Computational Resources: Hardware, Software & Training

Tom Nichols

Big assist from: Wilfred Kendall & David Firth

Thanks to ad hoc PhD IT induction team:
Camille Maumet, Seb Armstrong, Matt Moores, David Selby

0th Consideration: Platform

- Windows
  - Pro
    - It's sitting on your desk
    - Lots of software, often not free
  - Con
    - Greedy background/system processes
    - Limited/no ability to script programs
    - Interfaces, programs sometimes bloated

Microsoft® Windows®

0th Consideration: Platform

- Linux
  - Pro
    - Fast (e.g. compared to Windows on same hardware)
    - Lots of software, usually free
    - Build around command line, programing, scripting
  - Con
    - Software not always well documented
    - Can have steep learning curve, but many easy-to-use installations now available (e.g. Ubuntu)

Linux

0^{th} Consideration: Platform

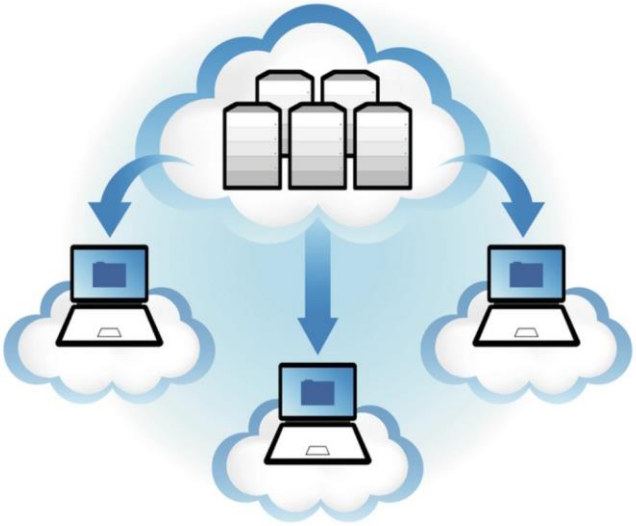
- Apple Mac
  - Pro
    - Beautiful interface, usually very intuitive
    - MacOS built on Linux, hence command line tools, programming, scripting, easy at hand
    - Coolness
  - Con
    - Hardware more expensive and *only* from Apple

0th Consideration: Platform

- Cloud/Web Solutions
  - I.e. browser based tools like Google Docs, Office Online
  - Pro
    - As long as you have internet, you have everything
    - Always backed up…
      No data loss due to lost/stolen laptop/drive
  - Con
    - Security issues… give someone your password, they have your ‘cloud’ life
    - Offline access
      - Some solutions, but never as elegant as just working on your own machine

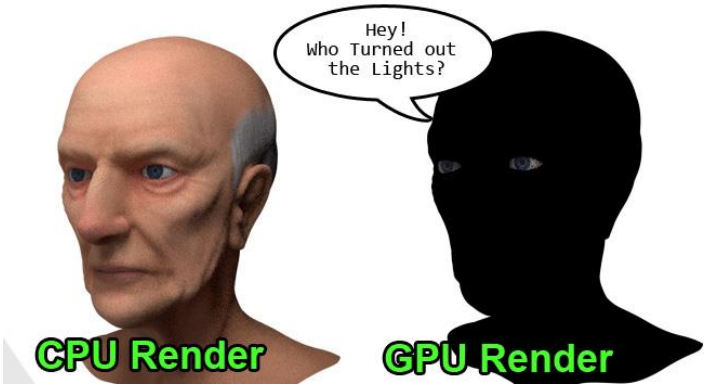
0th Consideration: Exotic Options

- Cloud Computing
  - Buy a ‘slice’ of a compute farm
  - Amazon Web Services
  - Only pay for AWS ‘Instance’ when it is used & computing
  - Pros
    - Linux hosts always updated, not your problem
    - Always accessible, known uptime
  - Cons
    - Metered usage – You pay for *exactly* what you use
      - Have a useless MCMC that runs for days? You pay!



