NATIONAL COMPUTATIONAL INFRASTRUCTURE

Providing Australian researchers with world-class high-end computing services

PROGRESS REPORT

July 2007 - June 2008

September 2008

Table of Contents

1. INTRODUCTION

This report outlines the progress and achievements of National Computational Infrastructure (NCI) during the twelve month period July 2007 – June 2008 and alludes to ongoing progress in the period July – September 2008 that will be reported in detail in next year's progress report.

The NCI project is funded as part of the Platforms for Collaboration (PfC) Investment Plan. Its purpose is to provide state-of-the-art national computational infrastructure facilities and services to the Australian research community and builds on the successful National Facility (NF) program of the Australian Partnership for Advanced Computing (APAC).

The NCI project was established and funded through a contract between The Australian National University (ANU) and the Commonwealth of Australia, represented at the time of signing (22 June 2007) by the Department of Education, Science and Training, and subsequently by the Department of Innovation, Industry, Science and Research (Department of Innovation) following changes to the structure of the Commonwealth Government after the November 2007 election.

The program has three main activities –Operational, Outreach and Policy/Planning — the status of which are summarised in Section 2 and expanded upon in Sections 3 and 4 of this report.

More information concerning NCI can be found at its website (www.nci.org.au).

2. PROJECT STATUS

The activities and achievements of the NCI programs and related activities are summarised below and detailed in the report on activities. This and the following section review progress against the Business Plan and comment on any variations from this.

Transition from APAC

An initial committee, put in place at the beginning of the project to oversee the transition from APAC and the implementation of NCI, completed its work in January 2008. In particular, research services derived from the National Facility (NF) are now delivered under arrangements overseen by the NCI Steering Committee (NCI/SC). The transfer of these arrangements has been entirely transparent to the research user community. The call for Merit Applications for 2008 (issued in October 2007) was conducted and assessed in the same way as in previous years, with the research projects allocated resources now under NCI governance and management arrangements. Technical projects, tool development and operating system enhancements are also under NCI governance and management.

Governance and Management

Steering Committee / National Planning

The NCI Steering Committee was established during the implementation phase of NCI and, for a period, it operated in parallel with the APAC-NCI Transition Committee. Professor Mark Wainwright was appointed as the independent Chair of the NCI/SC and Professor Jim Williams of ANU served as acting Director from July 2007 – April 2008. An NCI Merit Allocation Committee (MAC) was also established, with Professor Brian Yates from the University of Tasmania, and previously Chair of the former APAC Merit Allocation Committee (MAC), appointed as Chair of the NCI MAC. The current membership of the Merit Allocation Committee is listed in Appendix 1.

Since the establishment of the NCI/SC, representatives of four potential partners, CSIRO, the Bureau of Meteorology, Geoscience Australia, and the research intensive universities—all leading research organisations with a commitment to, and interest, in Australia's high-end computing infrastructure—have been members of the Steering Committee, with the Department of Innovation agreeing to this arrangement until September 30, 2008, or until an organisation becomes a formal partner of NCI (whichever is the earlier). The present membership of the NCI/SC is listed in Appendix 2.

As forecast in the Business Plan, ANU and CSIRO are anchoring NCI, with the total of their planned investments slightly exceeding the annualised contribution of the Commonwealth. The Bureau of Meteorology (BoM) and Geoscience Australia (GA) are fully engaged with NCI in the development of national strategy and policy for the advancement and uptake of high-end computing, with their future contributions to NCI being worked through. It is significant that BoM and ANU/NCI have proceeded with a joint tender for the acquisition of interoperable facilities to support earth systems science research and climate modelling.

The research intensive universities are similarly engaged with the directions being taken by NCI, with there being a strong understanding of the need to substantially upgrade the computational resources in targeted fields in which they have high impact / high profile research. Discussions regarding formal investment and partnership are underway with research intensive universities and we expect a number of these to be favourably concluded in coming months. Of significance is the recent joint announcement by the University of Melbourne and the Victorian Government of the Victorian Life Science Computational Initiative which will be a major addition to the national computational fabric. In the light of this, the NCI has invited the University of Melbourne to participate as an observer in recent Steering Committee meetings.

NCI has been established by the Department of Innovation as both an operational program and also as a source of advice on national policy settings for the advancement of high-end computational infrastructure— with this latter role reflected in the Steering Committee that brings together many of the national research organisations that have a strategic interest in the national computational fabric. Indeed, the need for a national council that brings together publicly funded research organisations, committed to the application of high-end computing in advancing the scope and excellence of their research, is clear.

At recent meetings the NCI/SC has been considering how best to advance the national planning and policy agenda and recommends bifurcating the strategic planning and operational roles of NCI. Specifically, it proposes the establishment of a national council to promote, plan and coordinate the advancement of high-end computing in Australia. All of the publicly funded research organisations that make substantial use of advanced computing infrastructure in the carriage of their research missions, and which make substantial institutional investments in such infrastructure would be invited to join the council and, in doing so, to commit to its national development in a collaborative and cost-sharing manner. This matter is canvassed in greater detail in Section 4 (Governance and Management) which proposes a revision of NCI to become a purely operational program, with its presently attributed planning functions to be ceded to a new national planning council.

The establishment of the envisaged council may take some time and so NCI is concerned that the representation of key research organisations on the present NCI/SC (i.e., some of the potential partners) may be lost, given that their present terms will expire on September 30, 2008, before the necessary changes can be implemented. Accordingly, a revision of the Business Plan is proposed, whereby potential partners continue as members of the Steering Committee until March 31, 2009, the date of submission of the next revision of the Business Plan, which should allow sufficient time for the necessary changes to be engineered.

Director, NCI
NCI commenced operations with Professor Jim Williams serving as Acting Director from July 2007. The position of Director was advertised early in January 2008, with interviews for the position held on 13 March 2008. Following the selection process, an offer was made to, and accepted by, Professor Lindsay Botten (Professor of Applied Mathematics of the University of Technology, Sydney and a former Board member of both ac3 and APAC) who took up his appointment on 26 May 2008.

The NCI Office has been established as an ANU organisational unit, with delegations to operate held with the NCI Director as Head of the Organisational Unit, along with delegations under which the Director can take responsibility for finances falling under the NCI project. The NCI Office also has access to ANU administrative, legal services, and support for maintaining all NCI official files and financial accounts.

Due to the late appointment of the substantive Director, the program has experienced some delays. However, these are now being recovered.

Planning, Policy and Access
Resource Allocation
Substantial work in the development of a Resource Allocation Model (i.e., a shares model), which determines the access shares on the NCI facilities, is underway, with policy documents being considered at the 24 June and 2 September meetings of the NCI/SC. Of critical importance is the need for the resource allocation model to leverage additional partner development, based on the strong nonlinearity of the price performance ratio (see Section 3.2.1).

The principle that underpins this is as follows: through Commonwealth funding alone, only quite a modest facility can be acquired; yet, through the quite nonlinear price-performance characteristic, additional funding from partners can substantially increment the capacity of the system. The NCI/SC has considered some preliminary financial and resource modelling, and has approved in principle, a model whereby partner investments are weighted relative to the Commonwealth funds which provide for Merit and Priority access. As is explained in Section 3.2.1, this decreases the MAS share relative to the partner shares, but on a much larger facility, thus leading to a "win-win" situation. This policy will be formalised in due course once the budget and allocations associated with the new peak facility are clarified.

Merit Allocation
The main Merit Allocation round for 2008 (determined by the MAC in December 2007) received some 73% of the available cycles for distribution in 2008 (12.8 M processor hours). In this round there were 142 applications, of which 29 were associated with new projects. These requested 19.5 M processor hours requested, which is approximately 50% more than was available to the MAS. The Merit Allocation Committee overcommitted by 10%, allocating 13.9 M processor hours, with the largest grant being 1M processor hours.

Normally, a minor MAS round is also conducted in May of each year to pick up new projects which arise in the interim. This year, however, the May round was not advertised due to a shortage of available cycles on the National Facility. (This is referred to more comprehensively in the National Facility Operational Reports at Appendix's 3 and 4.) However, some modest "out of session" requests were considered by the MAC Chair.

The resources available to the MAS for 2009 will be enhanced by the acquisition of the new transition system which, it is anticipated, will boost computational throughput by about 50-60% over that currently served by the present SGI Altix 3700Bx2 system. The proportion available to the MAS for 2009 allocations is to be decided by the NCI/SC at its November 2008 meeting, after consideration of the Resource Allocation Policy.

During the reporting period, the NCI/SC has revised and refined the merit allocation criteria to accord with the present research environment and priorities. In preparation for the December 2008 MAS round, an opportunity has been taken to review and enhance the Merit Allocation Scheme, following the revision of the assessment criteria by the NCI/SC to accord more closely with the new operating environment. The resources distributed by the MAS are presently valued at 50c / processor hour (a figure that reflects the total cost of ownership) and so with 25 applications being granted 240K processor hours or more, there are 25 groups receiving the equivalent of a medium-to-large ARC Discovery Grant in APAC/NCI support.

Accordingly, to ensure that NCI's assessment and allocation processes are rigorous, the MAS has been reviewed and the application processes are being redesigned following discussions with the ARC (Prof Ian Mackinnon, Executive Director, Engineering and

Environment, and Mr Anthony Murfett, Director, Linkage and Infrastructure). Section 3.2.2 outlines the intended revisions to the application processes which places increased emphasis on the quality of the scientific program and the track record of the applicants, enhances the collection of statistical information, facilitates some improvements to the grant accounting processes, whilst retaining the current computational assessment criteria. The ARC has further indicated that it will provide an observer to MAC meetings (at least for the first year or so) to assist with the development of our processes.

Outreach

Some reasonable progress has been made in initiating the Outreach program, the major focus of which is in raising the profile and uptake of high-end computing in Australia, and in making a real difference to the impact and scale of Australian scientific outcomes. To this end, the activities have been focussed on increasing NCI's engagement with key research fields, NCRIS capability areas and major research organisations. A report on activities and visits undertaken is provided in Section 3.3.1.

The transition from APAC to NCI has necessitated the development of a new visual identity and branding. A logo and emblem have been developed and new banners, consistent with the new identity, for use at conferences and meetings have also been developed. New research publicity materials (brochures and posters) from the National Facility with the new branding are also being produced.

A Plone-based website (www.nci.org.au) has been established and the domains nci.org.au and nci.edu.au registered, with these changes also being reflected on the National Facility website (nf.nci.org.au in parallel with the existing URL nf.apac.edu.au). While the website is functional and informative, more work is require to increase its content, attractiveness and utility, particularly through the development of a research gallery which highlights the excellence and impact of the research being undertaken through NCI facilities.

With regard to conferences, NCI's antecedent, APAC, ran a very successful conference (APAC07) in Perth in October 2007. This year, NCI is contributing a workshop to the eResearch Australasia conference, with the Director also participating in an AeRIC eResearch Forum.

Operations and Services

National Facility

The acquisition of the next-generation peak system for the National Facility, the performance of which should be approximately an order of magnitude (i.e., a factor of 10) greater than the current NF, is presently underway. This follows an agreement between the Bureau of Meteorology, ANU and CSIRO to proceed with a joint procurement of new operational facilities for BoM and a replacement of the present national research system (SGI Altix 3700 installed in 2005 and upgraded in 2006).

A joint tender, requiring interoperability between the BoM and ANU systems was released in April 2008 and closed in late May 2008. Since then, the evaluation process, overseen by a Joint Steering Committee of senior representatives of BoM and ANU/NCI, and undertaken by an Evaluation Committee (comprising BoM, ANU and CSIRO representatives) has proceeded according to a rigorous Evaluation Plan, with the aim of determining a single preferred tenderer to provide interoperable facilities to both BoM and ANU/NCI.

The Joint Steering Committee is expecting to receive the Evaluation Committee's final report in early October, at which time it can seek to formulate a recommendation for a single preferred tenderer. At that time, the system and the ancillary site work (i.e., additional power and cooling infrastructure) can be scoped according to the available budget and cash flow. As foreshadowed in the Business Plan, the installation will proceed in stages according to the cash flow, and is on track to commence in mid-to-late 2009.

Development/Transitional System

To meet the immediate need for additional computational capacity prior to the installation of the new peak system, and also to provide the National Facility staff with a transitional system that is characteristic of the next-generation peak system, the NCI/SC has allocated $800K to the purchase of a development / transition system. Tenders were let on 21 June 2008 and closed on 17 July 2008 with responses being received from six vendors. During this process, CSIRO/CAWCR indicated a desire to invest $150K to acquire a dedicated share for development and production using the ACCESS climate simulation. At the time of writing, an order for a system valued at $950K is to be placed before the end of September 2008.

Specialised Facilities

The Specialised Facilities program has been initiated, with a broad call for expressions of interest developed in June 2008 and circulated on 21 July 2008 to the Deputy Vice-Chancellors (Research) of all Australian Universities, major research organisations (CSIRO, Bureau of Meteorology, Geoscience Australia), specialist providers of high-end computing services (VPAC, SAPAC, iVEC, QCIF, ac3/Intersect and TPAC), various research groups/consortia with high-end facilities and key NCRIS capability areas (astronomy, geosciences, characterisation, bioplatforms). Expressions of Interest closed on 20 August 2008, with seven applications received. These are now being evaluated, with a few proposals being considered as a better fit for development under the associated Computational Tools and Techniques Program. Recommendations of the leading expressions of interest, following their finalisation with proponents, will be brought to the NCI/SC for ratification, after which they will be developed, jointly by applicants and NCI, into formal proposals that can be considered for funding by the Department of Innovation.

Computational Tools and Techniques
While guidelines for the CT&T program have been developed (Section 3.1.3), the program has not been formally initiated. This is planned for the fourth quarter of 2008, once the major work for the Specialised Facilities Program has been completed.

3. PROJECT ACTIVITIES

3.1 Operations and Services

3.1.1 National Facility

The NCI National Facility is hosted at the ANU through an agreement under which ANU provides computing services and specialist support for users. The present National Facility consists of a peak computing system – an SGI Altix ('ac') installed in 2005 and a Dell Linux cluster ('lc') installed in 2003. In addition, the NCI has access to large-scale data management facilities through facilities management agreements with ANU.

