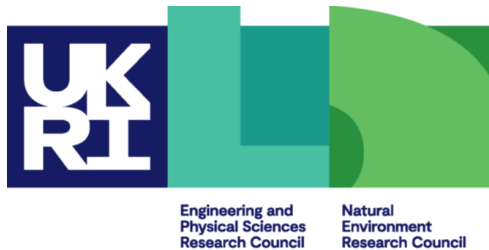


ARCHER2 for Data Scientists

Introduction



Reusing this material



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

This means you are free to copy and redistribute the material and adapt and build on the material under the following terms: You must give appropriate credit, provide a link to the license and indicate if changes were made. If you adapt or build on the material you must distribute your work under the same license as the original.

Acknowledge EPCC as follows: “© EPCC, The University of Edinburgh, www.epcc.ed.ac.uk”

Note that this presentation contains images owned by others. Please seek their permission before reusing these images.





- UK National Supercomputer Service
 - managed by UKRI/EPSC
 - Housed, operated and supported by EPCC
 - hardware supplied by Cray
- Training provided by the ARCHER2 Computational Science and Engineering (CSE) support team
 - 60 days per year at various locations
 - free to all academics



Located at EPCC's Advanced Computing Facility (ACF)



Key ARCHER2 Resources

- Upcoming courses
 - <http://www.archer2.ac.uk/training/>
- Material from past courses
 - <https://www.archer2.ac.uk/training/materials/>
- Virtual tutorials (online)
 - <http://www.archer2.ac.uk/training/>
- Documentation
 - <http://www.archer2.ac.uk/documentation/>

Who am I?

Adrian Jackson a.jackson@epcc.ed.ac.uk

- Teach at EPCC on HPC and program optimisation:
 - MSc, PhDs
 - PRACE Advanced Training Centre
 - ARCHER training programme
 - commercial training
 - ...
- Also do HPC research
 - Porting and optimizing community codes
 - new parallel programming models, accelerators, performance, memory hardware, ..



Other Resources

- Please fill in the feedback form!
 - you will be sent a link at the end of the course
- General enquiries about ARCHER2 go to the helpdesk
 - support@archer2.ac.uk
- EPCC runs one-year taught postgraduate masters courses
 - ***MSc in HPC*** and ***MSc in HPC with Data Science***
 - awarded by the University of Edinburgh since 2001
 - scholarships available
 - <http://www.epcc.ed.ac.uk/msc/>



Access during course

- Personal accounts for duration of course
 - will allow machine access for up to a month afterwards
- Accounts will be closed two weeks after access ends
 - all files etc. will be deleted
 - take copies of all your work beforehand!
- Course materials (slides, exercises etc) will continue to be available from ARCHER2 website
 - archived on ARCHER2 training pages for future reference

Code of Conduct

<https://www.archer2.ac.uk/about/policies/code-of-conduct.html>

- We expect all course trainers and attendees to:
 - Use welcoming and inclusive language
 - Be respectful of different viewpoints and experiences
 - Gracefully accept constructive criticism
 - Focus on what is best for the community
 - Show courtesy and respect towards other community members
- See web page for full details and incident reporting form



Funding calls

- Embedded CSE support
 - Through a series of regular calls, Embedded CSE (eCSE) support provides funding to the ARCHER2 user community to develop software in a sustainable manner to improve research on the ARCHER2 service. The funding allows the employment of a Research Software Engineer (RSE) to carry out software development of ARCHER2 software.
- See <https://www.archer2.ac.uk/ecse/> for details



ARCHER2 hardware / software setup

- 5,860 nodes each with 128 AMD CPU-cores
 - made up of 2x64-Core AMD multicore processors
 - more than 750,000 CPU-cores!
- Batch access via SLURM: sbatch, squeue, ...
- Two file systems – you have two directories:
 - /home/project/project/username
 - /work/project/project/username
- You must run all parallel jobs from /work/
 - I recommend you “cd” straight there every time you log in
- If required, reserved queues each day for fast turnaround



Timetable

- 10:00 – Registration and check connections
- 10:10 – L01: Introduction to ARCHER2
- 11:00 – Exercise: ARCHER2 filesystems and placement
- 11:15 – Break
- 11:30 – L02: Containers and installing software
- 12:20 – Exercise: Containers
- 13:00 – Lunch
- 14:00 – L03: Parallel Python and Dask
- 15:00 – Exercise: Dask on ARCHER2
- 15:30 – Break
- 16:00 – L04: Parallel R on ARCHER2
- 17:00 – Exercise: Parallel R
- 17:30

Course aims

- ARCHER2 for Data Scientists
 - Teach you what you need to know to:
 - Understand what ARCHER2 is suitable for
 - Run your workflows on ARCHER2
 - Install new software on ARCHER2
 - Use the hardware as efficiently as possible
 - Not aiming to teach you:
 - How to undertake data science
 - The one tool you need use
- Focussing on:
 - Containers
 - Python
 - R



I hope you enjoy the course



- ... and *please ask questions!*