Managing Supercomputing Resources at the University of Manchester

Michael Pettipher, *University of Manchester*

ABSTRACT: *The University of Manchester has provided a supercomputing service for over 30 years. The University also provides other national services and support a very large local user community. While the systems on which the services are provided have changed enormously during this period, the objective to provide the user community with the simplest and most flexible way to access such resources has not. The fact that staff have been involved in such activities for a long period, and that they are very familiar with the requirements of a diverse range of users, has resulted in the desire to provide a single, consistent, flexible system to manage the wide range of services provided. The system now in use satisfies all of these objectives, taking into account specific requirements of the different services. A major objective has been to devolve resource management as far as possible to the end users, thus reducing central management and increasing control by the users themselves. The whole system is web based, thus helping to make it as accessible as possible to the whole user community. This paper explains the rationale behind the system used, shows how it is used both at a user and an administrator level, and also indicates how it will be further developed.*

1. Introduction

The University of Manchester

The University of Manchester was established in 1851 and is now one of the largest universities in the UK, both in terms of student numbers and in research income. It is also one of the most successful, with respect to research quality , teaching and student employment.

Manchester Computing

Manchester Computing (MC) is the Computing Service department, which is distinct from the academic department of Computer Science, although there is substantial collaboration between the two departments, and Manchester Computing is involved in numerous research activities. The department is responsible for a large number of computing services, not just supercomputing. It supports local computing facilities from desktop machines to local supercomputers as well as two national/international services.

The MIMAS service (Manchester Information and Associated Services) hosts a number of dataset services, covering bibliographic references, electronic journals, scientific data, socio-economic data, and spatial data.

Supercomputing has been important at Manchester since 1948, when the first stored program computer was developed at the University. National supercomputing services have been provided since 1972, starting on CDC 7600's, followed by a series of vector processors. The current service, CSAR (Computing Services for Academic Research), is based on an SGI Origin 3800 and a Cray T3E, with an SGI Altix system being installed this summer.

Supercomputing in the UK

There are two national supercomputing services in the UK:

- CSAR. This is a private finance initiative (PFI) service, based on SGI and Cray systems, provided by the CfS consortium:
  - University of Manchester
  - Computer Sciences Corporation
  - Silicon Graphics Inc
- HPCx. This is a traditionally funded (not PFI) service, based on IBM systems, provided by the HPCx consortium:
  - Edinburgh Parallel Computing Centre
  - Daresbury Laboratory
  - IBM

2. Access to the CSAR Service

There are four classes of project available on the CSAR service:

- Class 1 – Full peer review
- Class 2 – Pump-priming
- Class 3 – New application areas
- Class 4 – 'Commercial' use

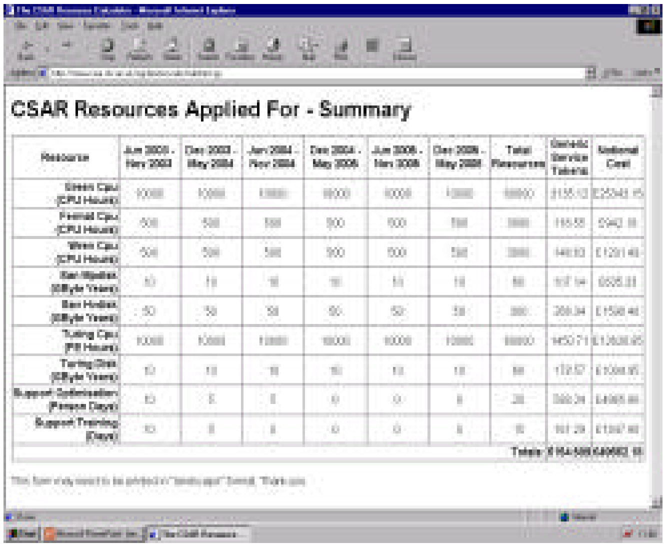
All major work is carried out under Class 1. The procedure for obtaining Class 1 resources on the CSAR service is similar in principle to that on most supercomputers: the Principal Investigator (PI) must submit a project application, indicating the resources required and providing justification for the request. The application is reviewed by technical referees and by CSAR, on behalf of the funding body – the UK Research Councils. If approved an allocation of tokens is made.

However, as this is a private finance initiative, resulting in a ‘pay for resources used’ approach, there are many differences in the way the system is managed and charged, in comparison with most other supercomputing services.