Usage of the Facilities: July 2007–June 2008

Appendices 3 and 4 present the six monthly National Facility Operations Reports for the periods July–December 2007 and January–June 2008, with further information about the National Facility is available at: http://nf.nci.org.au/ .

Over the course of these two periods, the Merit Allocation Scheme and ANU respectively received approximately a 73 % and a 27 % of the resources. Some 12.8 M hours (73% of the available resources) was available to the Merit Allocation Scheme for the December 2007 round of allocations for 2008, for which some 19.5 M hours had been requested (exceeding that which was available by a factor of approximately 1.5).

Table 1 summarises the Merit Allocations for 2008, according to the (following) explanatory notes.

Table 1: Statistical Summary for the 2008 round of NCI Merit Allocations

Rank	Organisation	SU Req.	% Req.	SU Grant	%Gr.	%CumGr	%Gr/Rq	Gr{#, Ave}.	FundGr{#,%#,%SU}
1	University of Sydney	3376200	17	2564200	17.85	17.85	75.95	{17, 150835}	{12, 71., 96.}
2	University of Queensland	2936000	14.79	1678200	11.69	29.54	57.16	{20, 83910}	{13, 65., 45.}
3	Royal Melbourne Institute of Technology	1702000	8.57	1550000	10.79	40.33	91.07	{8, 193750}	{7, 88., 96.}
4	ANU	1676000	8.44	1256000	8.75	49.08	74.94	{11, 114182}	{10, 91., 98.}
5	Monash University	1990160	10.02	1201550	8.37	57.44	60.37	{17, 70679}	{12, 71., 70.}
6	University of Adelaide	1403400	7.07	1143400	7.96	65.41	81.47	{4, 285850}	{3, 75., 100.}
7	Curtin University of Technology	987000	4.97	893990	6.22	71.63	90.58	{4, 223498}	{3, 75., 72.}
8	University of NSW	1363130	6.86	844550	5.88	77.51	61.96	{15, 56303}	{12, 80., 93.}
9	University of Melbourne	848000	4.27	565000	3.93	81.45	66.63	{8, 70625}	{7, 88., 72.}
10	CSIRO	766000	3.86	495000	3.45	84.89	64.62	{5, 99000}	{5, 100., 100.}
11	University of Technology Sydney	570000	2.87	460000	3.2	88.09	80.7	{3, 153333}	{2, 67., 91.}
12	University of Newcastle	439200	2.21	290200	2.02	90.12	66.07	{4, 72550}	{2, 50., 93.}
13	University of Tasmania	345810	1.74	273050	1.9	92.02	78.96	{5, 54610}	{4, 80., 85.}
14	University of Western Australia	304000	1.53	256000	1.78	93.8	84.21	{2, 128000}	{2, 100., 100.}
15	Swinburne University of Technology	183000	0.92	171000	1.19	94.99	93.44	{3, 57000}	{2, 67., 16.}
16	Latrobe University	284000	1.43	164000	1.14	96.13	57.75	{3, 54667}	{2, 67., 85.}
17	Murdoch University	240000	1.21	160000	1.11	97.25	66.67	{1, 160000}	{1, 100., 100.}
18	Macquarie University	152520	0.77	152520	1.06	98.31	100	{2, 76260}	{1, 50., 24.}
19	James Cook University	40000	0.2	60000	0.42	98.73	150	{1, 60000}	{0, 0., 0.}
20	ADFA	63200	0.32	55200	0.38	99.11	87.34	{5, 11040}	{3, 60., 60.}
21	Griffith University	40000	0.2	36000	0.25	99.36	90	{2, 18000}	{2, 100., 100.}
22	University of Wollongong	44000	0.22	32000	0.22	99.58	72.73	{2, 16000}	{2, 100., 100.}
23	Queensland University of Technology	40000	0.2	20000	0.14	99.72	50	{1, 20000}	{1, 100., 100.}
24	University of South Australia	15005	0.08	15005	0.1	99.83	100	{1, 15005}	{0, 0., 0.}
25	Deakin University	38800	0.2	14800	0.1	99.93	38.14	{2, 7400}	{1, 50., 46.}
26	University of Southern Queensland	10000	0.05	10000	0.07	100	100	{1, 10000}	{1, 100., 100.}
27	External Government	0	0	0	0	100	Indet	{0, Indet}	{0, Indet, Indet}
28	ANSTO	0	0	0	0	100	Indet	{0, Indet}	{0, Indet, Indet}
	Totals	19857425	100	14361665	100		72	{147, 97698}	{110, 75., 83.02}

(Indet denotes an indeterminate calculation.)

Notes

The columns of the table above are as follows:

- SU Req—number of system units (cpu hours) requested in all projects,
- % Req—the proportion of requests from each university,
- SU Grant - number of SUs granted for each university,
- % Gr—Proportion of granted SU for each university,
- % CumGr—Cumulative percentage for grants,
- % Gr/ Rq—percentage of SUs granted relative to SUs requested for each organisation,
- Gr (#, Ave)—number of grants for each organisation and their average size (in SU),
- FundGr (#, %#, %SU)—the number of grants acknowledging research funding (ARC, industry, other), the proportion of grants with funding, and the proportion of SUs granted with research funding acknowledged.

Table 2 complements the summary of Table 1 by showing the total allocations for 2008, i.e., both MAS and Partner shares. The key difference between the two is the very large partner share of ANU, and the very modest access to partner shares by all other organisations.

Table 2: Statistical Summary for the Total Allocations for 2008 (including MAS and Partner Shares)

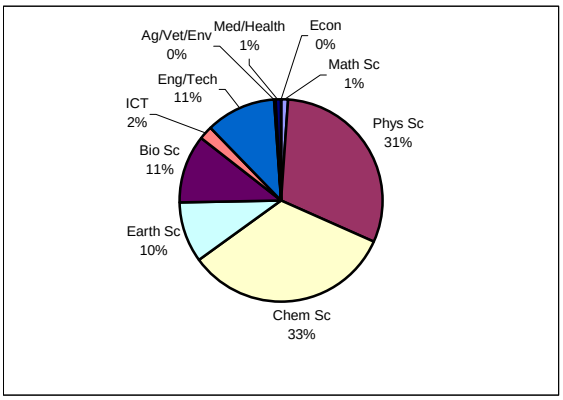
Rank	Organisation	SU Req.	% Req.	SU Grant	%Gr.	%CumGr	%Gr/Rq	Gr{#, Ave}.	FundGr{#,%#,%SU}
1	ANU	7774840	29.59	6201115	31.64	31.64	79.76	{57, 108791}	{42, 74., 81.}
2	University of Sydney	3376200	12.85	2564200	13.08	44.72	75.95	{17, 150835}	{12, 71., 96.}
3	University of Queensland	2960700	11.27	1699000	8.67	53.39	57.39	{25, 67960}	{17, 68., 45.}
4	Royal Melbourne Institute of Technology	1702000	6.48	1550000	7.91	61.3	91.07	{8, 193750}	{7, 88., 96.}
5	Monash University	1990160	7.57	1201550	6.13	67.43	60.37	{17, 70679}	{12, 71., 70.}
6	University of Adelaide	1403400	5.34	1143400	5.83	73.27	81.47	{4, 285850}	{3, 75., 100.}
7	Curtin University of Technology	1016300	3.87	923290	4.71	77.98	90.85	{9, 102588}	{3, 33., 70.}
8	University of NSW	1366130	5.2	847250	4.32	82.3	62.02	{16, 52953}	{12, 75., 92.}
9	CSIRO	904040	3.44	611354	3.12	85.42	67.62	{17, 35962}	{10, 59., 91.}
10	University of Melbourne	848000	3.23	565000	2.88	88.3	66.63	{8, 70625}	{7, 88., 72.}
11	University of Technology Sydney	570000	2.17	460000	2.35	90.65	80.7	{3, 153333}	{2, 67., 91.}
12	University of Tasmania	381810	1.45	312790	1.6	92.25	81.92	{6, 52132}	{4, 67., 75.}
13	University of Newcastle	439200	1.67	290200	1.48	93.73	66.07	{4, 72550}	{2, 50., 93.}
14	University of Western Australia	305700	1.16	257700	1.31	95.04	84.3	{4, 64425}	{4, 100., 100.}
15	Swinburne University of Technology	183000	0.7	171000	0.87	95.92	93.44	{3, 57000}	{2, 67., 16.}
16	Latrobe University	284000	1.08	164000	0.84	96.75	57.75	{3, 54667}	{2, 67., 85.}
17	Murdoch University	240000	0.91	160000	0.82	97.57	66.67	{1, 160000}	{1, 100., 100.}
18	Macquarie University	152520	0.58	152520	0.78	98.35	100	{2, 76260}	{1, 50., 24.}
19	James Cook University	40000	0.15	60000	0.31	98.65	150	{1, 60000}	{0, 0., 0.}
20	Queensland University of Technology	83000	0.32	55050	0.28	98.93	66.33	{3, 18350}	{1, 33., 36.}
21	ADFA	63200	0.24	55200	0.28	99.22	87.34	{5, 11040}	{3, 60., 60.}
22	External Government	40000	0.15	40000	0.2	99.42	100	{1, 40000}	{1, 100., 100.}
23	Griffith University	40000	0.15	36000	0.18	99.6	90	{2, 18000}	{2, 100., 100.}
24	University of Wollongong	44000	0.17	32000	0.16	99.77	72.73	{2, 16000}	{2, 100., 100.}
25	University of South Australia	15005	0.06	15005	0.08	99.84	100	{1, 15005}	{0, 0., 0.}
26	Deakin University	38800	0.15	14800	0.08	99.92	38.14	{2, 7400}	{1, 50., 46.}
27	University of Southern Queensland	10000	0.04	10000	0.05	99.97	100	{1, 10000}	{1, 100., 100.}
28	ANSTO	6000	0.02	6000	0.03	100	100	{1, 6000}	{0, 0., 0.}
	Totals	26278005	100	19598424	100		75	{223, 87885}	{154, 69., 80.75}

Table 3 and Figure 1 characterise the total usage on the National Facility by Research Field. Historically, usage of the National Facility has been dominated by the fields of Physics and Chemistry and this trend continues. Earth systems science / climate modelling appears under RCFD code 26 (Earth Sciences) and to date has been only a modest part of NF usage; this, however, is expected to change dramatically given the current priorities and CSIRO's intention to acquire a share of the National Facility, a large fraction (about two-thirds) of which will be used for climate modelling. This is likely to be further augmented by the possible partnership investment by a number of leading research universities, some of which are targeting additional resources for their climate research groups.

Table 3 Total Usage of the NF by Research Field

Field ID	Research Field	SU granted	%
23	Mathematical Sciences	205,314	1.05%
24	Physical Sciences	6,005,022	30.64%
25	Chemical Sciences	6,503,552	33.18%
26	Earth Sciences	1,934,557	9.87%
27	Biological Sciences	2,105,403	10.74%
28	ICT Sciences	431,826	2.20%
29	Engineering and Technology	2,212,907	11.29%
30	Ag, Vet and Env Sciences	29,183	0.15%
32	Medical and Health Sciences	159,561	0.81%
34	Economics	11,100	0.06%
Totals		19,598,425	100.00%

Fig 1 Total usage of the NF by Research Field



Acquisition of the next-generation National Facility

As detailed in the Business Plan, the central focus for developing the National Facility over the period 2008-09 is to increase the performance of the peak system by approximately an order of magnitude (i.e., a factor of ten).

High-level discussions in March 2008 involving ANU, BoM, CSIRO, and NCI led to a decision to put forward a joint procurement for high-end computing facilities to provide for an integrated modelling environment for the Centre of Australian Weather and Climate Research (CAWCR) and the Australian Community Climate and Earth System Simulator (ACCESS) user communities. In this process, the Bureau of Meteorology is refreshing its Operational System based at its Melbourne headquarters (replacing the present NEC SX-6 facility) while the ANU is refreshing the NCI National Facility (NF) that serves as Australia's peak computational research resource.

An agreed objective under the framework is to develop the National Facility so that it is capable of inter-operating with BoM's operational facility, creating a computational environment for the national earth systems science community that spans research, development and operations. This objective is tied to achieving a far higher degree of functionality and collaboration than would otherwise be the case. Although this earth systems science focus is having an major impact on the NF acquisition process, arrangements are in place to ensure that outcomes also benefit other research areas supported under the NCI funding Agreement.

The procurement process is being led by the Bureau of Meteorology under the Commonwealth Government Information Technology and Communications v4 Contractual Framework. Following the preparation of a comprehensive tender document, a Request for Tender was released on 3 April 2008 through AusTender. At the close of tenders on 29 May 2008, four responses had been received from leading vendors of high-end computing equipment. Following screening of the tenders for completeness by BoM's Contracts and Consultancy Unit, the detailed evaluation of the tenders commenced according to a comprehensive Evaluation Plan (covering technical, financial, contractual and legal matters) that had been developed by staff of BoM, ANU and NCI.

The evaluation process is overseen by a Joint Steering Committee of eight members, comprising equal representation from BoM (Phil Tannenbaum - Chair, Neville Smith, Tom Keenan, John de la Lande) and ANU/NCI (Robin Stanton, Doug McEachern, Alex Zelinsky and Lindsay Botten). Its role is to formulate a recommendation, based on a report from the Evaluation Committee, of a single preferred tenderer that can meet the needs of both BoM and ANU/NCI and provide for the required interoperability of the ACCESS model on both systems. The detailed evaluation of tenders is performed by a four member Evaluation Committee (Tim Pugh (BoM) - Chair, Ben Evans (ANU) - Deputy Chair, Marcel van Dyk (BoM) - Financial and Peter Shipp (ANU) - Financial) assisted by a substantial panel of advisors with technical, financial and legal expertise.

During July and August 2008, the Evaluation Committee has undertaken an evaluation of the four tenders and has produced a shortlist of three, with one vendor being excluded on technical/environmental grounds. The tenders from the three shortlisted vendors are presently being evaluated in a detailed manner against capability, feasibility, affordability and value criteria, with the final Evaluation Report anticipated in early October, 2008.

