Green job scheduling

Sadie Bartholomew, NCAS-CMS meeting 26/05/23

Note: the first two slides are new and by SB to put the topic into context; all subsequent slides were by the 'Climate-aware Research Computing' group of the SSI CW23 Hack Day (03/05/23) and made during the event.

Idea towards green [≈ energy-efficient, low(er)-carbon] computing

"A **job scheduler** is a computer application for controlling unattended background program execution of jobs." [Source: Wikipedia]

Examples in increasing order of sophistication and capability:

- 'at' command
- cron job
- Batch submission e.g. SLURM
- Workflow manager e.g. Cylc

"Carbon intensity is a measure of how clean our electricity is. It refers to how many grams of carbon dioxide (CO₂) are released to produce a kilowatt hour (kWh) of electricity" [Source National Grid website*]

The CO₂ emissions from some task requiring electricity depend on the time the task is done! Reason: renewable (low-carbon) sources aren't available in a steady manner, and demand on an electricity grid varies.

^{*}https://www.nationalgrid.com/stories/energy-explained/what-is-carbon-intensity

Climate Aware Research Computing: cats (climate-aware task scheduler) and towards an 'Environmental Impact of Open Research' chapter for The Turing Way handbook

CW23 Hack Day Presentation

Team members (10): Colin Sauze, Sadie Bartholomew, Andrew Walker, Loïc Lannelongue, Thibault Lestang, Tony Greenberg, Lincoln Colling, Adam Ward, Abhishek Dasgupta and Carlos Martinez

Image credits: infographic designed by Prof. Ed Hawkins, see showyourstripes.info (indicates global average temperatures from 1850-2022 and the rising trend in these)

Motivation behind our project

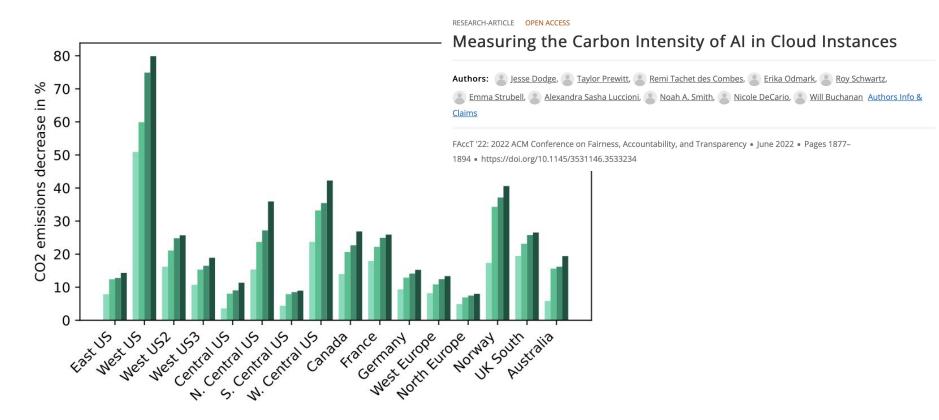


Image credits: from IT Crowd (Channel 4), linked from https://i.imgflip.com/208mpa.jpg

Many of us need to do computationally or data intensive science ...

... but can we do this and not set the world on fire?

Motivation behind our project



(a) Flexible Start optimization for Dense 201.

Climate Aware Research Computing: two project strands *⇒*



2. Towards an 'Environmental Impact of Open Research' chapter for The Turing Way handbook



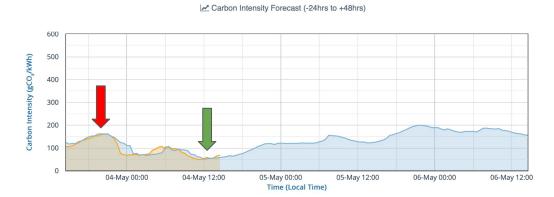
(One team as a final, fifth, sub-team)