The first major difference is that all resources are charged, rather than just cpu and possibly disk. The following list gives the ‘purchaseable resources:

- SGI Origins (3 different systems) and Cray T3E cpu
- SGI Origin (SAN) and Cray T3E disk
- Tape storage
- Training Courses
- Optimisation and application support
- Guest system resources.

All applications must include details of each resource expected to be used It is recognised that it can be very difficult to accurately estimate such resource usage over a three year project, so flexibility in the service allows projects to ‘trade’ some resources for others during the lifetime of the project. As an aid in producing these estimates, a resource calculator is provided on the CSAR web pages. This tabulates the resources requested and gives a total in ‘generic resource tokens’, as well as a notional cost, used by the funding bodies, as indicated in figure 1.



CSAR Resources Applied For - Summary

Resource	Jun 2003 - Nov 2003	Dec 2003 - May 2004	Jun 2004 - Nov 2004	Dec 2004 - May 2005	Jun 2005 - Nov 2005	Dec 2005 - May 2006	Total Resources	Generic Service Tokens	Notional Cost
Green Cpu (CPU Hours)	10000	10000	10000	10000	10000	10000	60000	[illegible]	[illegible]
Fermat Cpu (CPU Hours)	500	500	500	500	500	500	3000	116.55	[illegible]
Wren Cpu (CPU Hours)	500	500	500	500	500	500	3000	140.10	£1201.40
San Wpdisk (GByte Years)	10	10	10	10	10	10	60	[illegible]	[illegible]
San Hodisk (GByte Years)	50	50	50	50	50	50	300	209.34	[illegible]
Turing Cpu (PE Hours)	10000	10000	10000	10000	10000	10000	60000	1450.71	[illegible]
Turing Disk (GByte Years)	10	10	10	10	10	10	60	172.57	[illegible]
Support Optimisation (Person Days)	10	5	5	0	0	5	25	[illegible]	[illegible]
Support Training (Days)	10	5	5	0	0	5	[illegible]	[illegible]	[illegible]
								Totals: [illegible]	[illegible]

This form may need to be printed in "landscape" format. Thank you.

.

Figure 1: Output from the CSAR Resource Calculator

3. Resource Management System Requirements

General Requirements

The resource management system has been designed to cover registration and all resource management functions for a variety of services managed by Manchester Computing. The following were seen to be general requirements across most of the services:

- Minimise administration
- Make it ‘easy’ to register
- Provide a rapid response
- Web based
- Hierarchical structure to allow devolvement of tasks. ‘Approved’ users (e.g. PIs) should be able to manage as many resources as possible on behalf of their specific groups.

CSAR Requirements

Significant additional functionality was required for the CSAR service:

- Self registration
- Allocate any resource
- Ability to trade resources
- Facility for sub-project management
- Usage reporting
- Capacity planning
- PI is the primary administrative level. Must be able to:
    - Authenticate users within consortium
    - Sub-allocate resources within consortium
    - Set and change individual user allocations and disk quotas

Local Supercomputing Requirements

These are a subset of the CSAR requirements, but registration and resource management is managed centrally.

MIMAS Requirements

Registration of the MIMAS service (about 7500 users) requires central registration, but resource management (changing passwords, disk quotas etc.) is devolved to site representatives

One particular service, the Crossfire service. Requires self registration (for about 13000 users), but uses a different authentication process.

Thus there are shared and specific requirements for the MIMAS and CSAR services.

Other Requirements.

Other requirements identified include:

- Logging of all transactions
- Live and test sites to allow continual development
- On-line help
- Compatibility with Grid activities
- Integration with query management system
- Feedback from the user community

All of this functionality has been provided in the resource management system currently in use at the University of Manchester.

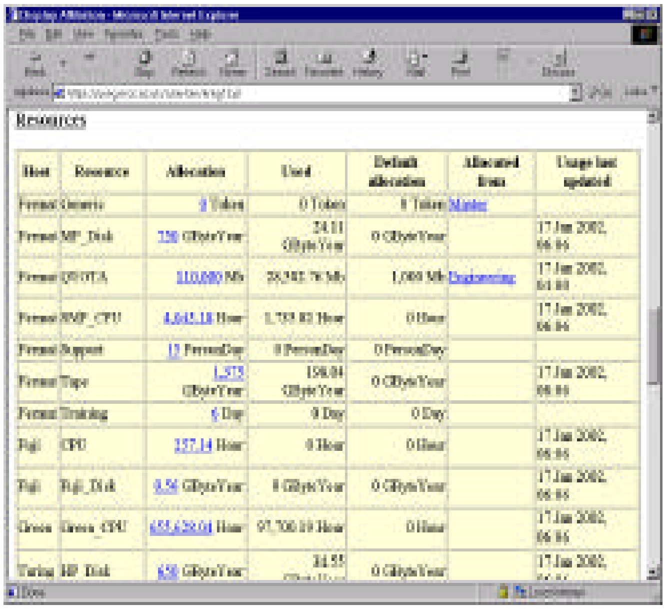
4. Structure and Implementation

One of the objectives in designing the system was to integrate with the query management system, used by the helpdesk service in managing all queries received at Manchester Computing. This query management system is based on the Action Remedy System (ARS), which utilises Oracle as the underlying database pacakage.