The new National Facility will be developed in stages—a consequence of the cash flow available to NCI—with the majority of the system to be installed in mid to late 2009. Once the choice of system is finalised, it will be possible to scope the system and the associated power and cooling infrastructure which is to be installed in the Leonard Huxley Building on the ANU Campus. While there are solid commitments from NCRIS, ANU and CSIRO, the additional funds that may be available from other potential partners is still uncertain. While all of the systems that might be acquired are highly scalable, NCI needs to estimate the capacity of ancillary (power and cooling) infrastructure to accommodate likely expansion of the facility as new partners join. Given the uncertainties that exist at present, it is not possible to present a revised Budget at this time. As soon as this is available, a revised Business Plan will be presented to the Department of Innovation.

Development/Transitional System

At its April 2008 meeting, the NCI/SC approved the acquisition of a development/transition system, allocating $800K to this purchase as foreshadowed in the Business Plan. Such a system is needed for two reasons:

- to provide much needed additional computational resources to address increased user demand associated with the Merit Allocation Scheme, and to provide additional resources to enable access to the National Facility by new partners and new research communities.
- to establish a system with similar characteristics to the next generation peak facility, allowing the National Facility staff the opportunity to acquire skills and expertise in configuring, managing and operating a large scale production cluster with an Infiniband interconnect and a Lustre filesystem.

With regard to the former, it is relevant that present workloads on the National Facility are very high, with the ratio of queued processors to available processors averaging at 3.7 during a quarter, and rising to a ratio of 5 at the end of the quarter. There is now a real shortage of processor cycles and hence the installation of this transitional system is urgently awaited.

A request for tender for this system was released on 21 June 2008, with tenders closing on 17 July 2008. Tenders from six vendors have been evaluated according to an evaluation plan which emphasised high computational performance, balance and uniformity of system performance, and reliability, robustness and manageability of the system (to be demonstrated during acceptance testing).

CSIRO/CAWCR has indicated a commitment to invest $150K in this system for a period of one year (prior to the installation of the new peak facility) in order to provide the Centre with a much needed increase in machine time to develop and test the ACCESS model, in preparation for the computational experiments that Australia is obliged to contribute to the 5th Assessment of the Inter-Governmental Panel of Climate Change, due in 2010. At the time of writing this report, tender evaluations were complete and an order for the systems valued at $950K was scheduled to be placed before the end of September 2008.

Data Storage

The National Facility makes substantial use of the high-end data infrastructure that is owned and operated by the ANU—for which NCI acquires access through payment of a facilities management fee identified as a line item in the Business Plan. Use of these resources arises from computational projects supported by the merit allocation scheme (MAS-C), and also the data projects merit allocation scheme (MAS-D), initiated previously by APAC. The latter scheme (see Section 3.2.2) is being discontinued given that alternative opportunities for data project support exist under the NCRIS environment, although existing users will be provided with support from NCI for another year to allow for their transition.

The real growth in data requirements is from the computational (MAS-C) projects, with storage requests growing exponentially with time (as can be seen from the National Facility report for January-June 2008). In this regard, any reference to the former MAS-D scheme is a distraction to the key point that NCI, unlike major computational facilities internationally, cannot directly acquire mass data storage infrastructure, but instead must rely on its host, ANU, for the provision of this infrastructure.

The current and anticipated growth in data requirements associated with computational projects places an unreasonable expectation on the ANU, and is a situation that is rapidly becoming untenable. The requirement that research organisations should provide for the storage needs of their researchers, together with no provision being made for the acquisition of storage infrastructure within NCRIS Platforms for Collaboration, overlooks the reality of operating a national peak computing facility.

ANU and NCI believe this matter needs to be resolved and thus look forward to discussions with the Department of Innovation in the course of the review of this Progress Report.

3.1.2 Specialised Facilities

In addition to the NCI National Facility, high-end computational research in Australia is supported by a variety of regional and discipline/application specific installations, a number of which provide services tailored to the needs of particular research communities. Through the Specialised Facilities (SF) program, NCI seeks to extend the range of high-end services available through its Merit Allocation Scheme by acquiring access, through subcontract agreements, to specialised facilities and services in particular application/discipline areas for which there exists a demonstrated need.

The NCI/SC at its 24 June 2008 meeting authorised the NCI Director to proceed with a broad call for Expressions of Interest (EoI) in the Specialised Facilities Program. A document describing the program and explaining the call for expressions of interest was developed by the Director and was circulated to NCI/SC members on 16 July 2008 for comment. Following minor amendments, the EoI call was distributed on 21 July 2008 (with a closing date of 20 August 2008) to the Deputy Vice-Chancellors (Research) of all Australian Universities, major research organisations (CSIRO, Bureau of Meteorology, Geoscience Australia), specialist providers of high-end computing services (VPAC, SAPAC, iVEC, QCIF, ac3/Intersect and TPAC), various research groups/consortia with high-end facilities and key NCRIS capability areas (astronomy, geosciences, characterisation, bioplatforms).

The call for expressions of interest focussed on organisations offering specialty computational services that were interested in enhancing these and extending their availability to the national research community, facilitated by investment from NCI. The business model adopted by NCI for this program is to acquire a fixed share of the resources of a specialised facility that could be distributed by NCI through its merit allocation process. This is in contrast to the more complex and unwieldy model referred to the Funding Agreement (and Business Plan) that relied on mutual leveraging of funds and in-kind contributions by NCI and the SF provider.

The NCI budget has allocated $3M to this program, with substantial additional funds also being available from CSIRO for projects in which it has specific needs. The EoI call, which indicated that up to $1.5M would be available for each Specialised Facility (up to mid-2011), emphasised that "services" referred to a package of computational resources, access to a range of specialist software, and the accompanying expert support that provides for the computational needs of a particular research community. In general, NCI did not seek to invest in raw machine cycles, unless there was some persuasive reason (e.g., a highly specialised machine architecture of importance to researchers) to do so.

At the close of applications, seven applications had been received, six associated with application areas and NCRIS capabilities covering the fields of bioinformatics, earth system science, geosciences, astronomy, characterisation (imaging and visualisation), and seventh proposing the development of special purpose hardware (based on graphics processing units). Of these, two applications, which sought user and software development support (rather than specialised facilities) for particular disciplines, and potentially a third, will be considered under the computational tools and techniques program.

At the time of writing, the applications are under review by a small evaluation panel according to the selection criteria:

- the extent of the demand for the facility by 'high-end' users in the Australian research community, including the NCRIS capability areas,
- the access to the specialised facility and user support that will be provided for researchers Australia-wide,
- the characteristics of the specialised facility and the manner in which it complements the capabilities of the National Facility,
- the proposed investment in the specialised facility represents good value in comparison with other possible investments,
- the host organisation is capable of providing the required level of service to users as a component of the 'national advanced computing infrastructure',
- the host organisation is able to implement policies, system scheduling and processes to accommodate projects granted resources under the NCI Merit Allocation Scheme.

The evaluation panel will report to the NCI/SC recommending which applications should proceed through to the stage of a formal proposal for presentation to the Department of Innovation. Those applications which are progressed to formal proposals must meet the following criteria for funding to be approved:

- acceptability to the Department of Innovation;
- an appropriate level of co-funding;
- established through contractual agreements linking accountability to NCI;
- arrangements for the allocation of resources to be made through the NCI-MAC;
- through NCI, the host of a SF may operate a shares based model to allow other investments to take place through the NCI program, similar to that of the NF;
- hosts of SF will provide the NCI Office with adequate accounting and reporting for the resources funded and will provide the NCI Office with appropriate audited accounts relating to the funding received, and report on and assist reviews of the performance and reliability of the facility and its services.

3.1.3 Computational Tools and Techniques Program

The Computational Tools and Techniques (CT&T) program is intended to support the development of software tools to improve particular applications or to provide targeted user support for particular user communities (e.g., astronomy, geosciences, earth system science/climate modelling, computational chemistry). With the formation of National eResearch Architecture Taskforce, and the growing internationalisation of the research software development effort, it also makes sense for the CT&T program to broadly support one or more specific user communities in their use of NCI facilities.

As for the Specialised Facilities Program, the primary driver for the establishment of a CT&T project is the need and demand from particular research communities. Of relevance is that the recent call for expressions of interest in the Specialised Facilities Program has identified two, and perhaps three, areas in which significant support might be provided to particular research communities (earth system science/climate modelling, astronomy, and geosciences).

The criteria for supporting a project are:

- the extent of the user demand for the tools and techniques,
- the extent to which it improves computational research for a targeted user community,
- the level of co-investment by that community,

- the track record of the key people in the project team and their relationship with the relevant user groups.

In view of the modest budget available to support the CT&T program, it is likely that the call for expressions of interest in CT&T in the fourth quarter of 2008 will be restricted and will take advantage of the needs that have emerged through the broad EoI call for Specialised Facilities.

3.2 Planning, Access and Policy

3.2.1 Planning and Policy

Resource Allocation

The importance of leveraging additional partner investment in the acquisition of high-end computing resourced is discussed and emphasised in a section headed "Partner Access" (see Section 3.2.2). The key to this is the extreme nonlinearity in the price performance ratio, in which modest proportional increments in the investment can lead to significant changes in the aggregate performance of a system. The Commonwealth funding alone can acquire only a relatively modest facility and so it is only is the additional investment from partners that can substantially upscale a facility for the benefit of all. Partner investment must thus be encouraged by leveraging the government's investment.

In recent months, NCI has been developing a resource allocation model, the general principles of which have been approved by the Steering Committee. In this, the shares are determined pro rata according to their investment, but with the contributions of each partner being multiplied by a scaling factor, e.g. 1.2. In this example, for every $1 provided, the partner receives $1.20 of services relative to the Merit Allocation Scheme that is funded by NCRIS money. The effect of this is to increase the partners' proportional share relative to the merit allocation scheme, but on a much larger system. In this way both the MAS and the partners benefit, but with the partner share doing proportionately better.

At this time it is not possible to present the details of the resource allocation modelling since the data on which this is based has been abstracted from the current tenders for the peak facility that are commercial-in-confidence. However, the following example may serve to illustrate the point. Consider a base level system in which our resource allocation model yields around 46% of the facility to the MAS and 54% collectively to the partners, with the actual time allocations (in this example) being 53M processor hours and 63M processor hours annually respectively. By attracting additional resources from partners sufficient to double the capacity of the machine, through a scaling factor of 1.2 to partner investments (i.e. a 20% loading on the value of partner shares relative to the MAS), the proportional shares available to MAS and partners are roughly 36% and 54% respectively, yielding approximately 80M hours for MAS and 141M hours for the partners. It is the extreme nonlinearity of the total investment-performance curve that allows this win-win situation to occur.

While the model is sound, there are many unknowns until the decision for the new peak system is taken and all of the direct and indirect costs can be determined. Accordingly, it will not be until the next substantive revision of the budget in the Business Plan that the resource allocation model can be transparently presented.

3.2.2 Access

Access to NCI resources is provided in two ways:

- Merit and Priority Access which is funded by Commonwealth resources provided through NCRIS,
- Partner Allocations, in which NCI partners and affiliates acquire resources (respectively a defined share or a fixed quantity of resources).

These are dealt with under separate headings below.

Merit and Priority Allocation

Background

Merit allocation is the direct successor to the long standing Merit Allocation Scheme (MAS) established under NCI's antecedent, APAC, while Priority Allocation has a direction association with the needs of national research priorities, including those supported through the NCRIS Capabilities. The NCI/SC is firmly of the view that all Commonwealth funded resource allocations must demonstrate appropriate research merit, irrespective of whether they are handled under the "Merit" or "Priority" allocation headings.

In the case of the Merit Allocation Scheme, there are, in general, two annual calls for application—a major call in October of each year, with allocation announced in December for the following year, and a minor call in April/May with allocations for the following six months announced in June. In 2008, however, due to the shortage of resources available, with the Merit Allocation Scheme effectively over allocated in December 2007, there was no call for applications in April/May. Applications are assessed by a Merit Allocation Committee established by the Steering Committee under principles ratified by the NCI/SC.

With regard to Priority Allocation, NCI will reserve (from December 2008 onwards) a fraction of the resources under the MAS for allocation to high priority projects that may come from national priority areas and which cannot be delayed until the two annual merit allocation rounds. This should provide a significant allocation of resources that will enable new projects to be commenced, after which these projects will need to acquire resources either through the MAS, or through additional partner investment, or a combination of both.

It needs to be recognised that the quantity of resources available under the MAS can never be sufficient to provide for the needs of leading groups working in internationally competitive research fields in which high-end computing is an essential tool. In such fields, additional partner investments are a necessity and NCI is strongly promoting this way forward.

As is alluded to elsewhere in this report, ANU has been a long-term investor in high-end computation and data in order to advance the excellence of its research profile in key areas, while CSIRO has recently decided to subscribe to NCI for similar reasons, and particularly to increase support for climate modelling and earth system science. With regard to CSIRO, it has indicated that it will acquire its resources through its "partner share" and has no intention of requiring its researchers to compete for resources through the NCI MAS, given the existence of rigorous processes of research review internally with organisation.

At the time of writing, a number of leading research universities is also contemplating investment in NCI for strategic reasons such as boosting the capability available to particular high profile/high impact research groups. In national priorities such as Climate Change, separate overtures to the Commonwealth Government, supported by NCI, are being made by various sectors of the earth systems science research community for additional resources (both infrastructure and support/expertise) to be invested in NCI.

The two Merit Allocation Schemes (MAS-C and MAS-D) inherited from APAC have been in operation during 2008. The former, and by far the larger is MAS-C which refers to the computational processing resources and any associated data storage requirements of the project. The latter (MAS-D), established by APAC during its second phase of operations (2004-06), anticipated the growing need for storage and accessibility for nationally significant data collections. Under the NCRIS environment, with the responsibility for data that is not associated with high performance computation (either generated or processed) being attributed to some combination of research institutions, ARCS and ANDS, it is no longer appropriate for NCI to continue with the MAS-D program.

