

DiRAC Sustainability

HPC User Forum fall 23

www.dirac.ac.uk

sburbidge@btopenworld.com

DiRAC centres

- 4 centres currently
- Hosted by partner Universities : Durham, Cambridge, Leicester, Edinburgh
- Each centre is autonomous, often hosting local systems and other shared resources as well as DiRAC ones. Different cooling and power approaches as well as age.

Centre M&E details

- All sites annually averaged PUE below 1.2
- New deployments since 2020 use direct liquid cooling
- Leicester - free cooling and DX
- Durham - Direct Liquid Cooling, free cooling, solar power for 5-10% in summer
- Edinburgh - Direct Liquid Cooling, free cooling, chillers when needed in summer. In Scotland away from city heat. 100% renewable energy supply
- Cambridge – water cooled, free cooling

But its not just hardware

- All the centres are reasonably efficient and the gains to be made are limited (although re-using the heat would be a major gain!)
- We are very focused on the cost of energy – higher than in the USA. And recently (due to changes in the energy market) it's become even more expensive. Around 5 times more expensive than in the USA.
- Also the UK government is aiming for "Net Zero" (carbon neutral) economy.
- So the pressure is on ...

Look at software optimization

- Recent well used genomics code, was ported to an Arm based system by a student intern. They found main inner loop wasn't vectorized – added in compiler directives gave a 50% improvement in speed (on Arm and X86).
- How many other codes out there have similar issues ? Most people assume that if it compiles without error then it must be optimized!
- Re-platform codes (eg to GPUs) as appropriate

Back to hardware ...

- Tursa system at Edinburgh
- System was procured for specific Lattice QCD applications such as *Grid*
- A lot of effort and preparation with the vendor concentrated on getting a system that would run this code faster and more efficiently than other systems
- In fact it was clear that a GPU would run this code much better than a CPU. Code was rewritten to run on GPUs, needing several person years of software engineering
- The result is that the new system delivers 5X performance using 50% less electricity

Look at hardware tweaking

- Tursa at Edinburgh Nvidia A100 GPU system study by Prof Antonin Portelli
- Clocking down the GPUs by 10% reduces power consumption by 16-24% .
- Have set this as default – users can choose to use full speed on a per job basis if they need to.
- <https://edinburgh-innovations.ed.ac.uk/news/supercomputer-sustainability-a-balance-between-performance-and-energy-efficiency>

Tools

- DiRAC RSE projects to measure energy of runs directly, so its immediately available to users.
- Trying to develop reporting tool that works on most/all systems (RSEs Tom Meltzer, Matt Walker at Cambridge)
- Alastair Basden at Durham sends monthly CO2 reports from the COSMA system to users and project PIs to improve awareness
- Aim to give users the data to make informed choices

Continuing research activities

- As well as our work at DiRAC, there is a researcher at EPCC, Prof Michele Weiland, who has a specific interest in the field :
<https://www.epcc.ed.ac.uk/about-us/our-team/prof-michele-weiland>