The resource management system is written in Perl and built as a layer on top of ARS. It is necessary for the system to execute commands on the host systems (for example to register a user or change a disk quota) It is also important that feedback is received from each host, to ensure integrity between the resource management database and the actual systems. (A system administrator could reset a disk quota directly without using the resource management system – this will subsequently be discovered by the resource management system, and its database will be amended accordingly.)

A key feature of the system is that it is 'person based'. Thus there is only one entry for an individual, but there may be any number of usernames on different hosts associated with this person.

The following figures show web pages, accessible by authorised users, demonstrating some features of the system



Resources

Host	Resource	Allocation	Used	Default allocation	Allocated from	Usage last updated
Fermat	Generic	8 Token	0 Token	8 Token	Master	
Fermat	MP_Disk	750 GByteYear	24.11 GByteYear	0 GByteYear		17 Jan 2002, 06:06
Fermat	QUOTA	110,000 Mb	28,282.76 Mb	1,000 Mb	Engineering	17 Jan 2002, 04:00
Fermat	SMP_CPU	4,645.18 Hour	1,737.82 Hour	0 Hour		17 Jan 2002, 06:06
Fermat	Support	17 PersonDay	0 PersonDay	0 PersonDay		
Fermat	Tape	1,575 GByteYear	198.04 GByteYear	0 GByteYear		17 Jan 2002, 06:06
Fermat	Training	6 Day	0 Day	0 Day		
Fuji	CPU	137.14 Hour	0 Hour	0 Hour		17 Jan 2002, 06:06
Fuji	Fuji_Disk	0.56 GByteYear	0 GByteYear	0 GByteYear		17 Jan 2002, 06:06
Green	Green_CPU	655,628.04 Hour	97,700.19 Hour	0 Hour		17 Jan 2002, 06:06
Turing	MP_Disk	650 GByteYear	34.55	0 GByteYear		17 Jan 2002,

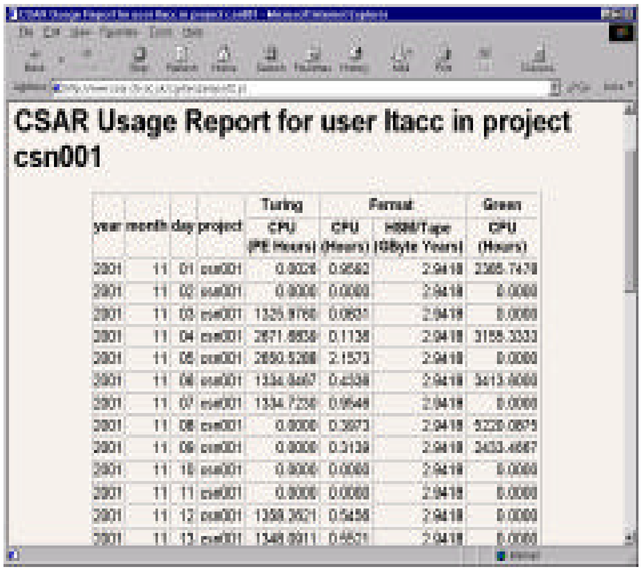
Figure 2: Details of some of the resources allocated and used.

Figure 2 shows details of some of the resources allocated and used:

- Fermat is an SGI Origin2000
- Fuji a Fujitsu VPP300
- Green an SGI Origin3800
- Turing a Cray T3E