Accordingly, from October 2008 onwards, the existence of the MAS-D scheme will no longer be promoted, with the MAS-C, renamed as MAS, being the only scheme under which applications may be sought. The list of projects supported by MAS-D is not large, and some projects are already receiving support under other schemes (e.g., NeAT). Accordingly, it is intended that NCI National Facility staff write to all holders of currently active MAS-D grants explaining the situation and evolving a strategy for their ongoing support beyond 2009. During 2009, NCI will continue to provide support for existing projects not supported elsewhere at the time, but will withdraw further support from the end of 2009.

Merit Allocation Scheme (MAS)

Under MAS, public sector researchers and research groups may apply for computational time (and associated data storage) in October (main round) and May (supplementary round) of each year. The MAS also accepts applications out-of-session, with this now being enhanced and formalised as Priority Access.

Applications are considered by the Merit Allocation Committee (MAC), the members of which have expertise in both computational methods and their respective research fields that span the range of applications / disciplines handled by the National Facility. The Committee is serviced by staff of the NCI National Facility, with the NF Manager being an ex-officio member. In addition to the spread of disciplines, NCI ensures that the membership is also geographically distributed so that researchers from each state/territory can easily access expertise and experience of the MAC processes.

The present membership of the MAC is listed at Appendix 1. The Merit Allocation Committee Chair, Professor Brian Yates, was confirmed in the position by the NCI/SC and approved by the Department of Innovation on 16 January 2008. The NCI/SC also agreed to a rotation of the MAC membership. This will commence with the longest serving members being replaced in 2009 for the 2010 rounds, and of the augmentation of the discipline expertise with the inclusion of earth systems science/climate modelling which is a growing area of demand for the service.

Applications for the MAS are assessed against the following criteria which have been refined during 2008 by the NCI/SC to fit in with the current environment under NCRIS. They are as follows:

Research quality

- *Research merit including the potential of the work to generate new knowledge in an important area, the comparative scientific merits of the work within its discipline, originality and international competitiveness.*
- *National benefit and research priorities.*
- *Experience and demonstrated research capacity of the applicant and the project team.*

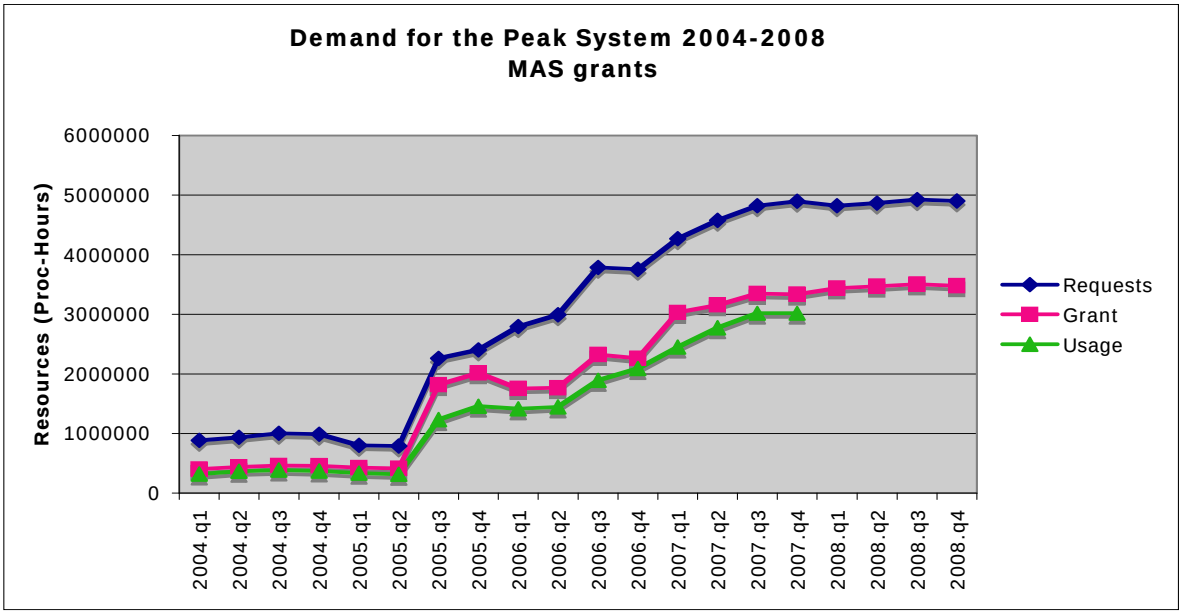
Appropriateness of the NCI computational resources
- *The need for such resources to conduct the research.*
- *Suitability of the system (hardware and software) and its operational environment to support the project.*
- *Evidence or experience to demonstrate that the project will use the facilities efficiently.*

Reasonableness of the level of resources requested
- *Relative to the total amount available.*
- *Needed to make adequate progress in the proposed research program.*

Track record of the applicant in using the NCI computational resources where relevant
- *In the case of proposals to continue a project, the record of achievement and efficient use of previous allocations.*

Highlights of the Merit Allocation round for 2008 resources include:
- The Merit Allocation Scheme (MAS) received 73% of available cycles for 2008 (12.8M processor hours)
- There were 142 applications, of which 29 were new projects,
- 19.5 M processor hours requested, which is 1.5 times the available,
- 13.9M processor hours allocated, which is 1.1 times the available,
- The largest grant was 1M processor hours.

The historical demand for NF resources continues to exceed supply, as shown in the following graph.



New Merit Allocation procedures for 2009 onwards

In preparation for the call for applications for the 2009 round of Merit Application (December 2008), the opportunity to refine the application and assessment processes is being undertaken. This is being done to enhance the rigour and the quality of our processes, given the magnitude of the grants awarded. At the December 2007 round of Merit Allocation (for 2008), the MAC distributed some 12.8M hours of processor time — the equivalent of $6.4M (at 50c / processor) hour, an amount equivalent to about one-tenth of that distributed annually by each of the ARC Discovery Grant panels. In this round, there were some 25 applications that were granted at least 240,000 processor hours, with the largest being 1,000,000 hours. Thus, there were 25 applications which received an allocation of at least $120K annually — the equivalent of a medium to large ARC Discovery grant.

Accordingly, NCI and National Facility staff are implementing a revised application process (and associated online form) outlined below, that increases the rigour of the application and assessment, places greater emphasis on the quality of research being supported, improves the alignment of the application form and the data collected with the assessment criteria, and increases NCI's engagement with the Australian Research Council.

These changes have been described in a paper from the NCI Director and MAC Chair (circulated amongst the present members of the MAC) and have also been discussed with the ARC (Professor Margaret Sheil, CEO, ARC, Prof Ian Mackinnon, Executive Director, Engineering and Environment, and Mr Anthony Murfett, Director, Linkage and Infrastructure). The ARC has indicated its willingness to assist NCI in these changes, agreeing to provide an observer at the Merit Allocation Committee meetings, and also indicating its

willingness and interest in contributing to national strategic planning discussions for high-end computing. Corresponding discussions with the NHMRC are also planned.

The revised application form outlined below, is consistent with the material provided by researchers when preparing ARC applications and so the revised processes should place little additional burden on applicants. The new form also allows NCI to accommodate projects in which the researcher team is distributed across multiple organisations (e.g., CRCs, Centres of Excellence) and so encourages block applications, rather than the multiple smaller applications that we see in some cases at present (and which may exhibit "double dipping").

A. *Administrative Summary*
   1. Project Title;
   2. Chief Investigators (including titles, positions, contact details) with the understanding that the first named CI is the project leader (and leader of the research group) and the one responsible for reporting etc.;
   3. Organisations—with percentages which will allow us to assign grants from groups with members in multiple universities amongst the multiple institutions, e.g. as in the in the case of distributed research centres /organisations which may apply as a block;
   4. Project Summary (100-150 words);
   5. Research Classifications (maximum of 5 with percentages, ARC records only 3);
      a. RFCD / FOR codes,
      b. SEO codes,
   6. Researchers attached to this project (as for now).

B. *Publication and Research Funding Data*
   For each listed CI, the following should be provided:
   1. Research Publications—For each CI, refereed publications for a period of 5 years classified under headings of Book, Book Chapters, Refereed Journal articles, Refereed Conference publications etc. This more ort less preserves the status quo and is precisely what the ARC requires. Publications relevant to the application should be indicated by (*) ;
   2. Research Funding (all agencies)—Summary of all research funding granted in the past five years, with each item including all named investigators on the proposal, project title, Funding Agency, Funding Scheme, years, amounts (similar to the summary required on ARC LIEF grants);
   3. Awards and Prizes.

C. *Project Description*
   All projects must provide the following information which has been structured so as to address the selection criteria:
   1. Research Proposal;
      a. Significance and Impact of the Research (around 2 pages),
         i. providing background to the research and highlighting the significance and innovation of proposed project,
         ii. referring to the list of research grants to be directly supported by this proposal (with (i) and (ii) allowing us to assess the "research merit including the potential of the work to generate new knowledge in an important area, the comparative merits of the work within its discipline, originality and international competitiveness" (from the selection criteria)),
      b. Experience and Capacity of the Research Team—comprising a statement of the most significant contributions and impact to the research field and of their computational expertise and experience - for each member of the CI team (1/2 page maximum per CI),
      c. Relationship to National Priorities (ticking a box),
         i. National Research Priorities and Priority Goals (as for the ARC),
         ii. Relationship to NCRIS Capabilities.
      d. National Benefit of the Research — a brief statement of around 100--150 words,
   2. Computational Resources and Justification of Needs;
      a. Description of the computational methods and algorithms used,
      b. Outline of why access to the National Facility is critical to the proposed research, e.g. parallel architecture, access to software, large memory etc. If the tasks involve truly distributed computation, then include a table characterising computational scaling should be provided (as is presently the case),
      c. Resource requested by quarter
         i. CPU resources (in SU — total amount and also the amount necessary for adequate progress),
         ii. Data storage.
      d. Justification of the total resources requested — explaining needs from a technical viewpoint and also in the context of the research objectives,
         i. CPU resources,
         ii. Data storage.

e. Progress report (1/2-1page),
   i. Highlighting research progress facilitated by previous grant(s), any difficulties experienced, any issues to be raised with NF staff etc,
   ii. Previous usage (from NF data) — inclusion of a table of previous usage and a text box allowing applicants to reflect on this, highlighting any issues/problems that have occurred.

Partner Access

Background

Partner access provides the mechanism by which organisations may acquire dedicated access to the facilities at full operating cost (that includes the cost of the system and upgrades, necessary ancillary infrastructure, facilities management fees, and costs associated with software, network, and system and user support).

Within the Business Plan, the NCI Steering Committee refined the definitions of *partner* and *affiliate* as follows:

- A *partner* must contribute at least an indicative $1M per annum over the life of the NCI program (i.e., until 30 June 2011) for which they receive a share, i.e., a fractional allocation of NCI resources,
- An *affiliate* is a co-investor which does not make this level of investment and, instead, receives a fixed resource allocation for the year in which they invest.

The pricing of resources is set by calculating the lifetime total cost of the system and dividing by the resources the system can deliver during that period. For the present peak facility (SGI Altix) this figure is approximately $0.50 per processor hour, with this being the rate at which access for publicly-funded research institutions is charged.

It is generally accepted that Australia can provide an appropriate level of high-end computing capability, deemed roughly equivalent to US Track 2 facilities, to its research communities only through the co-investment of a number of partners, including Government. It is vital that partnership, which implies long-term (rather than transient) investment, must be encouraged and so the NCI Director and the Steering Committee have been considering this issue closely, resolving on a number of broad principles that underpin the evolving Resource Allocation (Shares) Model described in a previous section (Policy Development).

Given the nonlinearity of the value equation in acquiring and installing high-end facilities, the Commonwealth's contribution alone can acquire only a modest facility. However, as is typical of many contemporary systems, an additional cash investment of around 50 percent over a set base price may, in fact, double the scale and computational performance of a system. This nonlinearity is further compounded by the underlying fixed or quasi-fixed costs (power and cooling infrastructure, staff costs etc), with only the power consumption scaling linearly with processing capacity. Accordingly, as is evident from the current round of tenders for the new peak facility, an additional increment of several million dollars over a projected three year period (2009-11) in a budget of $40-50M may serve to double the performance of the facility.

It is this that argues most persuasively that it is vital to encourage partner investment. To promote this, NCI has been active in working with prospective partners, particularly since the appointment of the substantive Director. Visits and presentations have been made to the Deputy Vice-Chancellors (Research) of the Group of Eight Universities and to CSIRO, the Bureau of Meteorology, Geoscience Australia, Monash University, the University of Sydney and the University of Melbourne, with more visits planned.

Thus far, NCI has commitments of:

- $3.4M p.a. from ANU ($1M in cash and $2.4M in kind for staff) for the four years 2007-11,
- $3.3M p.a. from CSIRO ($3M in cash and $0.3M in kind for staff) for three years 2008-11.

Contracts formalising these arrangements are under development. This has been delayed, however, until the decision for the new peak facility is taken and the resource allocation model can be finalised (thus determining the actual machine shares available to each organisation).

It is significant, however, that the sum of these two contributions ($6.7M) slightly exceeds the annualised contribution ($6.5M) from the Commonwealth.

NCI is aiming to attract at least $1.5M from research intensive universities and visits to universities thus far, and those planned for the future, are focussed on this outcome. A decision is expected from three major universities by the end of 2008, with the aim of signing partnership agreements effective from 1 January 2009. The discussions thus far have focussed on individual investments of approximately $0.5M p.a. Under the present definitions, however, each university would be classed as an affiliate, while collectively they might be thought of as constituting a partner. The thinking of the NCI/SC is that it is prudent to adopt a generous and long-term view, and, in doing so, to relax the $1M p.a. criterion.

Given that NCI wishes to encourage investment by research universities, the definition of partnership is evolving into one of long-term commitment, rather than be constrained by a particular number.

Accordingly, a revised definition might be that:

- A *partner* commits to a fixed level of investment over the lifetime of the NCI program (i.e., until 30 June 2011) for which they receive a share, i.e., a fractional allocation of NCI resources,
- An *affiliate* is a co-investor that receives a fixed resource allocation for the period in which they invest.

Partners, having bought in for the long haul (from 2009-11), benefit through the upscaling of the facility with time, mainly in the latter phases of the NCI program, through reductions in the effective cost per processor hour as the machine capacity is incremented.

