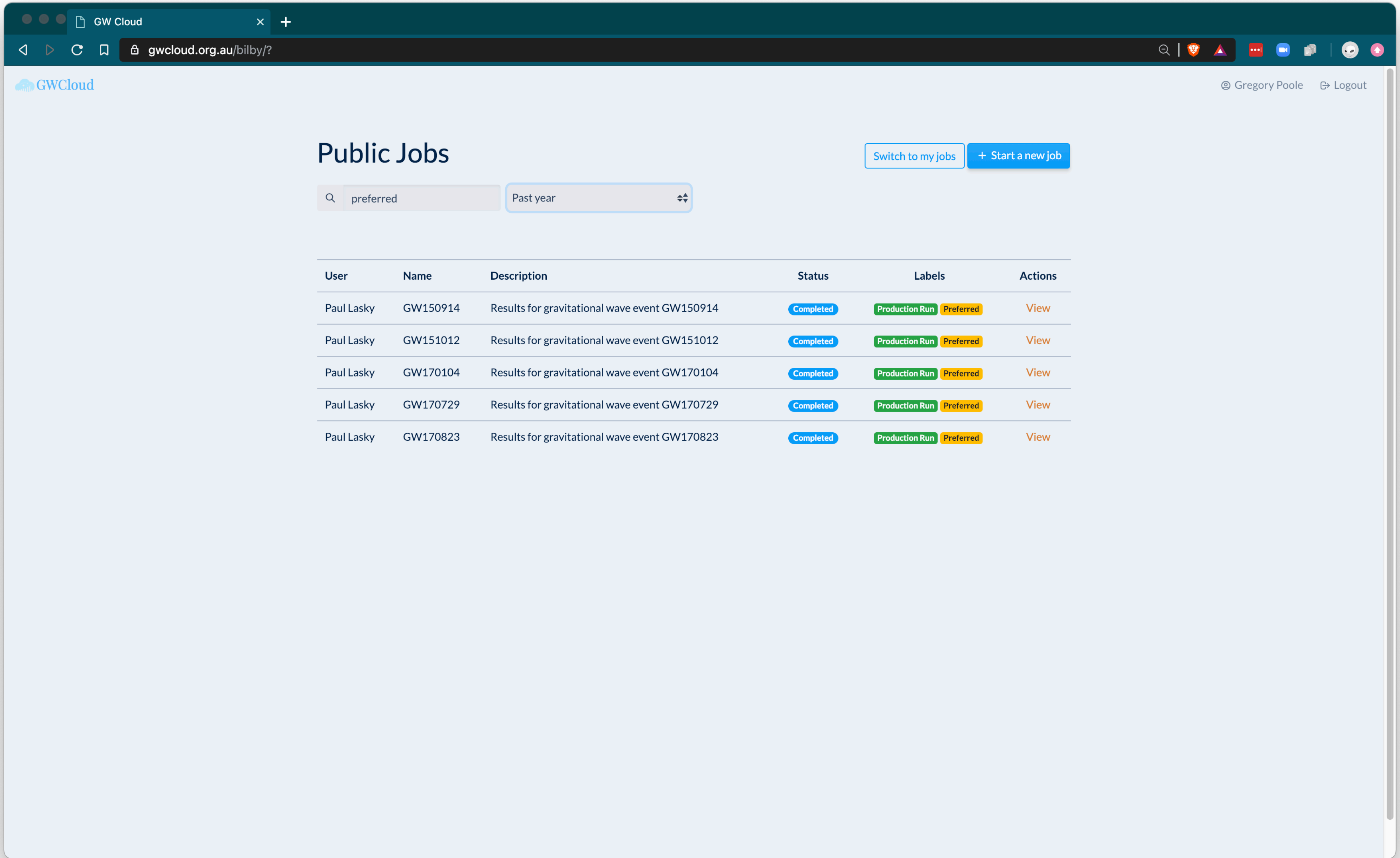
GWCloud

ASTRONOMY DATA AND COMPUTING SERVICES



GWCloud

ADACS
ASTRONOMY DATA AND COMPUTING SERVICES



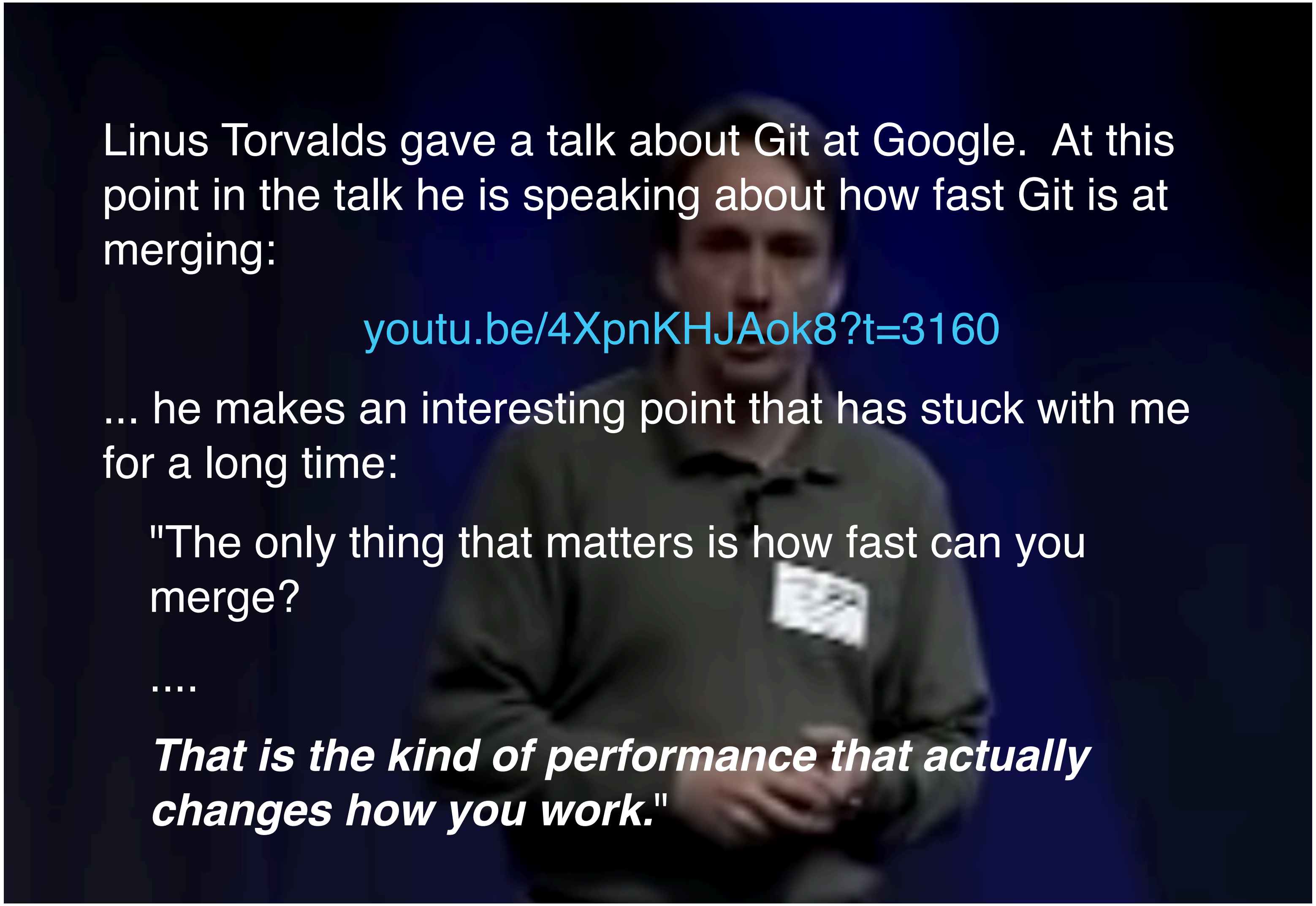
GWCloud

local ⌘1 local ⌘2

```python
from gwcloud_python import GWCloud

gwc_token = 'token_here'

gwc = GWCloud(token=gwc_token)

jobs, errors = gwc.get_preferred_job_list()

for job in jobs:
    print(job)
```

