The Social Life of Big Data - Pawsey resources

Luke Edwards
Pawsey Supercomputing Centre, Western Australia, luke.edwards@pawsey.org.au

Follow this and additional works at: https://ro.ecu.edu.au/slbd

Part of the Computer Engineering Commons, and the Social and Behavioral Sciences Commons

Recommended Citation

This Presentation is posted at Research Online.
https://ro.ecu.edu.au/slbd/3
The Social Life of Big Data
- Pawsey resources

Luke Edwards
(luke.edwards@pawsey.org.au)

2nd June 2015
Overview

• What is Pawsey? What services are there?

• Data Management tips

• How do I get access?
  • Level of expertise required.
  • Help and support that are available.
The Pawsey Supercomputing Centre

The Pawsey Supercomputing Centre is an unincorporated joint venture between CSIRO, Curtin University, Edith Cowan University, Murdoch University and the University of Western Australia and is supported by the Western Australian Government.

3 teams
- Data
- Supercomputing
- Visualisation

5 facilities
- CSIRO
- Curtin
- ECU
- Murdoch
- UWA

I’m based here
Pawsey services

1. Supercomputing
2. Data Storage
3. Visualisation

• NeCTAR Research Cloud (Pawsey Node)
• Hadoop for testing purposes (Data analytics)
1. Supercomputing

- **Zythos** - a large shared-memory machine
- **Galaxy** - operational processing for radio-astronomy
- **NeCTAR Research Cloud** - on-demand, research cloud computing
- **Magnus** - petascale computing
Data Storage
Storage

Tape libraries
2. Storage

Pawsey Data Storage (https://data.ivec.org)

- Web interface (navigate, search, edit, upload, download)
- Script upload
- Application process (>5TB)
- Designed for ‘active’ research, not ‘primary’ copy
3. Visualisation

The process of applying advanced algorithms and computer graphics to data allows researchers to better understand their raw data and translate it into meaningful forms.

Allow researchers to
- learn something new,
- to work faster and identify issues
- to communicate with peers / wider audience

Hardware, software and human expertise:
Research Data Management tips
(35) There shall be standard measures of wine, ale, and corn (the London quarter), throughout the kingdom. There shall also be a standard width of dyed cloth, russet, and haberject, namely two ells within the selvedges. Weights are to be standardised similarly.

Source: http://www.bl.uk/magna-carta/articles/magna-carta-english-translation
Data Standards

Source: https://xkcd.com/927/
Data Licensing

‘Be like a plumber’

Source: Baden Appleyard, Director AusGOAL
Data Documentation

Source: http://xkcd.com/1459/
Data ‘Open-ness’

Be F.A.I.R.

Findable
Accessible
Interoperable
Reusable

Source: https://www.force11.org/group/fairgroup/fairprinciples
Skills

• Some knowledge of ‘Linux’ / command line
  • Linux is free and open-source software
  • Run on most supercomputers
  • Command-line interface (not GUI)
  • Remote access to Pawsey from your computer
Support

- Training (http://www.pawsey.org.au/training/)
- Access (http://www.pawsey.org.au/access-and-support/)
  - Supercomputing (Competitive Merit Allocation)
  - Data (Online application form)
  - Visualisation (email)
- Intern program (10 week project - undergraduates)
- Pawsey friends (http://www.pawsey.org.au/email-list-registration/)