In the current context, we anticipate the cost per processor hour reducing from the present 50c, to around one quarter of that value in the first phase of the new peak facility (i.e., around 13c) and finally by a further factor of 2 to around 6-7c when the machine is fully incremented. While partners would effectively acquire resources at the 13c and 6-7c per processor hour in each of the two phases of the new facilities, affiliates should not be able to access resources at the 6-7c rate in the final phase (having not contributed for the full term of NCI) but instead should pay at a higher rate of (say) 10-13c (to be determined).

Through this change of definition, there is now the possibility of both "large" and "small" partners, with their strategic interests differentiating them. Large investors will have a greater stake in the NCI program and in the planning of the national high-end computing fabric.

Accordingly, we envisage that

- large partners will have representation on the NCI Steering Committee and on the national planning council (referred to as the AHECC—Australian High-End Computing Council discussed under changes to Governance and Management in Section 4), while
- small partners would have proportional representation on the NCI Steering Committee (i.e., a number of small partners would be represented by one of their number), and would not be represented on the national planning council.

3.3 Outreach

3.3.1 Raising the Profile of High-End Computing

A key focus of the NCI Outreach program is to raise the profile and uptake of high-end computing in Australia and, in doing so, to be seen to b e making a difference in the impact and scale of Australian scientific achievements. In this regard, NCI is mindful of its mission of providing Australian researchers with world-class high-end computing services.

NCI has commenced a program of increasing its engagement with key research fields, NCRIS capability areas and major research organisations. While the initiation of the program during the reporting period was held up by delays in the appointment of the Director, significant progress is now being made, with an extensive program of visits to leading research organisations now well underway.

The focus of these visits is to engage directly with key researchers, research communities, and with the senior management of research organisations and universities. At the time of writing this report, visits to the Bureau of Meteorology, CSIRO, Geoscience Australia, Monash University, the University of Melbourne, the University of NSW and the University of Sydney have been completed. A number of these organisations are contemplating their co-investment in NCI as partners, with others planning collaborative or co-operative relationships. This program of visits will be expanded in the coming year to build relationships and also the resource base of NCI.

NCI is working particularly closely with the climate modelling community (CSIRO/CAWCR and universities through the University Climate Consortium) to assist their investment in and uptake of the next generation National Facility.

3.3.2 Promotional Material and Marketing

As part of the establishment of NCI's identity, a logo and emblem have been established for use on letterhead, documents, websites and promotional material. New banners have also been produced for the NCI Office and the NCI National Facility for use at conferences and presentations. Signage outside the CSIT Building at ANU (where the NCI Office is located) and on internal directory boards has been rebadged to reflect the change from APAC to NCI, with corresponding changes also being made at the National Facility which is located in the Leonard Huxley Building on the ANU Campus.

Brochures and posters, for use at conferences and other publicity events, are being prepared by National Facility staff in a form that reflects the new identity. Derivatives of this material will also be used on the website to highlight the range and excellence of research being supported by NCI. NCI branded pens for distribution at publicity events are also planned.

3.3.3 Website

The website addresses of *nci.edu.au* and *nci.org.au* have been registered and a web developer was engaged to develop templates and establish a Plone content management system for simplified ongoing management and enhancement of the site. At this stage a functional website has been populated with material outlining NCI and its mission, the services and facilities that it operates, together with a range of news and information, and contact details for NCI and the National Facility. A private site has also been established on which material (agenda papers, minutes etc) for access by NCI/SC members is placed.

While the website is now in a functional state, more work is needed to improve its usefulness and attractiveness, in addition to establishing a significant outreach presence in the form of exciting, interesting and accessible applications of high-end computing. Staff of the National Facility are preparing publicity material (brochures and posters) for use at conferences and derivatives of these will be included on the NCI website to showcase the quality of the research being undertaken on NCI facilities.

The National Facility has its own extensive website, complete with detailed user guides and online applications for obtaining resources. Amendments have also been made to this website, rebranding it with the NCI identity and making it addressable as *nf.nci.org.au*, in addition to the familiar antecedent *nf.apac.edu.au*. The National Facility is planning a major revision of its website in the coming year.

3.3.4 Conference Activities

During the reporting period NCI's antecedent, APAC, ran the fourth of its conference series APAC07.

APAC07

The APAC Conference and Exhibition on Advanced Computing, Grid Applications and eResearch (APAC07) www.apac.edu.au/apac07 was held at the Rendezvous Hotel, Perth during 8-12 October 2007.

Highlights of the event were:

- The opening by The Honourable Francis M. Logan, MLA, Minister for Energy, Resources, Industry and Enterprise, WA Government, on Tuesday 9 October, 2007.
- The inclusion of eight leading keynote speakers in the program—Dr Brian Coghlan (Founder and Director, Grid-Ireland Operations Centre and Leader, Computer Architecture and Grid Research Group, Trinity College, Dublin), Prof. Thom Dunning (Director, NCSA), Dr Bill Kramer (NERSC Division Deputy and NERSC General Manager, Lawrence Berkeley National Laboratory), Prof. Bertram Ludaescher (Department of Computer Science and the Genome Centre at the University of California, Davis), Dr Erwin Laure (EGEE Technical Director), Dr Rick McMullen (Director, Pervasive Technology labs, Indiana University), Mike Netzband (Manager, Emerging Technologies, Technical Computing Department, Chevron Energy Technology Company), Prof Reagan Moore (Associate Director for Data Intensive Computing, San Diego Supercomputer Centre), Dr Dane Skow (Director, TeraGrid Grid Infrastructure Group, University of Chicago/Argonne National Laboratory).
- Presentations from 13 invited postgraduate students from around Australia at the Student Forum.
- Seven well attended workshops held over three 3 days.
- Conference support from thirteen companies, the Western Australian Government and thirteen research organisations.
- An international spread of conference presenters and attendees (Australia, Ireland, Scotland, New Zealand and the USA).

The conference was attended by over 230 delegates, with the tutorials and workshops having good audiences.

Feedback from participants indicated that

- the quality of the conference and exhibition was very high,
- the venue was attractive,
- the research exhibits added significantly to the exhibition.

eResearch Australasia

While there was broad support for events similar to APAC07 in the future, eResearch Australasia is becoming the broadly accepted annual meeting of eResearch practitioners. Accordingly, NCI will contribute to eResearch Australasia 2008 (29 September–3 October 2008) through the provision of a National Facility Workshop and the NCI Director's participation in the AeRIC eResearch Forum.

Given that the scope of eResearch Australasia is not in strong alignment with NCI's science oriented researcher base (the majority of whom would choose to attend discipline based conferences rather than a broadly based eResearch conference), NCI is investigating ways in which it can engage most effectively and directly with its communities of research users and stakeholders. This includes research forums and (institutional) meetings with key users and managers, which are reported elsewhere in this document. While the Business Plan refers to the possibility of an APAC09 conference co-hosted with eResearch Australasia, NCI needs to understand more clearly if there exists a real need before embarking on this path, given the focus of NCI activities, and the gradual evolution of the former APAC series of conferences away from computational methods and applications.

3.3.5 Training Programs

The NCI National Facility continues its outreach program in which staff of the NF provide user workshops and training visits to research communities and research organisations as and when required. Details of these can be found in the National Facility Reports (Appendices 3 and 4).

Attendance, possibly with displays, at major HPC events internationally will continue; examples for 2008 include SC08 (the premier international supercomputing conference) and WATOC08 (computational chemistry).

4. GOVERNANCE AND MANAGEMENT

Present Status

The Funding Agreement specifies that NCI will be governed by ANU on advise from a Steering Committee (NCI/SC) on the following matters:

- strategic plans for the national computational infrastructure and services;
- business and marketing plans for the Program;
- project plans for outreach activity;
- budget allocations within the Program;
- the general direction of implementation of the Program and associated delivery of services to users; and
- arrangements for promotion, collaboration and cooperation among the contributors, counterparts, and research organisations nationally and internationally.

The Steering Committee comprises:

- an independent Chair,
- one representative from each of the Partners,
- one representative elected by the Affiliates,
- the Chair, NCI Merit Allocation Committee,
- the Director, NCI,

in which partners are substantial investors over the lifetime of the program and affiliates are lesser investors, perhaps with transient interests in acquiring services from the program.

At the inception of NCI, the Steering Committee comprised:

- Emeritus Professor Mark Wainwright, Independent Chair,
- Professor Robin Stanton, ANU,
- Professor Brian Yates, MAC Chair,
- Professor Jim Williams, NCI Acting Director.

Consent was given from the Department of Innovation for the following prospective partners and affiliates to be represented on the NCI/SC as full members until 30 September 2008:

- CSIRO, represented by Dr Alex Zelinsky,
- BoM, represented by Dr Neville Smith,
- GA, represented by Dr Chris Pigram,
- Research intensive universities represented by Professor Doug McEachern, UWA.

In the 2007-08 financial year, the NCI/SC met on four occasions (13 December 2007, 14 February 2008, 15 April 2008), with a further two meetings planned for the calendar year 2008 (2 September 2, and 25 November). These meetings have focussed on initiating the NCI program and overseeing the transition from APAC. The work of the NCI/SC has encompassed all aspects of the program including the acquisition of new facilities (the peak system and the development system), the initiation of the specialised facilities and computational tools and techniques programs, the extensive outreach activities to encourage buy-in from potential partners and affiliates, the revision of the merit allocation program, the development of a resources allocation model that will underpin the growth and development planned for NCI, and the appointment of the Director.

NCI commenced operations with Prof Jim Williams as Acting Director. The position of substantive Director was advertised in January 2008, with interviews on 13 March, 2008 leading to an offer being made to Professor Lindsay Botten who took up his appointment on 26 May 2008. While the late appointment of the Director has delayed some aspects of the program, these are being rapidly made up.

The NCI Office is an organisational unit within the ANU and contains the positions of Director and Executive Officer. Members of the office during the reporting period were:

Director:	Professor Jim Williams (July 2007 – April 2008)
Director:	Professor Lindsay Botten (from 26 May 2008)
Assoc Director:	Dr Lindsay Hood (July 2007 – April 2008)
Executive Officer:	Ms Sue Cameron (from July 2007)

In its role as host institution, ANU is providing financial, human resource and legal services to the office. ANU is also providing accommodation for the NCI Office.

Evolution of the Role and Structure of the NCI Steering Committee

The key strength of the present Steering Committee is that it binds together many of the leading research organisations which regard high-end computing as an essential capability in achieving their research and development goals, and which invest substantially in infrastructure and services to achieve these ends. In doing so, it constitutes an influential council through which to formulate strategic policy settings for the advancement of national computational infrastructure and services. The underlying flaw in the compositional requirements, however, is that membership of this council (through the NCI/SC) is predicated on partnership with NCI—with the requirement that organisations invest a minimum of $1M p.a. in NCI.

This formal requirement of investment in NCI is likely to weaken the national planning framework by potentially excluding some current members of the Steering Committee, and also by failing to take into account major initiatives by organisations external to NCI—but which, in either case, make very substantial investments in high-end computing.

The following outlines the present situation:

- at this time only ANU and CSIRO have committed to investment in NCI (substantially exceeding the partnership threshold), with contractual agreements being developed,
- the Bureau of Meteorology is making a very large investment in its own new operational facilities, external to NCI, but through a joint tender with ANU, the aim of which is ensure that the BoM and ANU/NCI systems are interoperable,
- the University of Melbourne and the State Government of Victoria have announced plans to establish a large computational facility to support research in the life sciences, outside of the NCI framework,
- Geoscience Australia, a major investor in high-end ICT services, is contemplating future co-investment in NCI, but is not yet in a position to commit to a formal partnership,
- a number of research intensive universities, each of which invests substantially in high-end computing infrastructure and related eResearch services, is seriously contemplating direct investment in NCI, but is yet to commit formally to this.

While the uptake of computational science and engineering is growing, and there is increasing reliance on high-end computing to underpin advances in research, this trend needs to be accelerated if Australia is to maintain an internationally competitive position in key research fields. In order to ensure that Australia is able to compete successfully in internationally competitive research fields in which high-end computing is a vital element of the research infrastructure fabric, a shared vision for the advancement of Australian computational research, and the development of a coherent plan to carry the nation forward into the next generation of high-end computation are needed.

For this to occur, the base of national policy development, presently attributed to NCI, needs to be broadened and strengthened by drawing together all of the leading research organisations that recognise the importance of, and the need for, high-end computing in the achievement of their research goals.

A Bifurcation of the Roles of the NCI/SC

These matters have been addressed by the NCI Steering Committee which has agreed, in principle, subject to agreement from the Department of Innovation, to a course of action which sees:

- the establishment of an Australian High-End Computing Council (AHECC) or similar nomenclature— an organisation whose members, through their commitment to the application of high-end computing advancing the scope and excellence of their research, and through their ownership and stewardship of the nation's major computational infrastructure, come together to plan, coordinate and collaboratively fund the nation's high-end computing infrastructure.
- the transformation of NCI to an operational program of NCRIS Platforms for Collaboration, the responsibilities of which relate solely to the operation and management of the National Computational Infrastructure program, with its planning responsibilities ceded to the AHECC.

Under such a transition, membership of the NCI Steering Committee would be restricted to organisations that are financial partners and affiliates of NCI.

Role and Composition of the AHECC

The role of the AHECC, agreed to, in principle, by the 2 September 2008 meeting of the NCI Steering Committee is to:

- serve as the national peak body in high-end computational and data infrastructure, developing both national relationships and relationships with comparable international bodies,
- oversee the development of an internationally competitive high-end computing fabric for Australia through collaborative and cost-sharing arrangements,
- develop and update, on an ongoing basis, a coherent national strategic plan for the development of high-end computational and data infrastructure, advising the Commonwealth Government and other stakeholders appropriately,
- promote the development and uptake of high-end computational research, engaging with the requirements of the research community through a high-level Scientific Advisory Committee that would form a sub-committee of the Council,

- plan and co-ordinate the development of the skills and expertise requirements of computational science and engineering in Australia,
- oversee and coordinate national merit / priority access to the major HPC resources operated by members of the Council.

