HPC at Warwick and Beyond
Section 1 - Intro

Warwick RSE

17/10/2018

Introduction

What is HPC?

- High Performance Computing
- Custom bought (or built) computers for doing “hard work”
- “Hard work”
  - Anything that slows your desktop to a crawl for an hour or more
  - Things that run across multiple compute nodes, use custom hardware (graphics cards etc)

What's in the Box?

- Compute cluster
  - "Many" individual machines - **nodes**
    - These are basically good desktops
    - In different cases
  - Vital difference - **interconnects**
    - Nodes are networked together superfast
    - 32 GBit (real world)
    - Compare 1GBit "fast ethernet"

Many can mean anything, but at Warwick we have a few hundred
Good desktop here can also mean anything. At the moment usually means dual socket, 8-32 cores per chip, and a full load of RAM (4-12 GB/core). Machines for particular purposes can be quite different.
Main visual difference is that these aren't in normal cases, but mounted in a rack (https://en.wikipedia.org/wiki/19-inch_rack).

Ethernet can reach 5GBit on normal Cat-6 network cable if you have the kit, but normal computers, routers etc top about around a gigabit. Your home wifi caps out at about 36 megabit normally.

What these talks will cover

- Talk 1 (today): basics of what we have at Warwick, how to access it, and where to find help and support
- Talk 2 (in 2 weeks): using the clusters, submitting jobs, viewing and controlling your stuff
- Talk 3 (in 3 weeks): assorted bits of vital knowledge. Basics of scaling; how to understand errors; how to request packages etc; other facilities across the UK and internationally.

CSC, SCRTP, RSE, ITS, see?

- ITS - IT Services. University wide, ITS provides email, site builder pages and a managed Windows desktop. This is not us.
- CSC - Centre for Scientific Computing. Largely replaced by the SCRTP. Still might see pages refer to CSC but don't worry.
- SCRTP - Scientific Computing Research Technology Platform. This is us.
- RSE - Research Software Engineering. Part of SCRTP, support researchers in writing and running software.

In particular, the old CSC name crops up when accessing the cluster machines: you have to use {name}.csc.warwick.ac.uk
SCRTP website - warwick.ac.uk/scrtp
RSE webpage - warwick.ac.uk/rse

What can I use at Warwick?

- SCRTP Linux desktop. Quad-core (right now) managed desktop giving you hassle free access to compilers, packages etc
- Compute Clusters. Tinis (since 2015) and Orac (since 2017). Few thousand cores each, 4GB memory per core. See https://warwick.ac.uk/research/rtp/sc/hpc
- Other nodes - FAT (high memory), GPU, OpenPOWER

Desktop just migrated to Centos7. Offers variety of open-source stuff (gcc tools, mpi etc), Intel compiler and maths (imkl), Python 2 and 3 with the usual science libraries, some widely used packages e.g. COMSOL

Access to Systems - Theory

Who are we?

- SCRTP. Director David Quigley, 4 systems administrators, 2 RSEs and a couple of shared admin staff
- We support the compute clusters and managed Linux desktop
- Also manage access to HPC Midlands Plus (see Talk 3 for more)

What is there?

- Login Node - Godzilla. Provides remote desktop, file access etc
- Compute Clusters.
  - Tinis (since 2015)
  - Orac (since 2017)
- Dedicated Nodes - group owned nodes of the Cluster of Workstations (COW)
- Other nodes - FAT (high memory), OpenPOWER, GPU nodes

Dedicated nodes are SCRTP managed machines bought by and dedicated to a research group or individual. They’re accessed using custom queues on the COW. If your supervisor has one, ask them about access and make sure you’re using what you think you are!

Helpful Links

- SCRTP main page - https://warwick.ac.uk/scrtp
- Warwick RSE main page - https://warwick.ac.uk/rse
- Wiki
  - clusters https://wiki.csc.warwick.ac.uk/twiki/bin/view/Main/ClusterUserGuide
  - desktops https://wiki.csc.warwick.ac.uk/twiki/bin/view/Desktop2018/WebHome
- RSE forum. For help with writing and running code - https://warwick.ac.uk/rse/softwareforum
- Bugzilla. For help using the SCRTP systems, requesting software etc - https://bugzilla.csc.warwick.ac.uk/

Who pays for all this?

- Departments do
- https://warwick.ac.uk/scrtp/costs/
- Departments and Grant holders contribute to running and support. RIS can help you if writing proposals
- RSE training and support is supported by grants
- Generally your supervisor or PI should know what you can access

Mostly you shouldn't need to worry about this. Just keep in mind that somebody pays for core-hours (number of processors used times number of hours of use, the usual costing metric for compute time), and ask your supervisor/PI if you need details.
Note for any UG project students: your supervisor will know which machines and how much time you have, so if you're using SCRTP kit do be sure to ask them.

Access to Systems - Signing Up

How do I sign up - Desktop?

- Follow the steps at https://warwick.ac.uk/research/rtp/sc/desktop/gettingstarted/
- This takes you through signing up for desktop access, bugzilla, and the useful mailing lists

Please do sign up for the mailing lists and read them. They often contain vital information.

How do I sign up - Clusters?

- Once you've signed up for Desktop access, you can request cluster access at https://warwick.ac.uk/research/rtp/sc/hpc/register/
- Sign up for Orac or Tinis- pick whichever your group uses, or you can sign up for both by submitting two forms
- Note you have to provide a research group or department to manage the job prioritisation side of things

If your supervisor/colleagues uses one of the two, might want to pick that one. Otherwise, specs are very similar unless you want GPU nodes (probably Tinis) or one of the other test beds.
Orac has faster interconnect, and more cores per node, so for close coupled MPI codes (lots of comms) pick this
Orac is newer, so absent any other criteria, pick that.