0th Consideration: Exotic Options

- Graphical Processing Units (GPUs)
  - Formerly just for video games, now for scientific computing
  - Requires specialized code, thinking deeply about how to best parallelize your problem



  - Never stand-alone, usually installed into Linux box
- Field Programmable Gate Arrays (FPGA)
  - Custom-designed chips to accelerate computations
  - Bleeding edge

1st – 3rd Consideration: Backup, Backup, Backup

- Office Desktop
  - Everything on your H: drive is backed up by university
  - Everything local is not!
    - Your responsibility to back it up!
- Personal laptop
  - Buy a big external USB drive (e.g. 2TB for £50)
  - Use system's default back up 'history' tool
    - Mac: Time Machine, Windows: File History
  - Dropbox
    - Keep key files in synchronized folder or just drag a copy of key items (thesis.tex!) there now and again

Email

- University Solution
  - http://warwick.ac.uk/insite -> Email or http://webmail.warwick.ac.uk
- Can activate forwarding under settings [IT Help Doc]
  - They warn against this, as you then take responsibility for email loss, etc

Windows	**Linux**	**Mac**	**Web**
Outlook	Evolution (Ubuntu: free, open source)	Mail	Microsoft Office 365
	Kmail (Kubuntu: free, open source)	Microsoft Outlook	Gmail (free) Yahoo (free)
Thunderbird (free, open source) Mutt, pine (text-only, free, open source, for Geeks!)			

Calendar

- Pick one device/platform as ‘primary’, and make sure it synchronizes to all your devices
  - E.g. University’s Office 365, or Google Calendar or Apple Calendar
  - Sync all others to it
    - *** Not tested in general!***
      Only can vouch for: Google primary, sync’d to Mac & iPhone/Android

Windows	**Ubuntu**	**Mac**	**Web**
Outlook	Evolution (Ubuntu: free, open source)	Calendar	Google Calendar (free)
	Korganizer (Kubuntu: free, open source)	Microsoft Outlook	
Rainlendar (free) Mozilla Sunbird or Lightning (both free, open source)			

Computational Resources Available to You (1/2)

- Your desktop
  - Refreshed on regular basis, but some variation
  - If not fit for purpose, complain!
- Buster
  - Linux cluster
  - 12 high-performance Dell Linux workstations
    - 1 head node 'buster.stats.warwick.ac.uk'
    - Controls 11 other via SGE job control system
    - Share common file storage system
- Goldfinch
  - Linux host with GPU card

Session 3: Becoming a Power Buster User
Date: Tuesday 29 November, 4-5pm

Computational Resources Available to You (2/2)

- Centre for Statistical Computing (CSC)
  - Tinis
    - Newly cluster, 3500 Intel Haswell cores (c. 2015)
  - Minerva
    - Old cluster, 2500 Intel Westmere core (c. 2011) (To be decommissioned in early 2017!)
  - Apocrita
    - Cluster based at Queen Mary
  - Cluster of Workstations (COW)
    - Uses free cycles when desktop Linux system unused

CSC *also* covered in…
Session 3: Becoming a Power Buster User
Date: Tuesday 29 November, 4-5pm

Computational Courses in Other Departments

- MSc courses on computational methods are taught in various departments
  - Center for Statistical Computing (CSC)
  - Computer Science (CS)
  - Mathematics
  - Physics
  - … **In all cases**, need to inquire about ability of a Stats student to join… formally or informally.
- See also
  - IT Services
    - http://go.warwick.ac.uk/itstraining
  - Oxford's CUDA (GPU) school
    - https://people.maths.ox.ac.uk/gilesm/cuda/

Educational Resources: CSC Modules

- CY900: Foundations of Scientific Computing
  - Linux and programming basics (including Makefiles), and basics of scientific computing, incl. Linear Alg & Solving PDEs; including Matlab, R, etc.
- CY901: High Performance Scientific Computing
  - Optimizing serial and parallelized code; shared vs distributed memory programming; message passing; GPU programming
- CY903: Practical Algorithms and Data Structures
  - Algorithms for serial and parallelized coding; data structures for scientific computing; algorithm re-use; gauging algorithmic complexity

Educational Resources: CS Modules

- Foundations of Data Analytics
  - Statistics, Databases, Regression, Classification, data structures
- Data Mining
  - Models/algorithms & their accuracy

Educational Resources: MathSys Modules

- MA934 - Numerical Methods
  - Derivation, interpolation and extrapolation, integration, extremal finding, ODE; sorting, binary trees, search; MCMC
- CO923-18 - Computational Methods for Complex Systems
  - Systems of linear equations, eigenvalue problems, matrix problems; MCMC, the Ising model; generation of constrained complex networks; useful algorithms and data structures; networked dynamical systems

Educational Resources: Physics Modules

- High Performance Computing in Physics
  - Optimizing serial and parallelized code; shared vs distributed memory programming; message passing; GPU programming

Questions

- Look for these slides and links on the web