It is envisaged that the AHECC would be assisted in its work by a *Scientific Advisory Committee* comprising leaders of the research communities that constitute the major current and prospective users of national high-end computing facilities.

The role of the *Scientific Advisory Committee* would be to:
- provide the underpinning scientific direction that informs decisions of the AHECC,
- provide a standing forum for computational research communities,
- recommend strategies, activities and developments to the AHECC that will inform the coordinated development of high-end computing in Australia,
- consider any matters referred to it for advice by the AHECC.

Proposed Composition of the AHECC
The AHECC would comprise organisations
- whose mission includes publicly funded research and development,
- whose research and development objectives require high-end computing skills and expertise,
- whose research and development performance depends either directly or indirectly on high-end computing capability,
- which are willing to improve national high-end computing capability and infrastructure through cost sharing and collaborative arrangements,
- which have a strategic commitment to a nationally coordinated high-end computing fabric, and
- which are committed to contributing to the formation of a national high-end computing strategy.

Organisations which meet this description, which subscribe to the vision of a national high-end computing strategy and which have a commitment to, and can accept shared responsibility for, the development of the national computational capability and infrastructure through cost-sharing and collaboration would be invited to sign a charter of association (under development).

Since the Council seeks to bring together organisations with a strong interest and a strong financial stake in the advancement of the national high-end computing infrastructure, prospective members would need to demonstrate substantial levels of expenditure in high-end computing infrastructure and services (around $1M or more per annum).

Membership
The proposed membership of the AHECC is:
- An independent Chair,
- A representative of each organisation signing the membership charter,
- Chair, Scientific Advisory Committee,
- Chair, National Merit Allocation Committee,
- Director, NCI (Secretary of the Council),
- Representatives of the Australian Research Council and the National Health and Medical Research Councils (observer status).

Concomitant Revision of the NCI Steering Committee
With the formation of the AHECC, there would be consequential changes to both the role and composition of the NCI/SC.

Revised Role
A revised NCI Steering Committee would advise the ANU on the following matters that relate directly to the management and operation of the NCI Program:
- business and marketing plans for the NCI Program,
- budget allocations within the NCI Program,
- operation of the NCI Merit Allocation Scheme,
- the general direction of the implementation of the NCI Program and associated delivery of services to users,

with the strategic policy development functions of the originally constituted committee ceded to the AHECC.

Revised Composition
- An independent Chair,
- A representative of each of the major financial partners of NCI (*),
- Pro-rated representation from the minor financial partners of NCI (**),
- Chair, National Merit Allocation Committee,
- Director, NCI.

(*) major financial partners of NCI are long term investors in NCI with a substantial investment in high-end computing infrastructure and services that entitles them to membership of the AHECC.
(**) minor financial partners are long term investors in NCI, but whose investment would not be sufficient to entitle them to membership of the AHECC.

Necessary Immediate Action

The terms of the present prospective members of the NCI Steering Committee (i.e, CSIRO, GA, BoM and Research Intensive Universities) expire on September 30 2008. While CSIRO has indicated a strong financial commitment to NCI, BoM, GA and the Research Intensive Universities are still contemplating future partnership.

If the principle of bifurcating the national planning and operational functions of NCI (as outlined above) is accepted, this may take some months to implement. In order to avoid weakening the strategic expertise of the Steering Committee as it presently stands, an extension of the present terms of the prospective members is requested until either the establishment of the AHECC, or 31 March 2009, whichever is the earlier.

5. PERFORMANCE INDICATORS AND MILESTONES

The NCI Business Plan 2007-11 listed milestones for the July 2007 to June 2008 period. These and an outline of progress against these are detailed below.

Progress against Milestones—January to June 2008

Activity	Expected Date	Progress against milestone
Appointment of Director	Q2 2008	The NCI Director commenced duties on 26 May 2008.
Start procurement process for new peak system	Q2 2008	A joint tender between the Bureau of Meteorology and the ANU was let on 3 April 2008 and closed on 29 May 2008. A comprehensive evaluation of tenders has been underway since then.
Call for EoI for Specialised Facilities	Q2 2008	The call for Expression of Interests if the Specialised Facilities Program was issued on 21 July 2008 and closed on 20 August 2008, although almost all applicants requested and were granted a one week extension. A process of evaluation and clarification with proponents has been underway since then.
Call for EoI for Specialised Facilities	Q2 2008	The call for Expressions of Interest in the Computational Tools and Techniques Program has not yet been issued. It was decided to proceed serially with the two calls for EoI in SF and CT&T, and indeed the SF call has identified two or three areas that are candidates for target user support under the CT&T program. A call for Expressions of Interest will be issued in Q4 2008.

The performance of NCI against agreed performance indicators is provided in Attachment 1.

Updates to the Business Plan have been foreshadowed in Sections 2 and 3 and a revised Business Plan will be provided when the relevant matters have been clarified or have developed more fully.

Attachment 1: Report on Performance against Key Indicators

The NCI Agreement and Business Plan list key objectives, key performance indicators and outcomes. The report on the performance of NCI against these indicators for the period July 2007 to June 2008 is given below.

Objective: Develop a national strategy for advanced computing to support eResearch in Australia.

A national advanced computing strategy to support eResearch is in place and maintained.

- NCI builds on the successful National Facility program of its antecedent APAC, informed by its mission of "providing Australian researchers with world-class high-end computing services".
- NCI is building partnerships with leading research organisations (CSIRO, research intensive universities, Geoscience Australia, Bureau of Meteorology) to establish it as a service integral to the advancement of cutting edge science in Australia.
- Partnerships informed by particular science drivers (particularly earth systems science / climate modelling) are being established (with BoM, CSIRO and research intensive universities).
- The framework for a national advanced computing strategy is evolving through growing engagement with leading national research organisations and universities, with co-investment (e.g., ANU, and CSIRO commitments) and co-operation (the intention to install interoperable facilities at BoM and ANU).
- High-end computing needs involving the interoperation of peak systems at NCI (ANU) and the Bureau of Meteorology (and other facilities such as that at the University of Melbourne in the future) will inform advances in the development of national data fabric, the national grid, and network provisioning. This will involve engagement with ARCS, ANDS and AARNET as necessary.
- The NCI Steering Committee is evolving its plans to strengthen the national planning capacity for high-end computing infrastructure through the creation of a Council for Australian High-End Computing, bringing together research organisations committed to the application of high-end computing advancing the scope and excellence of their research, and which give effect to this through substantial infrastructure investments.

Demonstrated progress towards the implementation of the strategy.

- The NCI National Facility has consistently provided a world-class high-end computing service for Australian researchers, although the current facilities are no longer competitive in international comparisons. The replacement of the peak system at the National Facility is in train with a tender having been let and now being evaluated. The aim is to build facilities that rank favourably with our comparator nations (Canada, Korea and Sweden) and which are comparable with US Track 2 facilities. The planned cash flow from the Commonwealth, however, is less favourable to an early installation of a high-end system, in contrast to that available previously to APAC.
- A development/transitional system to augment processing capacity and provide the National Facility staff with experience on a large scale cluster comparable to that of the next generation peak system has been tendered for, with an order being replaced by the end of September 2008.
- Resource allocation strategies that encourage partner investment are being developed.

Objective: Strengthen relationships to provide national cooperation on Australia's advanced computing infrastructure.

Membership and participation in NCI categorised by degree of involvement and contribution of organisations.

- Present commitments for co-investment in NCI are $6.7M p.a. — $3.4M p.a. from ANU and $3.3M from CSIRO, slightly exceeding the annualised contribution of $6.5M from the Commonwealth.
- Substantial work is in train to raise partner investments of around $1.5M p.a. from leading research universities.
- An extensive program of visits to potential partners is underway, the primary aim of which is to build relationships (co-investment, collaborative and co-operative) with leading researchers, high impact research groups, and senior managers of these organisations. At the time of writing, highlighting visits to BoM, CAWCR, CSIRO, GA, Monash University, the University of Sydney, the University of NSW and the University of Melbourne had been undertaken.
- NCI is working closely with the climate community (BoM, CSIRO, CAWCR, UCC) to assist them in increasing the computational and support resources available to them on the National Facility.
- A researcher forum is planned for 2009 associated with the installation of the new peak system.

Funding received from sources other than NCRIS.

- Present commitments for co-investment in NCI are at $6.7M p.a.

Objective: Improve the peak computing capabilities of the NCI National Facility to serve the demands of Australian researchers.

Key research achievements obtained by users of the National Facility.

- A summary of all NCI (MAS) projects including brief descriptions is also available at http://nf.nci.org.au/accounts/grants/.
- The Principal Investigators of these projects are required to submit an annual report on the project. The reports for 2007 (and earlier years) are available on-line at http://nf.nci.org.au/annual_reports/.
- National Facility staff have developed publicity materials (posters and brochures) highlighting the excellence of the research conducted on NF systems. Some of this material will be used, in a revised form, in the Research Highlights Gallery on the NCI website.

Significant developments in computational tools and techniques on the NCI and partner facilities.

- The Computational Tools and Techniques program has a different orientation to that formerly conducted within APAC. Under NCI's program, highly targeted support will be provided to key research areas (e.g., climate modelling, astronomy, chemistry etc). NCI is working with research communities to obtain additional funds via alternative channels (e.g., other Government funding sources) to augment the modest CT&T budget within the NCI program.

Capability of the National Facility relative to countries of similar size and development such as Canada, Sweden and Korea.

- Based on the June 2008 release of the TOP500 list (www.top500.org) for June 2008, Australia has only one system (from industry—finance sector) within the top 500 list, ranked at position 468. The current National Facility has fallen off the TOP500 list, last appearing at rank 200 in the November 2007 release of the list.
- The decline in Australia's representation on the TOP500 list is demonstrated in the table below, commencing with June 2005, the first period in which the current National Facility was recorded in the list.

List Date	Australian Systems in TOP500 List	Australian Aggregate Performance (ΣRmax Gflops)	Highest Ranked Australian System	International Rank of National Facility	Australian Rank of National Facility
June 2005	5	19,592	26	26	1
November 2005	11	34,460	36	36	1
June 2006	9	31,023	47	47	1
November 2006	4	20,670	71	71	1
June 2007	4	22,522	107	107	1
November 2007	1	8,974	200	200	1
June 2008	1	9,239	468	-	-

- A closer analysis of the data with time reveals a decline in both the aggregate HPC capacity of the nation (relative to the rest of the world) and of the capacity of mid-range systems available to Australian researchers.
- In June 2008, our comparator nations of Canada, Korea and Sweden all had research systems listed in the TOP500.

  Canada had two systems at ranks 249 and 395, respectively operated by a government department and a university.

  Korea listed one system operated by its national science and technology research organisation at rank 130.

  Sweden listed 9 systems in the TOP500, the highest at rank 11 operated by a government defence agency, the next two at ranks 39 and 40 associated with a university and the national supercomputer centre, and the remaining six at ranks 132, 212, 268, 276, 384, 419 associated with industry (most commonly in finance).

Capability of the National Facility relative to the needs of the Australian research community as indicated by extent and type of demand.

- The demand for the National Facility is demonstrated by:
  - The large number of projects, principal investigators and associated researchers supported by the National Facility.
  - The geographic distribution of users for the MAS (coming from 28 research organisations in every State and the ACT).
  - The broad base of application areas and research disciplines.
    These are detailed in the Tables of Section 3.1.1 and in the National Facility Operations reports in Appendices 3 and 4.
- The demand for MAS grants in the December 2008 round exceeded the available resources by a factor of approximately 1.5. This figure has saturated as researchers are aware that there is no point in applying for resources which are not available.
- The resources available under MAS allocated at the December 2007 round constituted 73% of the facility, in contrast to the 42% allocated in 2004-06 (with this increase associated with a corresponding decrease in the partner shares over this time).
- The increased demand for data storage and data management services is following a similar trend to the computing demands.

- The quality of the research projects supported is indicated by the high proportion of grants with ARC and external support.
- Further details on the demand for the National Facility are given in Section 3.1.1 and in the operations reports in Appendices 3 and 4.

Extent of use of the National Facility indicated by the number of States represented by users, projects, users and resource allocations.

- During the six month period from January-June 2008, there were 224 projects through the MAS and partner shares (mainly ANU) associated with 935 researchers with login accounts (and a further 77 researchers without accounts).
- The MAS scheme accounted for 147 projects, of which 110 (75%) were supported by research funding (ARC, industry, or other). Recasting this in terms of allocated system resources associated with research grants, this figure rises to 83%.
- Further details on the extent of use of the National Facility are given in Section 3.1.1 and on the National Facility reports (Appendices 3 and 4)

Availability and performance of the National Facility indicated by the available system units, used system units, efficiency of operation and extent of parallel computation on the National Facility.

- During the six-months from January to June 2008, the SGI Altix system was available for 97.4% of the total time theoretically available. The loss of time was due largely to problems with internal system routers which were diagnosed and replaced.
- Overall system utilisation for the period has been 97.4% which is extremely high and international best practice, nearing the capacity of the system for the current job mix. The corresponding figure for the July-December 2007 period was 93%.
- The present workloads are very high with the ratio of queued processors to available processors averaging at 3.7 during a quarter and rising to a ratio of 5 at the end of the quarter. There is now a real shortage of cycles and the installation of the transitional system and the new peak system are urgently awaited.
- The percentage of projects using the parallel capability of the peak system is around 85% of the available resources of the system. This has increased from 75% reported in 2006. This percentage accounts for parallel jobs utilising the high speed interconnect of the peak system.
- All jobs utilise the global filesystem, application packages and other software. The performance scales uniformly across the entire system and resource allocation is based on underlying hardware resources rather than dedicated nodes for specific (software) purposes.
- Further details on the availability and performance of the National Facility are given in Section 3.1.1 and in the National Facility Reports (Appendices 3 and 4).

Summary of surveys of users of the National Facility showing their level of satisfaction.

- The Director meets regularly with the Manager of the National Facility to review the operation of the National Facility and to discuss plans for ongoing development of the systems.
- A customer survey was last conducted in March 2006 to gain insight into the level of user satisfaction with the facility and the new SGI system that had been installed. The overall customer satisfaction was at a very high level (4.6 out of a possible 5) indicating that the National Facility is well supported by its users.
- A new customer survey is planned for 2009 to assess user satisfaction and emerging needs following the addition of new capacity through the installation of the development system, and again in 2010 following the installation of the new peak facility.