NORMAL script.py[+] pyt… 100% ≡ 10/10 ln : 1

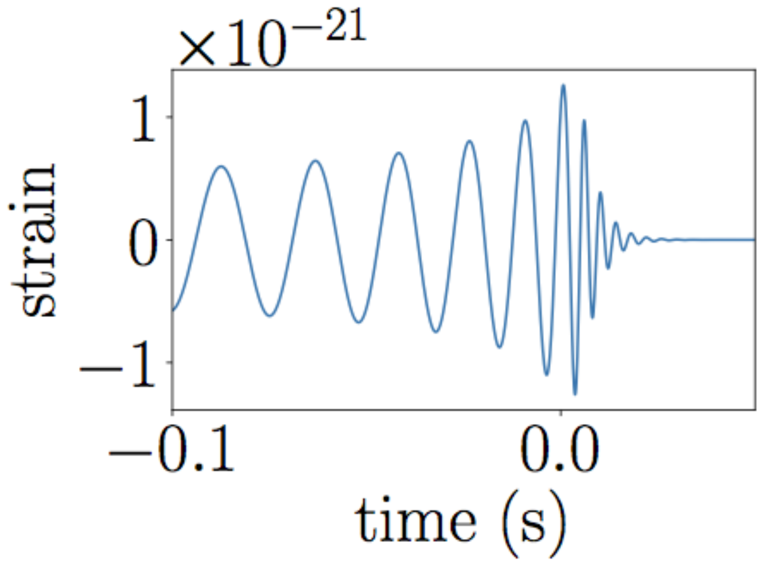
Software Optimisation: New capabilities & New responsibilities



So ... the benefit of performance isn't really about doing things faster, it's about opening-up ***whole new capabilities***.

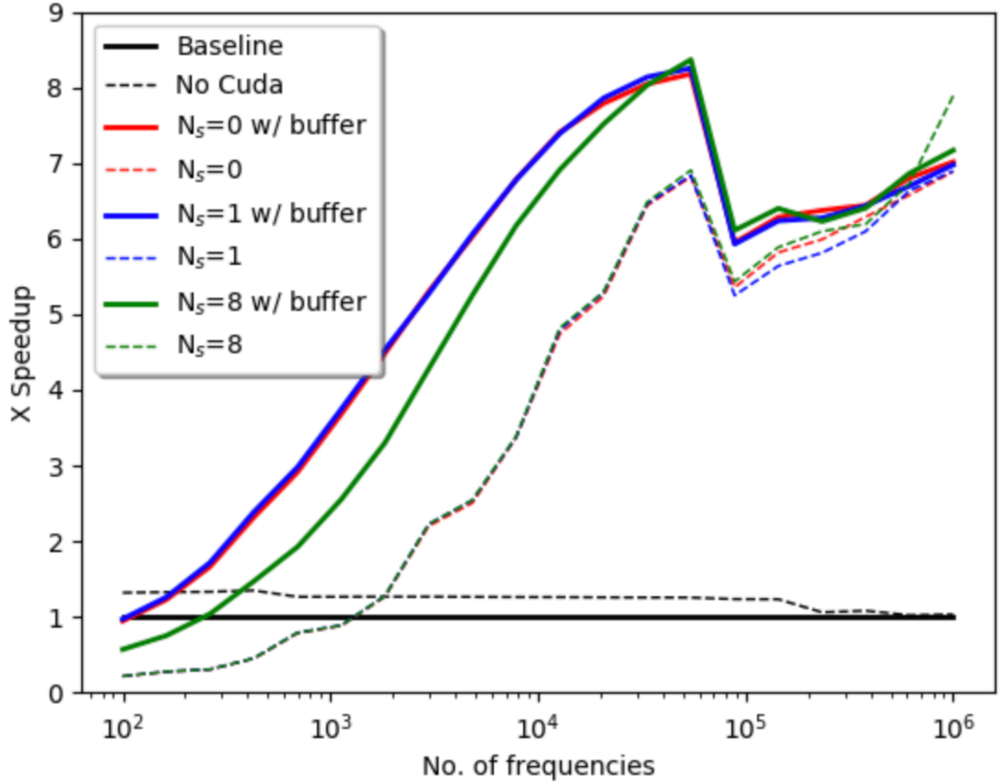
Optimisation: Example 1/2

Rory Smith (Monash) won an ADACS MAP allocation to optimise a "workhorse" GW waveform generator for coalescing binaries



