ECDF Services Update Feb – May 2009

Mike Baker, ECDF Middleware Team
Orlando Richards, ECDF Systems Team
21st May 2009

1 Introduction
This paper describes developments and work carried out by both the Middleware team and the Systems team. The ORG are requested to
• Comment on services
• Suggest improvements or further services

2 Systems Team Report
The systems team have mainly been working on Subversion and GPGPU services over this time, as well as doing planning work on the compute cluster successor and on NAS services.
The hosting facility required a power down in April to allow for essential maintenance work and compliance testing on the building's power supply, which involved a complete shut down and restart of all the ECDF services.

3 Middleware Team Report
During this period it was agreed that the middleware team should migrate from NeSC to IS during 2009, and that a complement of 3 FTEs would provide a sustainable resource for the University. Current staffing level is 2, one staff member having left post, which puts current services at risk as the team relies on the goodwill of staff to work out of hours and during holidays in order to keep the existing services running. The process is under way to replace the 3rd team member.

4 Pre-existing services
Developments with the ECDF compute cluster service since February 2009 include:
• Moved to a new image management system, which will allow us to deploy Scientific Linux 5 (and, potentially, other operating systems) in the near future.
• Changed the architecture of the GPFS parallel file systems to allow for greater separation of bottlenecks, without reducing resilience.
• Deployed and incorporated fast disk storage for critical components of the parallel file systems, with continuous service.
• Continued to deploy updates to managed applications.
• Streamlined the user creation processes.
• The backup and SAN services have remained satisfyingly static.

A large user (and income-stream) of the ECDF (GridPP) relies on support of the EGEE glite middleware on ECDF. It is hoped that this middleware will be increasingly used by other disciplines within Edinburgh.

Our collaboration with the NGS continues, and the national infrastructure that is available to Edinburgh researchers is now the basis of projects in BioInformatics and Geosciences, and Edinburgh's participation within the NGS is seen as key to helping collaborative bids.

Members of the middleware team have carried out consultancy work for several groups over the past quarter, typically becoming an embedded member of the project for a period, then providing support for the projects during their lifespan.

5 New services
A GPGPU service has been deployed, adding an additional 4Tflops of computing power (on top of the cluster's current 17Tflops). This is provided by a single Nvidia Tesla S1070 system, which is connected to two of Eddie's phase 1 worker nodes. This technology offers very large potential for accelerating some computational tasks. We are continuing to push the use of this technology to our users where it might be useful.

We have deployed a web server cluster to serve two main purposes:
• To provide a platform for publishing live system information, such as queue status and system load monitoring.

• To allow users to make data held on Eddie's cluster file system accessible over http (and https).

Reduced staffing in the middleware team has hampered our ability to deliver new services this quarter. We continue to await permission from the University to advertise for two replacement staff.

The middleware team and systems team have worked together to provide a hosting infrastructure for research projects that wish to develop alternative access methods (such as portals) to access the ECDF. We are making use of virtualisation technology to provide this service, and are in the process of rolling this out to our first research group (see Genomics portal development below).

6 Developing services

Work continues on the subversion service, and we are confident that we will have the service up and running by early summer. Significant technical challenges have been encountered, particularly with respect to integrating the service with the University's authentication systems.

We have a release of the accounting and reporting service deployed, and it is capable of reporting on all the past ECDF accounting records. We are working closely with Stephen Booth in EPCC to help with the development of this service - both to meet our requirements, and the requirements of the JISC funded project. There is ongoing consultation with the user community and with the schools to ensure that the system will provide the reporting that is desired.

We are currently developing a Scientific Linux 5 image for our worker nodes. This should be ready in June, at which point we will provide a test environment for users to explore, and consult on requirements for both Scientific Linux 4 and Scientific Linux 5 to come up with a suitable migration plan.

We are currently working with Cell Biology, helping them build a portal for the analysis of genomics data. Several other groups within biology have been identified as future users, and have indicated a desire to fund ECDF compute time used via this service.

After extensive requirements gathering meetings with the School of Economics, we have specified a new statistical analysis service for their users. Currently limited by memory on their desktop computers, the new service will allow them to process much larger data sets than is currently possible.

7 Future projects

These NAS services and compute cluster refresh projects are in relatively early stages of planning and resourcing. Further detail is available in the accompanying papers.

The Condor trial (utilising unused CPU cycles on desktops and clusters for researchers) had to be put on hold during 2008 due to loss of team members. This is seen as a way of providing more computation to researchers at no capital cost, and at a very low environmental/power cost, and is even more important in the current economic climate. This trial can be resumed once the middleware team is able to recruit additional staff.

There is growing interest amongst researchers in virtualisation of compute resources, both using the Cloud Computing metaphor as popularised by Amazon and Google, and for running services that would otherwise leave servers under-utilised. The middleware team has attracted funding from the National Grid Service to investigate Cloud Computing, and once the necessary staff are in place will investigate future services for researchers.

Traditional command-line access to resources such as the ECDF is seen as a barrier to many. During requirements gathering a number of groups have been identified who would like to build discipline-specific portals to enable access to these resources for large numbers of researchers. The cost to build such portals has been seen as prohibitive in the past. A NeSC-developed technology, RAPID, has the potential to simplify the process, and several groups are interested in building services using it. In order to lower the cost of all these groups building their own infrastructure/hosting, the middleware team intends to build a service to host these portals, and make the ECDF available as a resource for the users.