

CCP-EM Working Group 2

Manchester, Thursday 30th April 2015

Minutes

Talk1.

Alan Roseman (Manchester) showed how organic growth of a cluster of workstations can become cumbersome. They have good access to centrally run archive services, and compute clusters, but low network bandwidth is a limitation.

Talk2.

Corinne Smith (Warwick) explained their unique requirements in their current project. Low particle numbers but large images. This illustrates a point about a standard computing setup that will be expanded later – no one-fits-all solutions. The equipment needs are not always understood by IT department, and they need to argue a strong case with them to get the right equipment. Main software used: Frealign, relion, UCSF software (motioncorr, SAM ctf), spider.

Talk3.

Morgan Beeby (Imperial, London) has a cluster of workstations used for subtomogram alignment. Integrated with this are terminals in the lab for electronic notebooks. The compute needs for his tomography applications are not as high as for single particles. Programs used: Peet, imod, leginon, tomocf

Talk4.

Bettina Boettcher described her setup in Edinburgh. She has her own small compute cluster with windows machines as terminals. The windows machines can also run most of the interactive software, such as chimera, eman2,... Finding the suitable staff to support her cluster is a problem, and it is in a server room she does not have access to. Most computing is done on a cluster or Edinburgh University clusters. An interactive node on the cluster is reserved for some computing and is better than PCs or workstations because of bandwidth limitations for accessing data.

Talk 5.

Peter Rosenthal described his set up at the NIMR (Mill Hill, London). As they plan their move to the Crick Building, everything might change. He emphasized the importance of the human personnel element in setting up and delivering a good computing environment and software tools. He gave an example of a useful package developed by a post doc, but is now harder to use since he has left. Other packages used: Amira for bespoke segmentation. Peter motivated an idea to capture and keep track of workflows for certain tasks or projects.

Talk 6.

Neil Ranson (Leeds) professes he has excellent compute support. He does not allow group members or himself root passwords. All computing issues are resolved via his sysadmin. Hardware upgrades and systems are jointly decided with the sysadmin, with good operational success in mind. He said reduction of choice and standardization of systems made the job easier, and allowed more efficient use of resources. His lab uses Mac front ends to the cluster computing. Computers linked with Gigabit network. They have a local cluster for 'on demand' processing, and access to larger clusters. Moving large datasets across the network is also an issue.

This point was debated by the participants. We have seen the diversity between our own systems, and that in part all of them have inherited some historical component. We can't restart from somewhere else. It was generally agreed that while in each lab we could apply this principle of simplification and standardization, ccpem should serve the community as a whole. Alan explained how ccpem was adopting "dev tools" and "build bot" from ccp4, and these tools were designed to allow a release that could be distributed successfully to a wide range of compute platforms.

Neil also suggests reliable RPMs or distribution vehicles would make their systems admin job easier.

Summary of important issues raised:

Make distribution of packages easy to install. How: help the developers. Rewrap? Use docker type solutions? Bring into the ccpem distribution. Problems are: obsolete/outdated/unavailable libraries, outdated nvidia drivers, etc.

Standard ISO for an EM workstation:

Follow the CCP4 model. Have everything possible available. Sort out any potential clashes. Have a menu at the beginning to allow selection or de-selection of packages. Alan floated the idea to have a standard place for the software packages that could not be distributed by CCP4, so they could be installed at a standard point, which would keep the directory path standard, and help with standard scripts.

Ideas for projects:

Database for tomography and/or subtomogram averages
Virtual machine/iso/some delivery of working software system.
Keep track of workflows.

Specific software issues:

CCPEM packages to be able to be installed without sudo rights
Graphic cards problems. Problems with drivers or Cuda versions.<Wiki/forum/notes.

General topics and ideas for CCPEM to help on:

Advice
Contact point

Packages need frequent updating in this fast moving field.
Advise on hardware purchases.
Advice on negotiating with local IT departments on computing requirements, hardware and configuration.
Block grant for EM time
UK wide service contracts for EM's
Training courses
Wiki workshops, downloadable guides/primers.
Specify a recommended OS and standard set up for a new lab. i.e. similar to a ccpem local system so best support can be provided.

Parallel computing

Cluster problems and setup, can there be a general solution or are they all too individual?
Cloud computing advice and help. Forums. Scripts exchange?

Possible topics for primers, notes or introduction courses

Setting up group working
Background of file systems
Mounting external file systems
Remote working tools
Data Security
Intro to sys admin, or how to talk "sys admin language".
3D screens

Thank-you also to international colleagues for their email contributions.

Alan Roseman, Manchester, May 2015