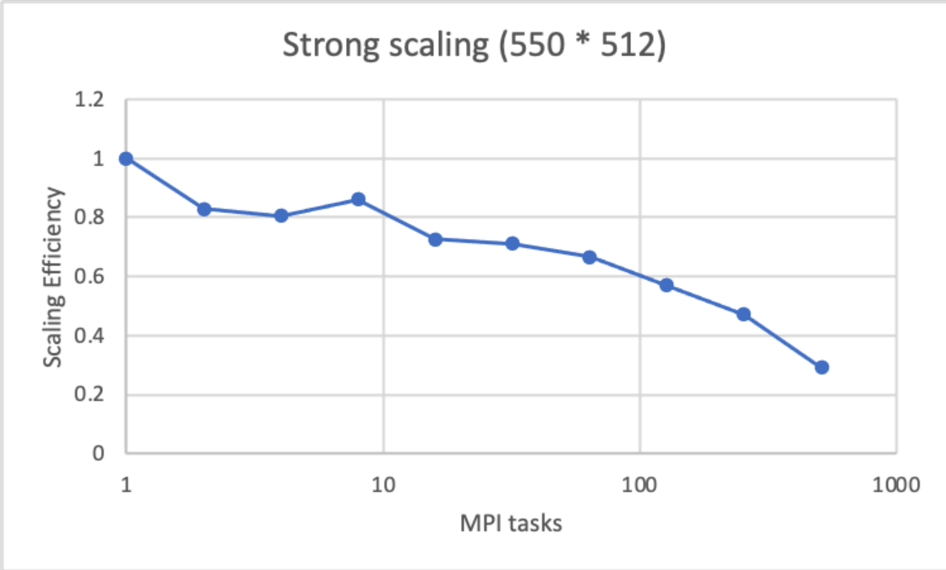
ADACS achieved an order of magnitude speed-up.

➡ Enables MC Bayesian inference on large samples of LIGO triggers



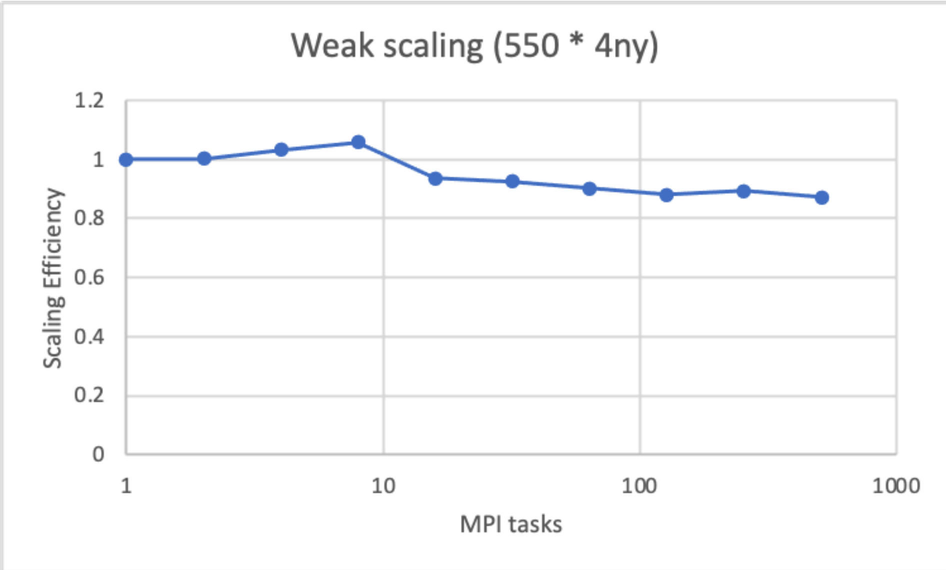
Optimisation: Example 2/2

Bernhard Mueller (Monash) won an ADACS MAP allocation to optimise a a 6-D Boltzmann solver developed to model neutrino transport in core-collapse supernovae



Excellent weak scaling, and reasonable strong scaling up to 100 MPI tasks

➡ paves the way for the first multi-dimensional (2D) supernova simulations with Boltzmann neutrino transport



Optimisation: New responsibilities

ADACS
ASTRONOMY DATA AND COMPUTING SERVICES

- For smaller/repetitive/regular execution, compute savings from optimisation can add up
  - ➡ drives a *reduction* in demand
- But for some cases - particularly large-scale execution:
  - ➡ ... we tend to just run larger problems
  - ➡ ... it allows us to scale to larger infrastructure
- Both of which just sustain or even drive an *increase* in demand

Should we/how can we organise as a community to manage this?