Objective: Implement specialised computing capabilities through NCI Specialised Facilities to serve the demands of Australian researchers.

Key research achievements obtained by users of Specialised Facilities.

Capability of Specialised Facilities relative to the needs of the Australian research community as indicated by extend and type of demand.

Extent of use of Specialised Facilities indicated by the number of States represented by users, projects, users and resource allocations.

Availability and performance of the Specialised Facilities is comparable to the National Facility.

Summary of surveys of users of Specialised Facilities showing their level of satisfaction.

The Specialised Facilities Program is still in the process of being implemented. Accordingly there is no progress to report against these indicators.

Objectives: (1) Develop and deliver outreach services for users of advanced computing infrastructure. (2) Participation and contribution of organisations in the Outreach program. (3) Number and types of outreach activities conducted.

Extent and nature of working linkages between industry and research organisations.

- NCI is working closely with the climate research community and is also building relationships with leading research organisations (research intensive universities, CSIRO, Geoscience Australia and Bureau of Meteorology) as reported elsewhere in this document. Additional linkages with priority areas will evolve in the coming year.
- The substantial investment of CSIRO shows strong commitment to increasing its use of high-end computing. Development activities will be planned in conjunction with CSIRO, for both the National Facility and the Specialised Facilities in which it is investing.

Increased skills of users of advanced computing infrastructure.

- National Facility workshops have continued to be run.
- Additional support for targeted user areas is planned through the Computational Tools and Techniques Program.

NCI is refining its definition of outreach, with a particular focus on building capacity and capability in research organisations, and in building the uptake of high-end computing.

Revised objectives will be presented in a future business plan.

Appendix 1: Members of the APAC Merit Allocation Committee

The Merit Allocation Committee determines the users of NCI's resource share of the National Facility on the basis of merit.

The following is a list of the current members of the Merit Allocation Committee as at 30 June 2008.

Professor Brian Yates (Chair)
School of Chemistry
University of Tasmania

Professor Mark Ragan
Institute of Molecular Bioscience
The University of Queensland

Professor Julian Gale
Nanochemistry Research Institute
Curtin University of Technology

Prof Geoff Bicknell
Research School of Astronomy and Astrophysics
The Australian National University

Associate Professor Jim Denier
Department of Applied Mathematics
The University of Adelaide

Associate Professor Louis Moresi
Department of Computational Mathematics and Geophysics
Monash University

Members of the MAC contributing to the December 2007 allocations included the above list and also;

Professor Lindsay Botten
Department of Mathematical Sciences
University of Technology, Sydney

who resigned in April 2008, prior to taking up the position of Director, NCI.

There are presently two vacancies to be filled: one in the general area of physics (arising from Professor Botten's resignation) and a new position in the growing area of earth systems science modelling. These two vacancies are to be filled from nominations from NSW and CSIRO to maintain geographical and partner diversity.

Appendix 2: Members of the NCI Steering Committee

The members of the NCI Steering Committee at 30 June 2008 were:

Emeritus Professor Mark S. Wainwright AM (Chair) The University of New South Wales SYDNEY NSW	Independent Member
Professor Robin Stanton Pro Vice-Chancellor The Australian National University CANBERRA ACT	ANU
Professor Brian Yates Chair, Merit Allocation Committee University of Tasmania HOBART TAS	Independent Member
Professor Lindsay Botten Director, NCI The Australian National University CANBERRA ACT	NCI

Prospective partners and affiliates represented on the NCI Steering Committee at 30 June 2008 were:

Dr Alex Zelinsky Director, CSIRO ICT Centre SYDNEY NSW	CSIRO
Dr Neville Smith A/Deputy Director (Research & Systems) Bureau of Meteorology DOCKLANDS VIC	Bureau of Meteorology
Dr Chris Pigram Deputy CEO and Chief of Geospatial and Earth Monitoring Division Geoscience Australia CANBERRA ACT	Geoscience Australia
Professor Doug McEachern Deputy Vice-Chancellor University of Western Australia PERTH WA	Deputy Vice-Chancellors (Research)

Appendix 3: National Facility Operations Report – July 2007 to December 2007

This Appendix outlines the systems in the National Facility and the systems and user support provided by ANU staff working in the National Facility during the period July 2007 to December 2007.

Merit and time Allocations

Merit Allocation Scheme (MAS)

Following an APAC board decision, some partner shares were not continued in the second half of the year. This allowed a small increase in cycles to be used for other purposes such as the MAS. However, since the increase was so small there was no new mid-year call made for the Merit Allocation Scheme. Instead these cycles were required for managing the existing projects. The MAC Chairman reviewed projects submitted to MAC, many of which had resulted from projects moving over to the MAS due to the non-continued partner shares. The Chairman subsequently allocated 15 new projects. The arrangements formal share of the MAS was increased from 56% to 60% for the second half of the year following this change in shares. The Merit Allocation Process for Data projects (MAS-D) is reported under the section on data activities later in the document.

Committee Time Allocation Summary

As for the first half of the year, the time granted for 2007 as a percentage of the overall MAS share was:

	2006	2007
Top 10 projects	42%	40%
Top 20 projects	60%	58%
Top 40 projects	78%	80%

There were 8 new projects in the top 40 projects and all 8 made significant progress in the second half of the year.

The committee has noted that the planning process for shares in 2008 will consider the increased need for cycles under the MAS share given the increased demand.

Reports on Individual Projects

A list of all MAS projects including brief descriptions can also be found at http://nf.apac.edu.au/accounts/grants/

Grant and Usage Date

In this period the National Facility serviced a total of 217 computational projects with 932 attached users. There were an additional 93 researchers associated with projects but who did not have a login-names, bringing the total number of researchers involved with the projects to 1025.

The tables below present the overall number of projects and researchers and their grants and usage during the period broken down by projects under the MAS and other shareholders. More details on the individual projects are available on the NF website. The table also shows that 159 of the projects had support from ARC or external grants. (Note that CSIRO projects are not eligible for ARC grants).

The increase in the formal MAS share on the facility allowed a more manageable allocation to take place. The MAS was 1.82 times over-requested and the committee granted 1.3 times the formal MAS share.

Overall Usage of Computational Time

Division	No. of PIs	No. of Projects	No. of Users with log-in Names	No. of Researchers (total)	No. with ARC Support	No. with NHMRC Support	No. with Industry Grant	No. with Other Funding	Grant as % of Total Resource Available	Usage as % of total Resource Used
MAS	137	137	690	759	80	3	13	39	75.8	72.4
MAS-D	2	2	18	18	2	0	0	0	0	0
APAC	3	3	5	5	0	0	0	1	0.1	0.1
Other Shares	74	80	357	389	35	3	2	12	37.8	27.6
Start-up	11	11	20	23	0	0	0	0	0	0
TOTALS	213	233	972	1068	117	6	15	52	113.7	100

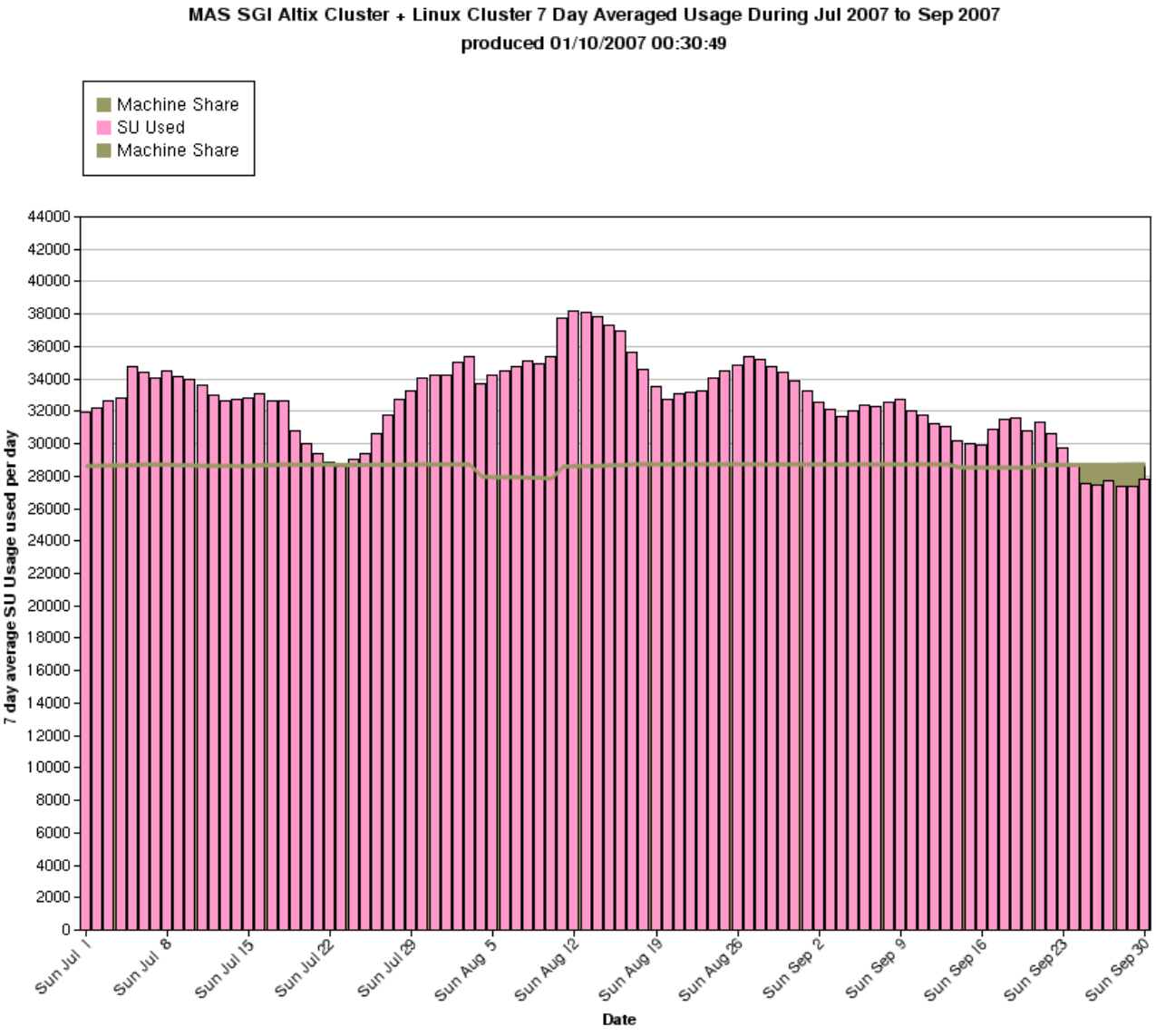
Other Shares

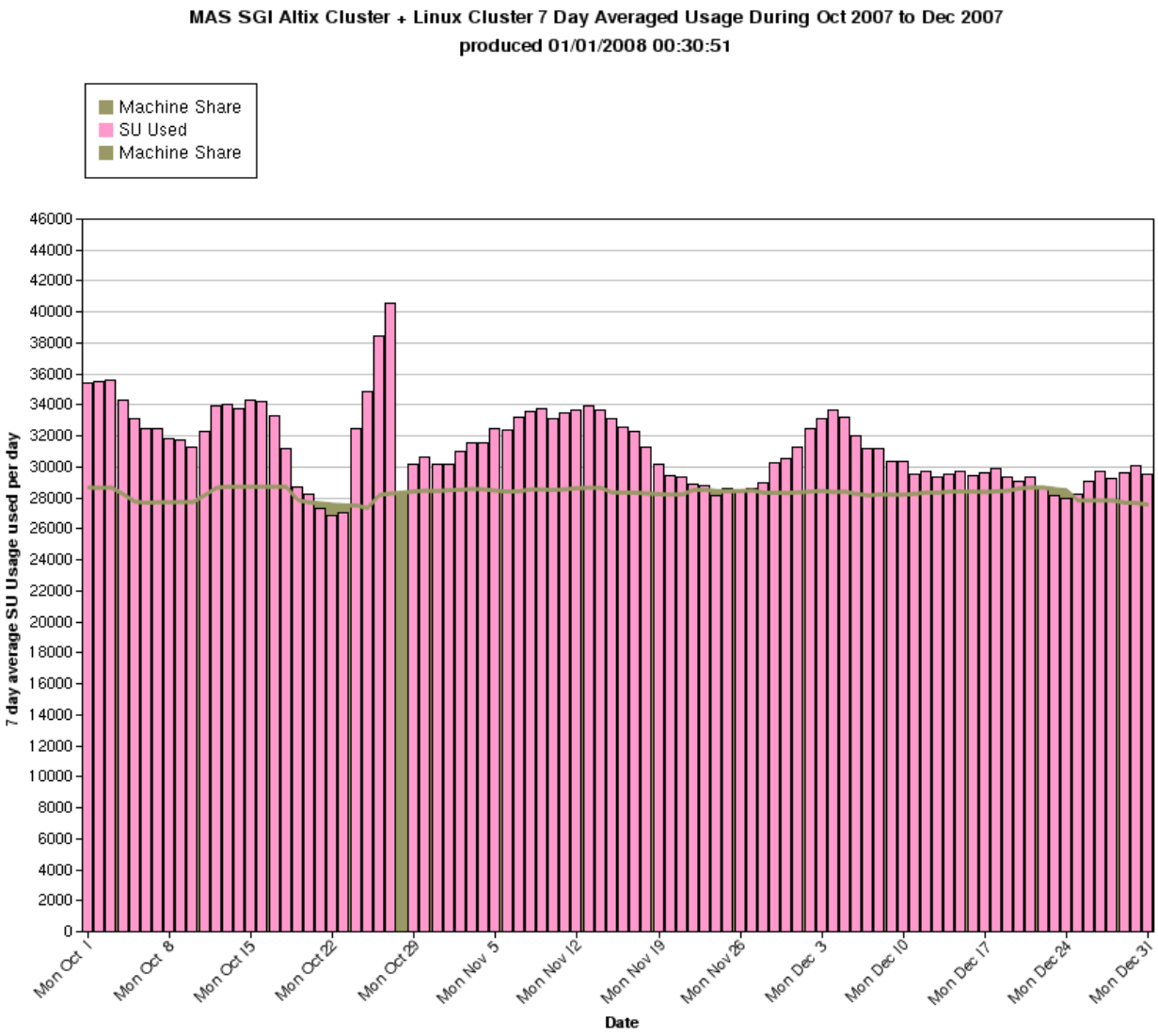
Division	No. of PIs	No. of Projects	No. of Users with log-in Names	No. of Researcher (total)	No. with ARC Support	No. with NHMRC Support	No. with Industry Grant	No. with Other Funding	Grant as % of Total Resource Available	Usage as % of total Resource Used	Format % Share of the System
ANU	45	51	268	285	29	3	2	8	34.7	24.7	37
CSIRO	11	11	39	43	0	0	0	2	1.0	1.1	1
IVEC	8	8	30	36	1	0	0	1	1.0	0.9	1
QCIF	10	10	22	27	5	0	0	1	1.0	0.9	1

Note: Totals represent unique PI, researchers, and users.