Access to Systems - Read Then Proceed

Finding the Wiki

- The first place to go if you have problems is the Wiki - see https://wiki.csc.warwick.ac.uk/twiki/bin/view/Main/GettingStartedCategory
- You'll get access to this when you sign up for a desktop account
- Use the quick start guides and search function
- See also the Links page at the end of these slides

Bugzilla

- Bugzilla is the primary support channel
- Use if you have trouble getting access (e.g. lost ssh keys)
- Use for problems with installed packages
- Request new packages
- Access at https://bugzilla.csc.warwick.ac.uk
- You have to sign up for this separately to the desktop/cluster accounts - we recommend using your @warwick email address when you do

More on bug reporting in Session 3. But:
Always search Bugzilla for existing bugs before submitting your own. But do bear in mind that sometimes (e.g. in the case of the ssh keys) you're better making your own bug than commenting on somebody else's. It's also better to open a new bug with a reference to something which is related but not the same. Don't worry about the exact rules, just try and be helpful.

RSE Forum

- For issues that aren't really specific to the SCRTP facilities, try the RSE support forum
- We can give basic advice on techniques, bugs, profiling etc
- Might refer you to Bugzilla and vice versa
- Access at https://warwick.ac.uk/research/rtp/sc/rse/softwareforum

It's a bit tricky to say in general what should go on the forum versus bugzilla. Basically we as RSE's have no special access to machines, but we do know a lot about them, and might know of existing bugs etc.
E.g. the RSE forum can help with installing Python packages as a user, but not at system level

Coffee And Cookies

- Fridays at 2pm in the Physical Sciences common room
- Usually we (RSE) are there, often sys admins
- Can get *brief* pointers on clusters, software, techniques, recommendations
  - Informal discussions only!
- **Not a replacement for the official support channels!**

The csc-events mailing list posts this and other useful seminars etc, or see https://warwick.ac.uk/fac/sci/csc/news/calendar/

The Flowchart of Help

1. Yourself.
    1. For questions about access and available compute time, try your supervisor/PI
    2. Pay some attention to the mailing lists - sometimes there are outages/known issues
2. Try a search engine. See 3rd talk on error messages. Search the Wiki and Bugzilla
3. File a bug/post on Bugzilla or the RSE forum

A Few Essentials

Computer Terms

- Memory and Storage
  - Memory is RAM. At the moment measured in GB, or TB for a very large machine. Fast.
  - Storage is disk space. Can be of various sorts but nearly always refers to “permanent” storage. Much slower than RAM, usually much larger.
  - May hear the word “swap”. This uses disk space to get extra memory, at cost of slowness.
  - Cache usually refers to even faster, even lower capacity memory, “closer” to the processor.

Hopefully this is nothing new, but it can lead to real confusion if things aren’t absolutely clear.

Memory is (usually) temporary, cleaned up when a program exits. It’s specific to a program (unless using advanced techniques, or suffering e.g. the current crop of major vulnerabilities, Meltdown and Spectre).

Storage can be on all sorts of disk types and use all sorts of access protections, but it persists even when you turn off the computer (except /tmp/ directories, which are deliberately cleaned up).

A “swapping” computer is having to swap data from disk to RAM, so it usually grinds to a halt, but at least you have a chance to intervene.

Program data has to be moved from “main memory” into cache before the CPU can work with it. Cache also holds programs etc. You may hear “fits in cache” as a Very Good Thing™ because it means less moving back and forth of parts of a data set/program etc

Computer Terms

- Crashing and Hanging
  - Crashing usually means the program stops before expected, often with an error message
  - Hanging is when it does nothing, and fails to show expected signs of progress
  - Hanging is not running slow
  - Giving the wrong answer is a different sort of problem
- If in doubt when reporting problems, provide all the information you can, but not so much you swamp somebody trying to help

This one is also sort of obvious, but you'd be amazed how often error reports come in the form of "The code doesn't work" with no further clarification.
Also, try to distinguish true hanging from running but very slowly. The causes and solutions are usually very different. This isn't always easy.

If possible when reporting errors, use file attachments rather than pasting long errors into a posting. And remember, if you don't know what the problem is, it's unlikely you can identify which parts of the error are relevant, so it's often better to give everything.

Wrap Up

Summary

- Desktop and Cluster computing is available for you to use
- All the info for signing up etc are on SCRTP pages
- The Wiki contains almost everything you'll ever need to know about the machines etc
- But if you do have problems there are a few support channels

Helpful Links

- SCRTP main page - https://warwick.ac.uk/scrtp
- Warwick RSE main page - https://warwick.ac.uk/rse
- Wiki
  - clusters https://wiki.csc.warwick.ac.uk/twiki/bin/view/Main/ClusterUserGuide
  - desktops https://wiki.csc.warwick.ac.uk/twiki/bin/view/Desktop2018/DesktopComputingCategory
- RSE forum. For help with writing and running code - https://warwick.ac.uk/rse/softwareforum
- Bugzilla. For help using the SCRTP systems, requesting software etc - https://bugzilla.csc.warwick.ac.uk/

Before Next Time

- Register for a Desktop account
- Register for Cluster use
- Make sure you can access the cluster(s) you signed up for
  - See e.g. https://warwick.ac.uk/research/rtp/sc/rse/training/linuxdesktop/remoteaccess
  - Email rse@warwick.ac.uk if you have trouble with ssh keys etc

username@godzilla.csc.warwick.ac.uk

username@orac.csc.warwick.ac.uk

username@tinis.csc.warwick.ac.uk

By the way, although your viewer might think those are email addresses, they aren't. You login to a machine as {username}@{machinename}.{domain}
Note the machine names use the CSC domain.