–

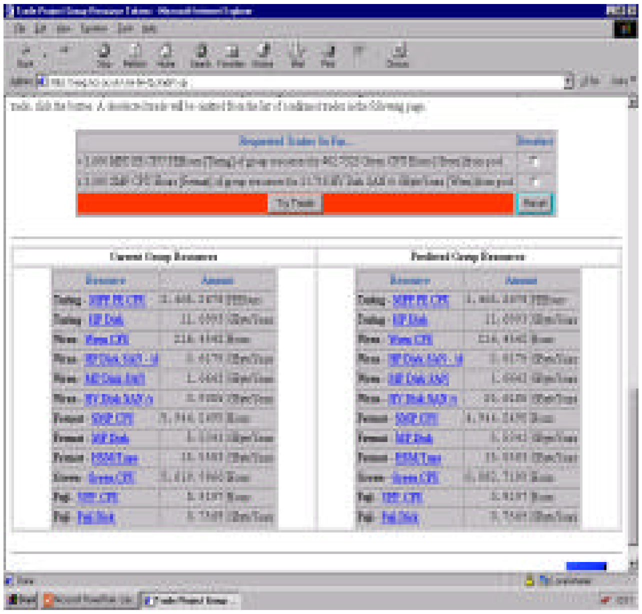
This information is available directly to the PIs.



CSAR Usage Report for user ltacc in project csn001

				Turing	Fermat		Green
year	month	day	project	CPU (PE Hours)	CPU (Hours)	HSM/Tape (GByte Years)	CPU (Hours)
2001	11	01	csn001	0.0028	0.9582	2.9418	2385.7478
2001	11	02	csn001	0.0000	0.0000	2.9418	0.0000
2001	11	03	csn001	1325.9760	0.0821	2.9418	0.0000
2001	11	04	csn001	2871.8856	0.1136	2.9418	3158.3333
2001	11	05	csn001	2850.5208	2.1573	2.9418	0.0000
2001	11	06	csn001	1334.8467	0.4338	2.9418	3413.0000
2001	11	07	csn001	1334.7236	0.9648	2.9418	0.0000
2001	11	08	csn001	0.0000	0.3873	2.9418	5220.0875
2001	11	09	csn001	0.0000	0.3139	2.9418	2433.4667
2001	11	10	csn001	0.0000	0.0000	2.9418	0.0000
2001	11	11	csn001	0.0000	0.0000	2.9418	0.0000
2001	11	12	csn001	1359.3621	0.5458	2.9418	0.0000
2001	11	13	csn001	1348.0911	0.6621	2.9418	0.0000

Figure 3: Resource usage on a daily basis, available to all users.



Requested Trades In Pool	Deselect
[illegible]	□
[illegible]	□

Current Group Resources

Resource	Amount
Turing - MPP PE CPU	1,466.2478 PEHour
Turing - MP Disk	11.0993 GbyteYear
Wren - Wren CPU	216.4142 Hour
Wren - MP Disk SAN - [illegible]	0.6179 GbyteYear
Wren - MP Disk SAN	1.0662 GbyteYear
Wren - [illegible] Disk SAN [illegible]	0.0000 GbyteYear
Fermat - SMP CPU	5,944.2455 Hour
Fermat - MP Disk	5.5561 GbyteYear
Fermat - HSM Tape	15.5161 GbyteYear
Green - Green CPU	5,418.5960 Hour
Fuji - VPP CPU	5.8197 Hour
Fuji - Fuji Disk	5.7569 GbyteYear

Predicted Group Resources

Resource	Amount
Turing - MPP PE CPU	1,466.2478 PEHour
Turing - MP Disk	11.0993 GbyteYear
Wren - Wren CPU	216.4142 Hour
Wren - MP Disk SAN - [illegible]	0.6179 GbyteYear
Wren - MP Disk SAN	1.0662 GbyteYear
Wren - [illegible] Disk SAN [illegible]	[illegible] GbyteYear
Fermat - SMP CPU	5,944.2455 Hour
Fermat - MP Disk	5.5561 GbyteYear
Fermat - HSM Tape	15.5161 GbyteYear
Green - Green CPU	[illegible] Hour
Fuji - VPP CPU	5.8197 Hour
Fuji - Fuji Disk	5.7569 GbyteYear

Figure 4: Trading pool – for PIs to trade allocations of different resources.

5. Future Developments

A new version of the resource management system is now being developed. There are three areas in which particular consideration is being given.

General Enhancements

The structure of the system is being made more modular, in particular by divorcing the system from the underlying ARS and Oracle software. Thus it will be easier to port to an environment without this software. Changes are also being made to improve the performance of the transactions.

A significant enhancement will be the inclusion of ‘role association’, as well as ‘personal association’. At present entries are person based, but a person may be given the authority to perform certain tasks because of their role, such as a PI. The role association would allow the role to be transferred to another person if appropriate (because of changing job, for example).

