JuMP-dev 2025

JuMP on demand: Creating your own compute cluster for solving optimisation problems

James Foster

- Abstract
  - I will give a brief overview of the why, what and how of solving JuMP models via an on-demand cloud-based compute cluster. Inspired by the NEOS Server for solving optimisation models as a service, we also make use of the HTCondor software system through the Google cluster toolkit, but with resource limits that can be specified by the user based on their needs and budget.

- About me:
  - Research scientist in Energy with Australia's national science agency, the Commonwealth Scientific and Industrial Research Organisation (CSIRO).
  - Research interests include energy system models, modelling languages and data, power flow formulations and convex relaxations, linear algebra, quadratically constrained and mixed-integer optimisation.

james.foster@csiro.au  @jd-foster

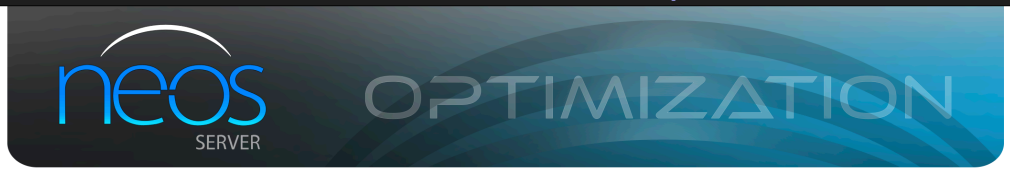
CSIRO

Background

- This talk is about:
  - challenges and techniques for scaling workflows
  - guiding those working through the same process
  - discussing the potential of a flexible JuMP-focused modelling workflow.

NEOS!

https://neos-server.org/neos/



NEOS Server: State-of-the-Art Solvers for Numerical Optimization

The **NEOS Server** is a free internet-based service for solving numerical optimization problems. Hosted by the Wisconsin Institute for Discovery at the University of Wisconsin in Madison, the NEOS Server provides access to more than 60 state-of-the-art solvers in more than a dozen optimization categories. Solvers hosted by the University of Wisconsin in Madison run on distributed high-performance machines enabled by the HTCondor software; remote solvers run on machines at Arizona State University, the University of Klagenfurt in Austria, and the University of Minho in Portugal.

- NEOS has been a long-standing and useful service to the optimisation community.
  - Web API/web submission front-end site
    - Other 3rd-party interfaces
    - see https://github.com/jump-dev/NEOSServer.jl !
  - **HTC**ondor: for managing and distributing the 'job queue' (problems to solve)
    - "**HTC**" stands for High-Throughput Computing - more later
  - **Solvers** (> 60*)
  - Built and hosted on different servers at U. Wisconsin-Madison, Arizona State, U. Klagenfurt & U. Minho.
    - **including commercial solvers that may otherwise be hard to access or obtain.*

CSIRO

Limitations of NEOS

- 10-15 jobs at a time
- Not all input formats are available to every solver
- Best case available memory: 3 GB of RAM
- Maximum run time: 8 hours
- Maximum of 4 threads per job
- Maximum output produced by solver: 100MB
  - "For security reasons, jobs are not allowed to read additional input files after they have been received by NEOS."
  - "In general, jobs are not allowed to write additional output files."

All these are well-founded design choices for the service.

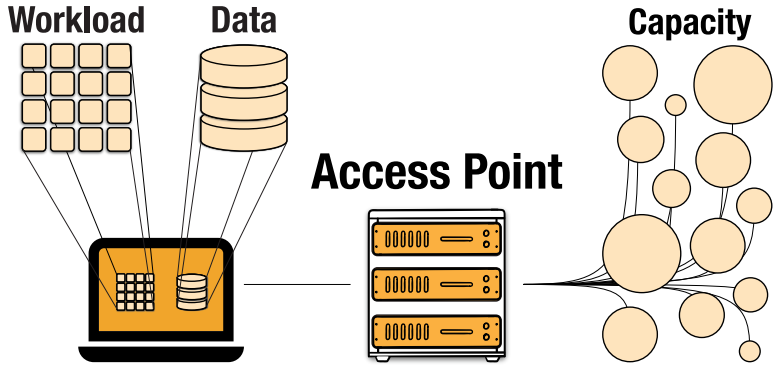
But what if we want a little more?