openclipart.org/detail/183412/eco-green-solidarity-icon

cats: the climate-aware task scheduler

1. Users submit a runtime and postcode

2. CATS figures out the best time to start the job



Carbon Intensity API



TADAA, you've reduced your carbon footprint by **70%**

cats: a community implementation of the climate-aware task scheduler

- Created under a new GitHub organisation 'GreenScheduler' and repository 'GreenScheduler/cats'
- Developed as four semi-independent components targeting UNIX 'at' command (as a basic scheduler)
- Made use of CI for linting and running tests before merging
- Issues and development at https://github.com/GreenScheduler/cats
- Intended as proof of concept for users who do 'small scale' computing (e.g. a few hours on a workstation overnight)



Demo - please hold your breath

```
$ python -m cats --duration 60 --loc RG1
{'timestamp': datetime.datetime(2023, 5, 16, 11, 30), 'carbon in
tensity': 82.0, 'est total carbon': 82.0}
Best job start time: 2023-05-16 11:30:00
202305161130
$ ls | at -t `python -m cats --duration 60 --loc RG1`
{'timestamp': datetime.datetime(2023, 5, 16, 11, 30), 'carbon in
tensity': 82.0, 'est total carbon': 82.0}
Best job start time: 2023-05-16 11:30:00
warning: commands will be executed using /bin/sh
job 28 at Tue May 16 11:30:00 2023
```

```
##### Your carbon footprint on the CW23 platform #####
```

By running at the suggested time: 4 kgCO2e

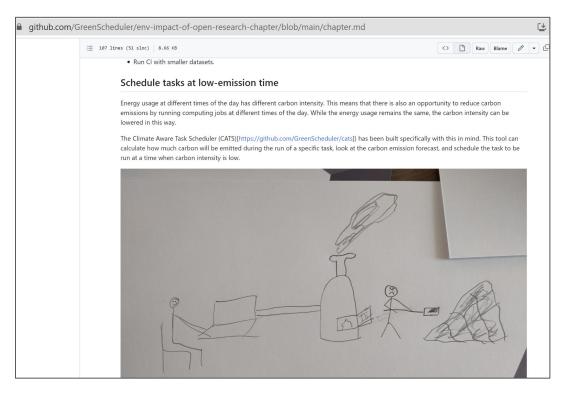
...vs running now: 12 kgCO2e (- 8 kgCO2e)

Towards an 'Environmental Impact of Open Research' The Turing Way chapter

- Original issue:
 https://github.com/alan-turing-institute/the-turing-wa
 y/issues/2803
- As with the Turing Way (TTW): "The process documents and data are made available under a CC-BY license." hence use CC-BY for an intermediate repository, to ease possible transition. Committed paragraphs to:
 https://github.com/GreenScheduler/env-impact-of-open-research-chapter
- Drafts of structure of chapter, text for the sections, plus sketched-out ideas for illustrations. We wanted to ensure there was plenty of non-text content!



Towards an 'Environmental Impact of Open Research' The Turing Way chapter





https://greenscheduler.github.io/env-impact-of-open-research-chapter

Team work

- Broke into small teams (2-3 people) for components
- 10 min online round table at the top of each hour
 policy decisions and discussion around integration
- Early discussion around scope, tooling, longer term plan, licence and code of conduct
- Lessons learned around hybrid hack day round table meetings need more formal 'chair', single audio stream from in-person participants etc...
- Cakes!



Future work, potential and ideas

- For the handbook, the future work is to further develop the content until it becomes a coherent chapter, eventually submitting this to TTW via their well-documented and established process for adding new content.
 - Pick this up for the next book dash
 - Key to building understanding and engagement
- <u>For cats</u>, we aimed to create a minimally viable product (so-called MVP), but we acknowledged many possible extensions along the way:
 - Extend it to make it more configurable, so that other schedulers can be wrapped and utilised rather than the basic 'at' command.
 - Support further alternative algorithms for estimating the 'greenest' time to run a job
 - Make available via PyPI