Project Unity

The University of Manchester is merging with the University of Manchester Institute of Science and Technology (UMIST), resulting in a single even larger institution. This is a major project for both universities. In the context of user registration and resource management, it provides an opportunity to unify the various schemes currently operating at both institutions. At both sites, there currently exist a university registration scheme for all staff and students, and another scheme for specific systems managed centrally. The new resource management scheme will provide the capability to manage all of these separate schemes.

Grid-motivated Enhancements

User registration, which could be regarded as one aspect of authentication and authorisation, and resource management are major issues in the exploitation of the Grid. Thus in the development of a new resource management system, one of the objectives is to ensure that this system is as compatible as possible with requirements for using the Grid.

In order to describe the sort of steps being considered, it is useful first to provide some detail of the current status and challenges for effective Grid use in the UK

The UK is pursuing major initiatives in the area of e-Science and the Grid, with the underlying objective of encouraging the exploitation of the Grid for ‘real science’.A number of e-Science centres have been set up around the country, and a variety of e-Science projects have been initiated:

- Industrial projects
- Pilot projects
- Interdisciplinary research projects
- Demonstrator projects
- International projects ,

UK Grid Resources

Most e-Science centres have committed resources for Grid use. In addition, both national supercomputing services are committed to supporting activities on the Grid, and will make some national resources available. (Globus and UNICORE are already in use via the CSAR service.)

In the short term, the major UK Grid users will be those with e-Science pilot projects, most of whom have been awarded significant resources on the national systems, and have access to other local and e-Science facilities.

Grid Registration – The Challenges

There will be potentially thousands rather than hundreds of HPC users, and tens of thousands social science and library users. It will be impractical and inefficient for registration of such numbers of users to be managed centrally.

In general, centres do not own resources being made available to the e-Science programme, nor do they necessarily have direct administrative control of them. Thus again, the ‘standard’ administrative procedures developed for managing clearly defined and controlled resources are inappropriate.

For example, consider a Globus-based Grid use wanting to use resources under the UK e-Science programme. This user will require, for each system to be used:

- A username on the system
- A personal certificate (X509), issued by the UK e-Science certificate authority.
- An entry in the grid-mapfile (to map the certificate name to the username).

Typically entries in the grid-mapfile are managed by the system administrator (by default, the grid-mapfile is owned by root). It is necessary to ensure that the user making the request is the same person as owns the private key of the certificate. Note that a user may legitimately possess multiple certificates, issued by different certification authorities.

The current, informal procedures where grid-mapfile maintainer knows users personally cannot scale for the projected numbers of Grid users.

Similar model exists in UNICORE, with the UNICORE User Data Base (UUDB) replacing the grid-mapfile.

This can clearly lead to complicated and excessive administration.

Accounting

Most sites providing resources on the Grid need to know to whom the usage should be charged, which means knowing at least the user and the amount of each resource used.

In the longer term, additional details will be required:

- A logging infrastructure that records resource usage.
- A commonly understood means of describing charges.
- A mechanism to negotiate charges between the consumer and the service provider.

- A secure payment mechanism.

These issues are currently being pursued through the UK e-Science project ‘A Market for Computational Services’.

How can the resource management system help?

The University of Manchester Resource Management System contains features which are applicable to many of these Grid related issues.

For example, as username creation is already delegated to the PI, it makes sense to similarly delegate control of the grid-mapfile or UUDB entries.

It would also be possible to add digital certificate authentication, as an alternative means to access the resource management system (user imports X 509 certificate into browser).

Similarly, one (or more) certificates can be associated with the person who owns the accounts.

A charging mechanism exists for a variety of resources, with the ability to trade resources as required.

Capacity planning may help to assess future requirements.

6. Summary

The Resource Management System currently in use at the University of Manchester handles a wide range of resource management activities.

A major objective has been to devolve administrative tasks as much as possible, to reduce the burden on system administrators and to give the users more control.

A new version under development will provide more flexibility and be capable of supporting a wider range of services both for local and national users of systems based at Manchester, and hopefully also in the Grid environment.

The developers are open to further suggestions

7. Acknowledgements

Two members of staff at Manchester Computing have been particularly helpful in producing this paper:

Phil Stringer is the author of the resource management system described in this paper. He has been a system administrator of numerous systems based at the University of Manchester, and this experience has enabled him to build a system which is sufficiently flexible to meet a very diverse range of services.

Stephen Pickles has a thorough understanding of the resource management system, and how it may be used or adapted for use on the Grid.

8. About the Author

Michael Pettipher is the manager of the High Performance Computing Services team in Manchester Computing, at the University of Manchester, U.K. E-mail: m.pettipher@man.ac.uk