CSIRO

High-Throughput Computing (HTC) and HTCondor

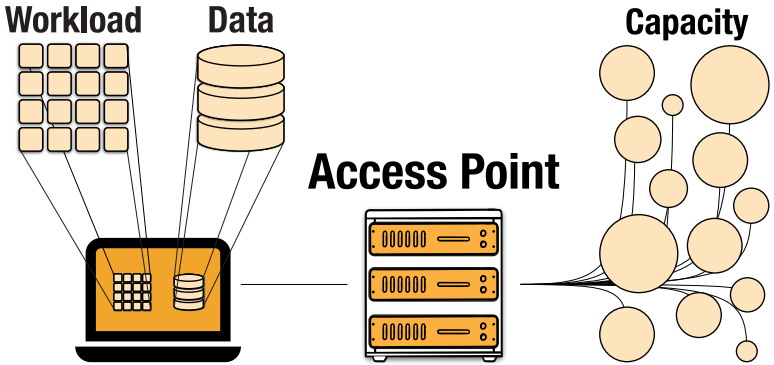
- HTC ≠ HPC
- Designed to manage effectively up to hundreds of thousands of jobs (rather than a small number of large compute task).
- The HTCondor 'Formula':
  - "describe your computational tasks as a series of independent, asynchronous “jobs.” [an atomic unit of work]
  - " access computational resources by submitting (or “placing”) job descriptions at an HTCondor “access point” (AP), also known as a “submit node.” "
  - "HTCondor locates an appropriate machine for each job, packages up the job and ships it off to ... execution points (EP).

CSIRO



https://osg-htc.org/services/access-point

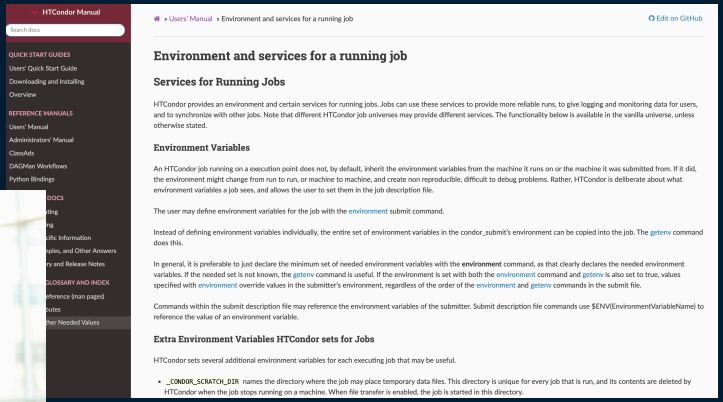
```
executable              = /bin/sleep
arguments               = 4
log                     = simple.log
output                  = simple.out
error                   = simple.error
request_memory          = 1GB
request_disk            = 1GB
request_cpus            = 4
queue
```




```
executable              = compare_states
arguments               = wi.dat us.dat wi.dat.out
should_transfer_files   = YES
transfer_input_files    = us.dat, wi.dat
when_to_transfer_output = ON_EXIT
log                     = job.log
output                  = job.out
error                   = job.err
request_cpus            = 1
request_memory          = 20MB
request_disk            = 20MB
queue 5
```

'Features' of HTCondor

- Great community supported by the Center for High Throughput Computing (CHTC) at UW-Madison.
- Training materials and high-quality documentation
  - Throughput Community Week
  - OSG Summer School



https://htcondor.readthedocs.io/en/latest/users-manual/env-of-job.html

https://www.chtc.cs.wisc.edu/events/2025/12/10/Bioinformatics-Cafe

- Consistency with researcher experience through the OSG Consortium and use of the Open Science Pool

One Researcher's Leap into Throughput Computing: Bringing Machine Learning to Dairy Farm Management

