DiRAC and IRIS

IRIS Collaboration Meeting

December 2023

Alastair Basden

DiRAC

- STFC-funded HPC Facility for theory community
 - Cosmology, particle physics, nuclear physics, solar physics etc

DiRAC Federation projects

- ~£1.9m funding for a 6-12 month spend
 - UKRI DRI preparation
 - Advanced level training materials
 - Allocation intervention strategies
 - Solar panel deployment
 - Multi-site data federation tools
 - Open-source system deployment tool

UKRI DRI preparation

- On-boarding activities
 - 6 non-STFC communities approached
 - 2-3 2-day workshops to explore future DRI needs
 - Time allocations given on appropriate systems
 - e.g. high memory workloads, GPU workloads
 - CompBioMed, HECBioSim, MRCGlasgow, UKAEA, Materials Science/PAX
- In preparation for design of future cross-community large systems

Advanced training materials

- Update of DiRAC HPC training course
 - HPC skills training
- Machine learning courses
- GPU courses
- Hackathons
- Innovation projects

Allocation intervention strategies

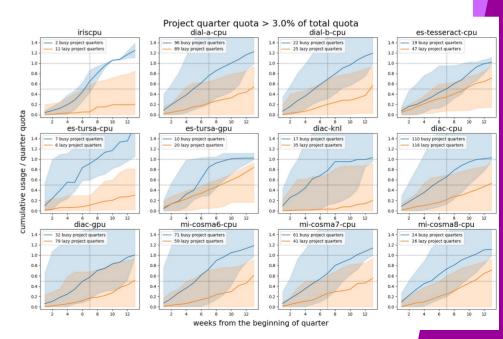
In-depth study of historical DiRAC usage

- Can a dynamic re-allocation process improve system

usage?

Median project 2-3 users

 Strong evidence to support a mid-quarter re-evaluation



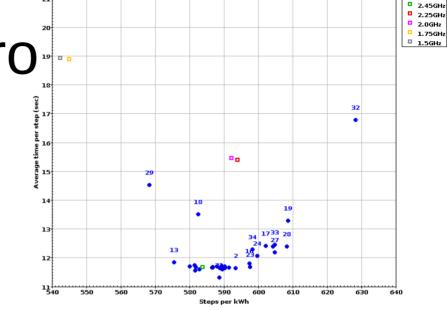
Solar panels

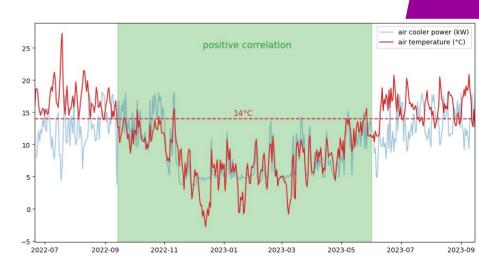
- ~£1m deployment at Durham
 - To demonstrate feasibility of coupling DRI with net-zero
 - Lessons learned: 6 month timescales are very challenging longer-term funding required
 - Requires a lot of good will from Estates
 - Good to have a pre-prepared plan



Net-zero.

- Several studies related to net-zero
 - Internal and external to DiRAC
 - GPU frequency
 - Green500
 - CPU BIOS settings
 - NERC scoping project (CEDA)
 - Carbon-aware scheduling workshop
 - User CO2 feedback and awareness
 - Leading edge cooling facilities
 - Historical power-cooling analysis
 - Code power-draw studies
 - Idle node power-off





Multi-site data federation

- Data federation between Edinburgh and Durham
 - RUCIO deemed not appropriate after some study
- Atempo ongoing pilot between Durham and Leicester
 - Moving data from Leicester to tape storage at Durham
- StorJ
 - Distributed cloud storage on-site cloud storage

StorJ distributed storage cloud

- Distributed cloud storage
 - Data is "sharded" and copied to ~100 locations
 - · Opt-in servers anywhere in the world
 - ~30+100 erasure coding
 - · Geo-fencing is an option
 - On-site system being set up (Durham, Edinburgh)
 - Data can be copied to "buckets" on storage hosted at these sites
 - Less sharding, ~8+3 erasure coding
 - Workflow:
 - Set up a bucket (typically via web interface)
 - Copy data from local (DiRAC) file system onto StorJ system
 - e.g. using rclone or web interface
 - Make this available to collaborators, or copy to another site
 - Note, not mounted as a parallel file system on the HPC service
 - Meta data tagging can be added and can be scripted

Open-source deployment: Ubiquity

- Ubiquity open source HPC deployment and management tool
 - QAssociates
- Based on open-stack
- Promises dev-ops style management
- Used for CIUK student cluster competition
 - Test cluster spun up at Durham

Conclusions

- DiRAC highly aligned with IRIS
 - Federation
 - User support
 - User training
 - DRI provision