# Proposal for use of Condor at DL

John Kewley

August 17, 2004

# 1   Scope

The aim of this document is to make an initial proposal for the use of Condor to utilise unused computer resources within CCLRC with initial focus on Daresbury Laboratories, and in particular the e-Science Centre.

# 2   Background

Condor is a research project of many years standing from Miron Livny's group at the University of Wisconsin in Madison. Its goal is (quote from Condor site):

> "To develop, implement, deploy, and evaluate mechanisms and policies that support High Throughput Computing (HTC) on large collections of distributively owned computing resources. Guided by both the technological and sociological challenges of such a computing environment, the Condor Team has been building software tools that enable scientists and engineers to increase their computing throughput"

More simply, it can also be seen as enabling the use of otherwise unutilised computing resoures to enable scientists to perform additional calculations that they would not otherwise have the resources for.

## 2.1   Potentially Available Resources

Here in eSC, and indeed throughout Daresbury, there is a variety of potentially available computing resources, much of it used only during working hours with little overnight or weekend use. These can be categorised as follows:

1. "Personal" workstations, often running Windows XP/2000, but also Linux; some of these machines may be dual-boot. In general the Windows machines have a common

setup and are administered by DL PC Support, whilst the Linux ones are truly personal and are administered locally, often by the owner and are of many "flavours" including RedHat, Gentoo, Suse and Whitebox.

2. Servers - e.g., eSC's Blade Servers, these are used for "serious" and externally visible uses ... Some of these run Linux; the rest, other types of Unix

3. Miscellaneous machines: IBM cluster, machines in the training laboratory and access grid. These machines have intermittent serious use, but have clear times when they are effectively "free".

4. Laptops: by their nature, these resources are not always on-site or online, especially overnight. For this reason, they will not initially be targets for inclusion in a Condor pool.

## 2.2   Supported Architectures

Until quite recently, for Windows machines, Condor was only advertised as supporting NT4. Now, however it is supported (albeit limited functionality) on Windows NT4.0 Workstation and Server, Windows 2000 Professional and Server, Windows 2003 server and Windows XP Professional. This makes it significantly more useful to potential DL users.

For Linux, Redhat 7.1, 7.2, 7.3, 8 and 9 are all now fully supported on Intel x86. SuSE Linux Enterprise Server 8.1 is also supported on Intel Itanium. Other Linuxes are also likely to work, but are not officially supported. For a given Linux, you must match the architecture with the kernel and glibc version numbers of your Linux flavour to get the correct download (for fuller details see "Report on Condor Installation").

# 3   Proposal

## 3.1   Rationale

The original plan was to create a single large pool at DL. This has proved initially impractical for the following reasons:

1. If we are to expect DL staff to release their workstations for Condor use, we should minimise any security implications. Therefore all "personal" workstations which are inside our firewall should not be available to run external user's jobs.

2. There is potential to use Condor for MPI jobs, there are Condor installation implications for this, and so Condor MPI-capable resources would be better kept separate from the "personal" workstations.

3. During the course of the evaluation, although Condor continued to release new versions of theor software which, although not adding much additional functionality, expanded

on the list of platforms that supported certain features. Being able to test such releases in isolation is useful.

4. Providing an externally visible resource to the UK Condor effort will be good for DL visibility and would be a suitable activity for the e-Science Centre.

## 3.2 Initial Solution

Combining these points, it can be seen that these aims can be met through the formation of two Condor pools, one totally inside the site firewall containing many "personal" and spare Linux workstations, and using the IBM Xeon Cluster (which is outside the firewall) as the core of an MPI/test machine.

### 3.2.1 Internal Pool

To form the initial pool, we need to utilise resources owned by people who are not expected to be "Users" of the system. This requires "buy-in" from these machine owners. They gain no benefits from adding their machine to the pool, therefore it is essential that we ensure that any impact on their machines is an minimal as possible. The should be largely unaware of any Condor jobs that run on their machines.

Once a critical mass of these machines is brought together, we can allow users access to the pool, on the condition that they add machines of their own.

Priority for inclusion in this initial pool will be Redhat and Windows machines.

### 3.2.2 Test Pool

The IBM cluster is outside the main firewall and should be setup initially as a Test Pool and at a later date as an external pool to provide additional (albeit a small) resource to the UK Grid and for use in large scale condor flocking experiments.

Prospective uses include:

- Test Condor releases as they are available

- Test installation scripts

- Test a wide range of Condor functionality, e.g. MPI support, as this will not be supported in the main pool.

- Gain sufficient understanding of Condor to be able to field support centre queries.

- Flocking testing to and from the new CCP4 cluster

# 4　Future Directions

- Flocking

  - allow internal pool jobs to flock to the test pool
  - allow external access to the Test pool, ie make it an externally visible resource which can be accessed through a gatekeeper and/or flocked to from other pool outside of CCLRC. This effectively evolves the test pool to be a fully fledged (sic) External Condor Pool.

- RAL

  - Once DL pool is showing some use, setup a pool at RAL

- CCP4

  - Evaluate how we can help CCP4 with their new cluster
  - Maybe their old mini condor pool can be joined up to either our internal or test pools, likewise for the new cluster.

# 5　Documentation

Additional documentation will be needed to support this work in the near future:

- Some notes on the multitude of Condor releases, release formats and installation possibilities for the Linux and Windows platforms.

- A User Guide. This is basically a "Managing expectations" guide and will explain to both Users and resource owners their respective roles.

- There is likely to be a future need for an installation notes/script/guide. This would be neccessary to help systems administrators who wish to setup condor nodes, this will be refined as more experience is gained on the different Linux and Windows machines.

- DL Security considerations. This will include a discussion on the security implications of having Condor running on your machine, hopefully with some input from PC Support on the various security configuration that are currently in use at DL.