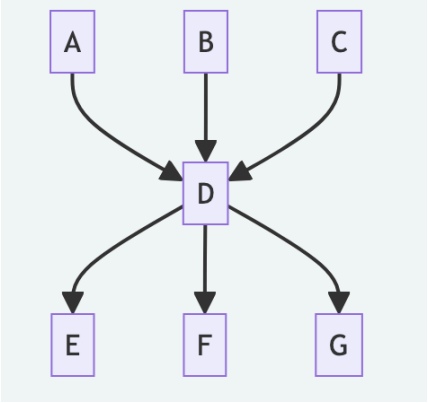
Sophie Dorros Nov 17, 2025

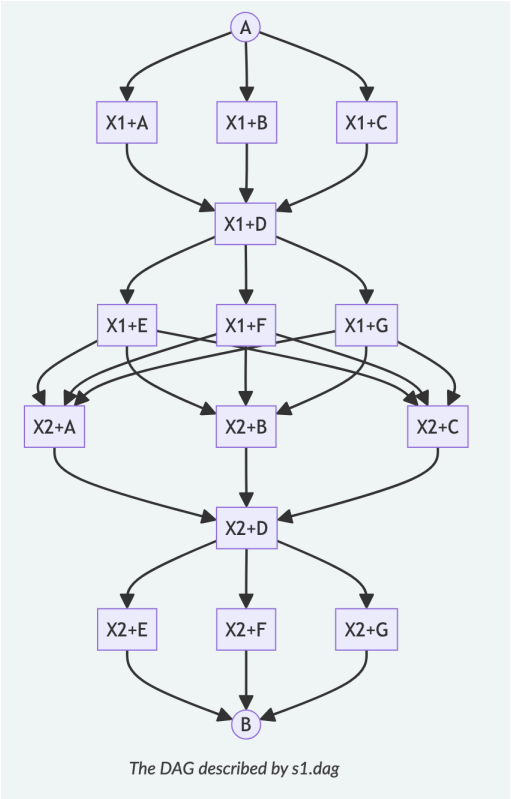
Ariana Negreiro, Ph.D, in the UW-Madison Digital Livestock Lab, discusses her work using images of cows to develop a machine learning application to monitor their health.

Pictured: Ariana Negreiro credits: Ariana Negreiro

The goals for Ariana Negreiro's work were clear: healthier dairy cows with an improved quality of life. Working under the direction of her lab director Dr. Joao Dorea, Negreiro's research utilizes machine learning and deep learning models to develop tools for dairy cattle management. She used an on-site processing system to produce top down, infrared and depth images of the cows, which are then processed and used to monitor the heifers' health and milk production potential. One strong advantage of the computer vision approach is that it eliminates the need to manually weigh and monitor cattle, which is

DAGMan Workflows



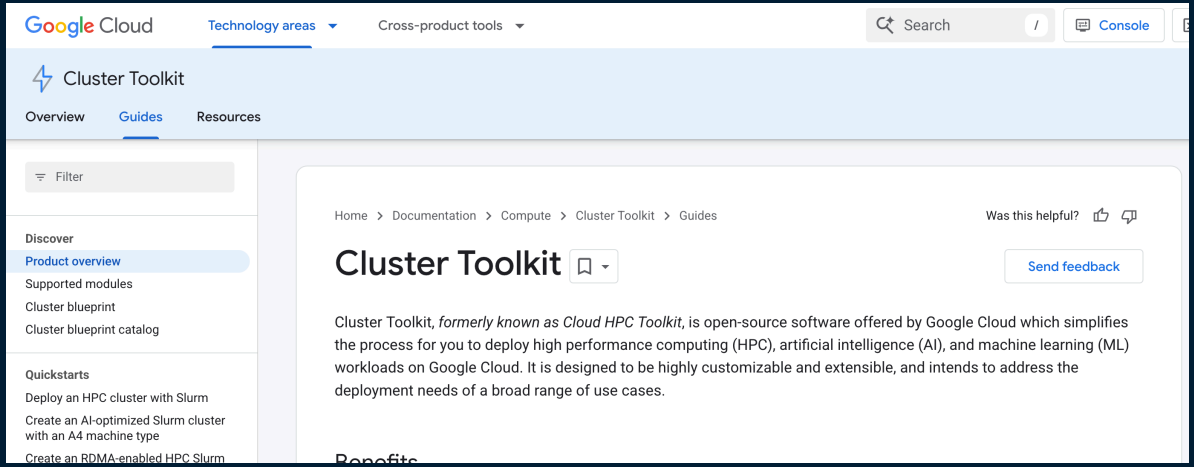


The DAG described by s1.dag

https://htcondor.readthedocs.io/en/latest/automated-workflows/dagman-using-other-dags.html#dag-splicing

CSIRO

Running your own cluster



Cluster Toolkit HTCondor Tutorial

Cluster Toolkit is an open-source software offered by Google Cloud which makes it easy for customers to deploy HPC environments on Google Cloud.

This tutorial will walk you through deploying a simple HTCondor pool on Google Cloud using the Cluster Toolkit.

Select a Project

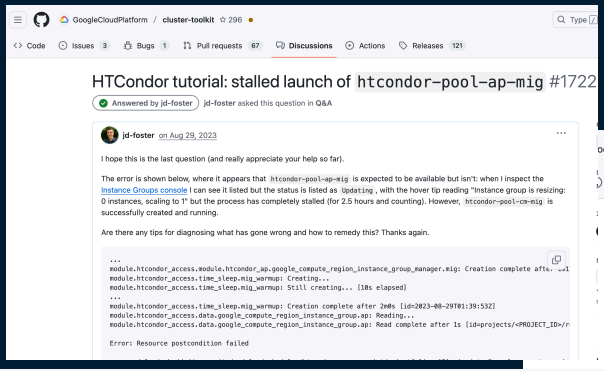
Select a project in which to deploy an HPC cluster on Google.

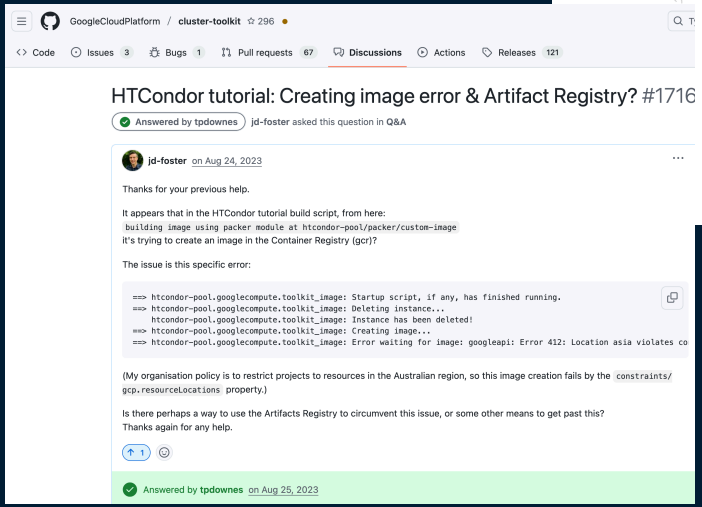
Once you have selected a project, click START.

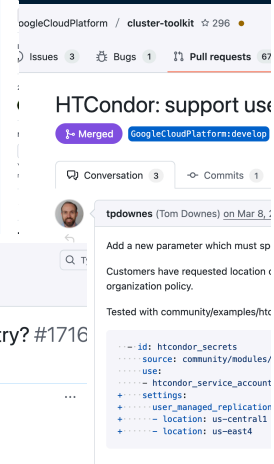
Add Credits to the Project

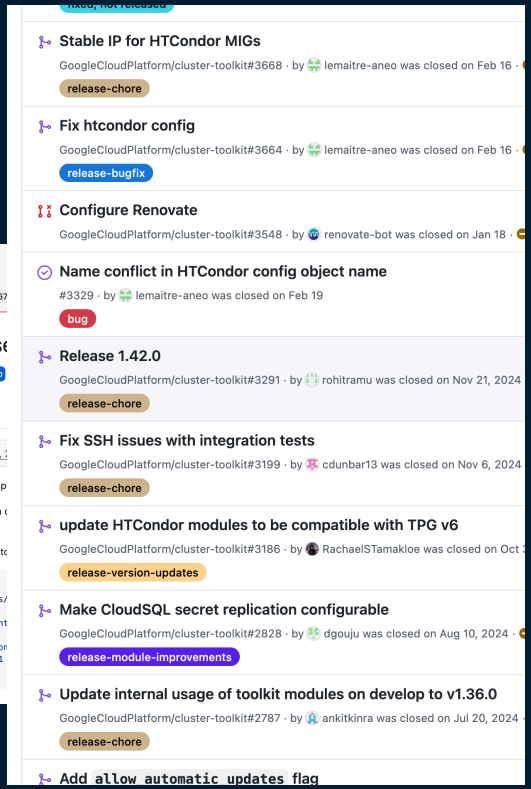
Talk with your tutorial leader to see if Google Cloud credits are available.

Batteries included*









CSIRO

Potential applications

- Programmatically running different variations on the same problem
- Sub-problem solving on medium-sized instances
- Testing solvers
  - benchmarking updates and features
  - exploring problem-specific settings and customisations
    - linear algebra libraries
    - new features in dependencies

CSIRO

Final thoughts

Condor calls for aid!

Final thoughts

- I believe this is an accessible route for general users to access additional computing resources.
  - This is particularly something that could assist general JuMP modellers.

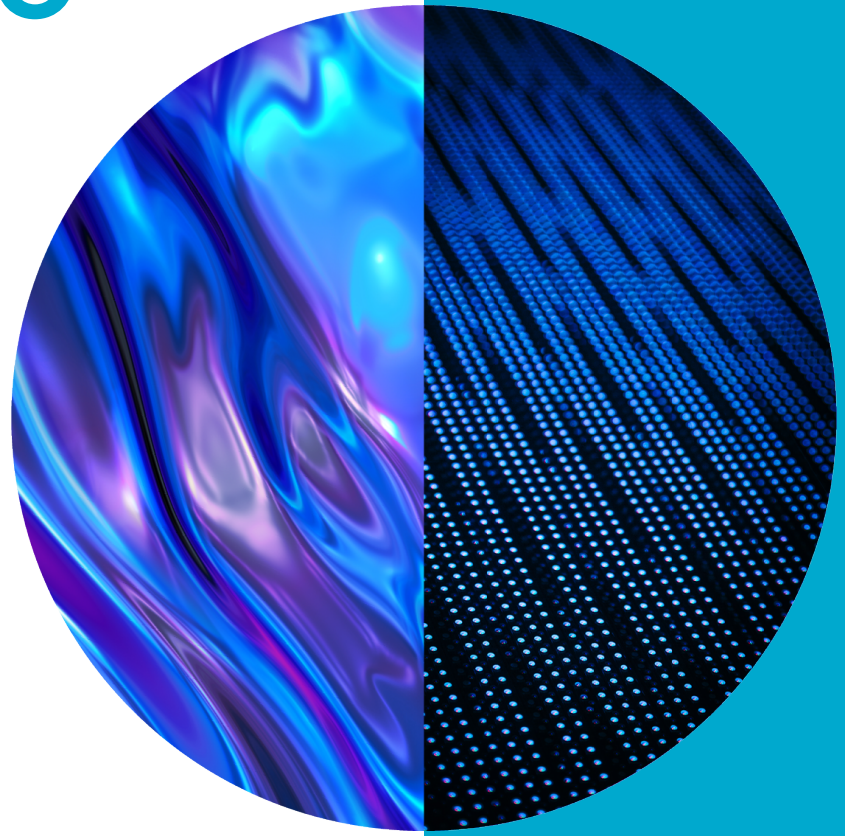
- I hope to have interested you (enough) to try this, and please reach out to chat.

Thank you.

@jd-foster

CSIRO

CSIRO



Australia's National Science Organisation