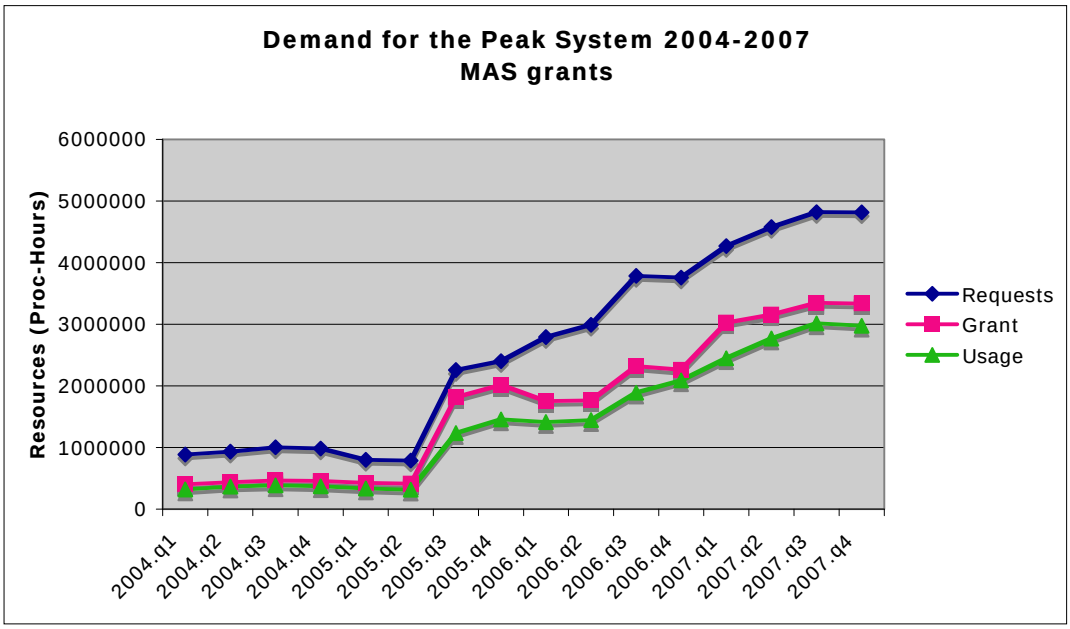
Demand for the System

The demand for the system continues to be high. System utilisation was an average of 93% and the MAS share has been fully used on the system over the whole period. The typical amount of queued work on the system has been approximately 2.5 times the number of cycles on the system, with an average load of 5000 CPUs, with a high of just under 8000 requested CPUs.



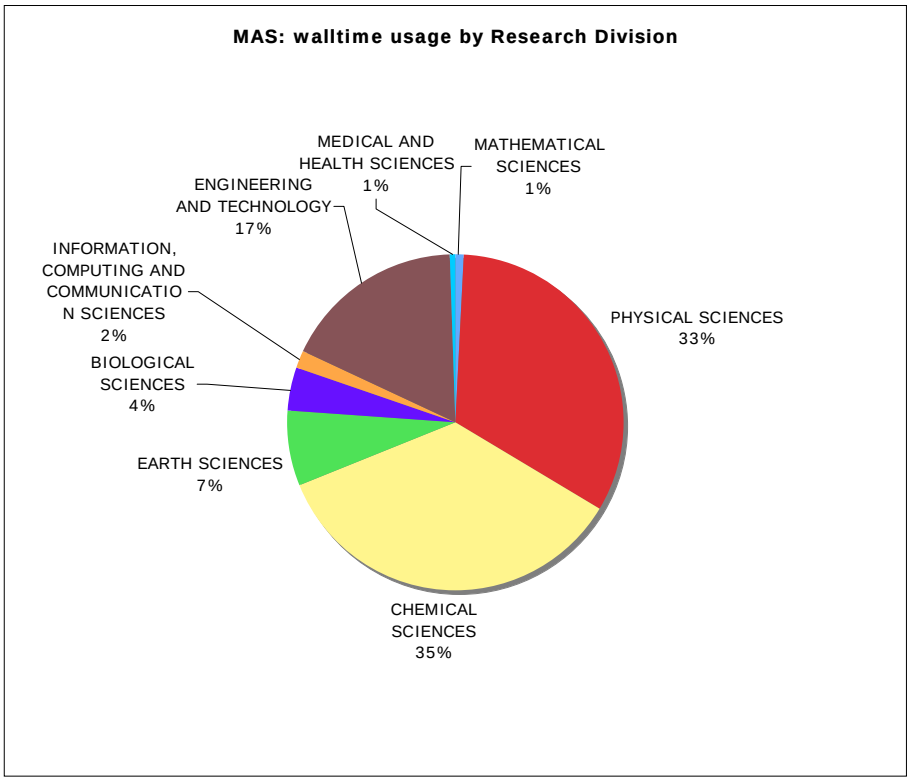


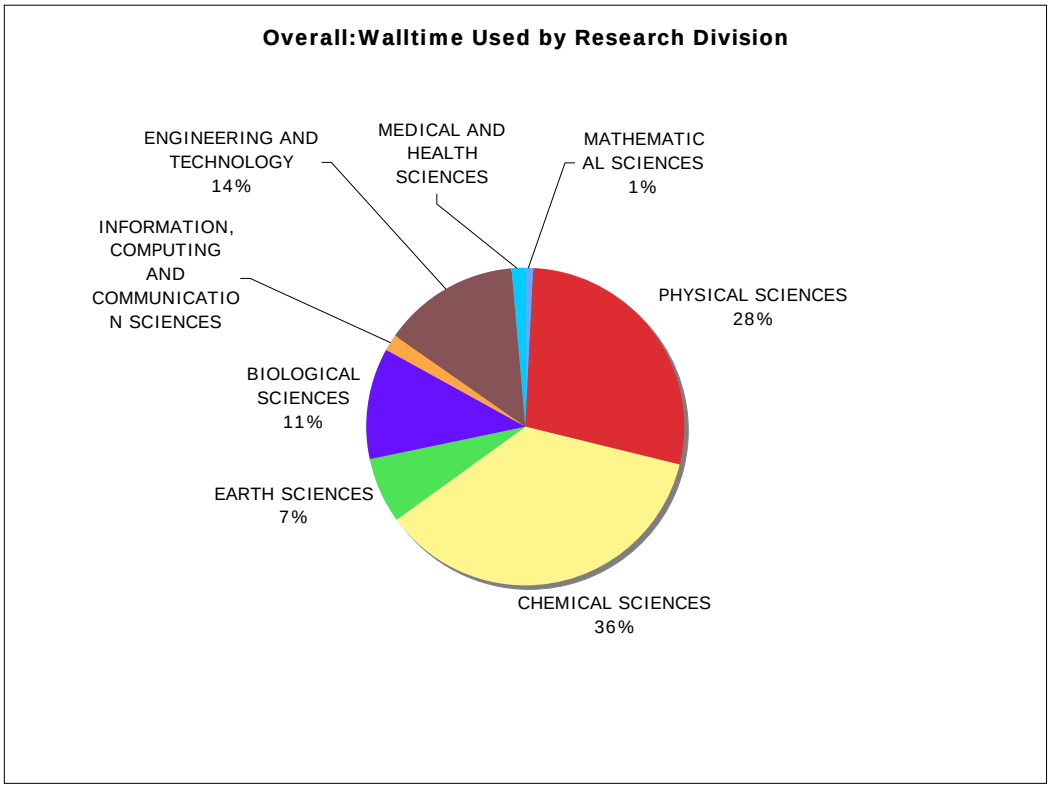
The following graphs show the demand for MAS share over the period of APAC-2 (2004-7).



Usage by Research Field

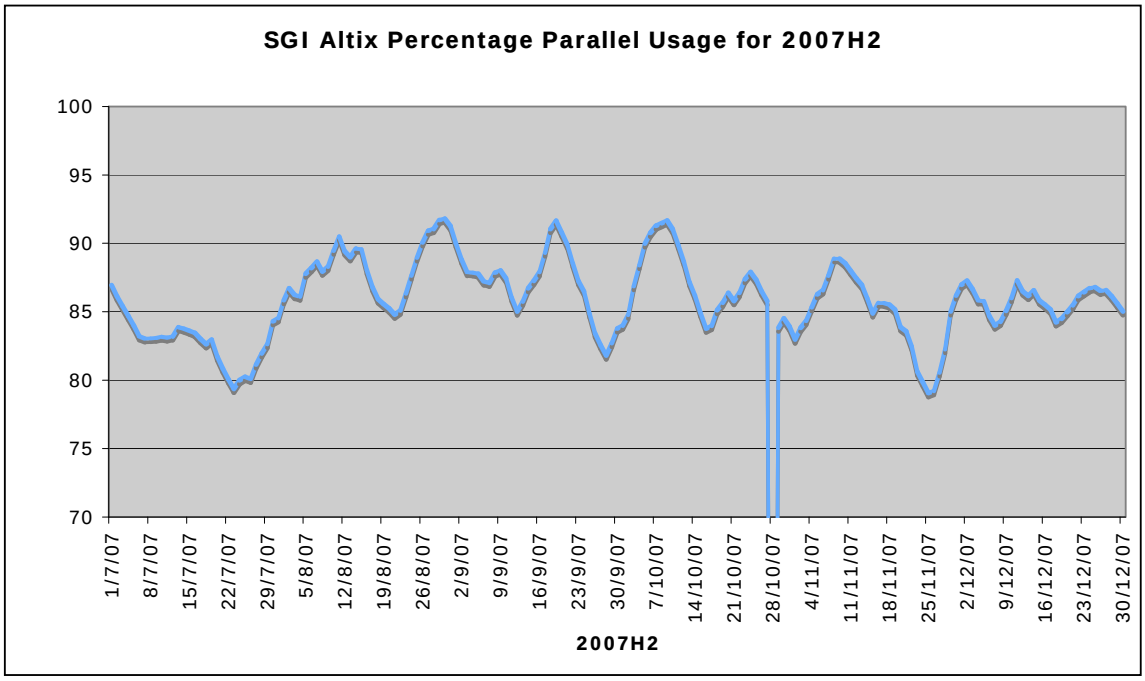
A summary of the usage by research disciplines during the period is given in the two pie-charts below. The first is restricted to MAS computational projects only while the second shows usage over the entire user base.





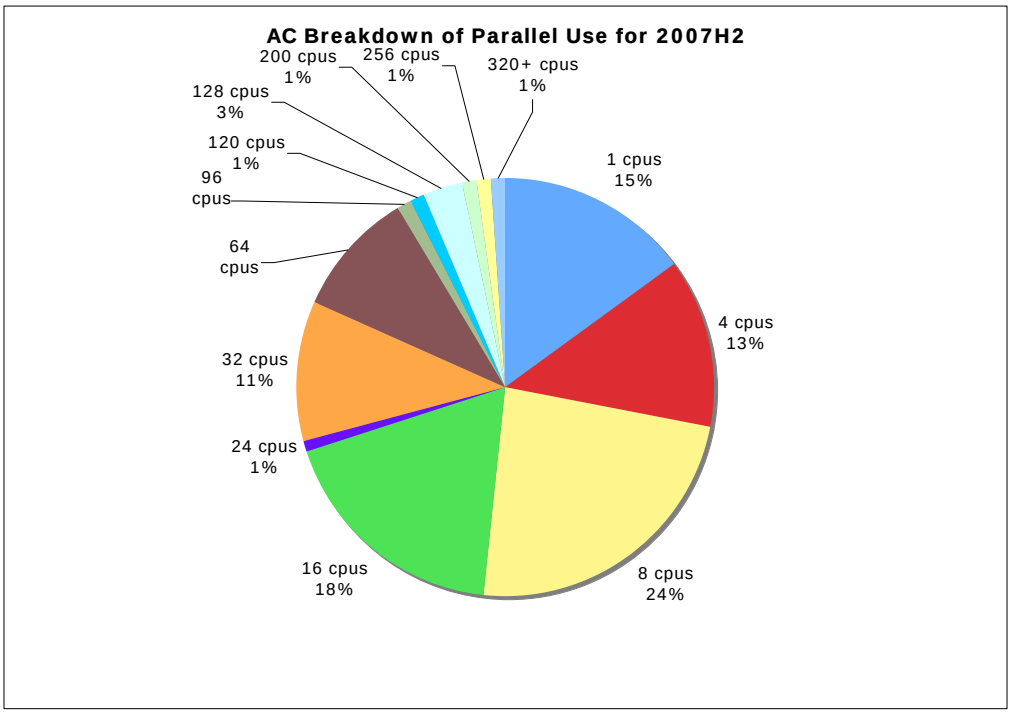
Measure of Parallel Use of the AC

The following graph indicates the percentage of the ac system resources used by parallel jobs. The NF encourages users who might otherwise run an ensemble of tasks as a parallel task, to run the jobs serially to allow the scheduler to make maximum use of the system.



Breakdown of Number of Processors Used on the AC

The following pie-chart shows the fraction of the total usage consumed by jobs making use of the number of processors. The number of CPUs or processors per job increases clockwise from the top, generally in multiples of 2, from 1 CPU / job to 320+ CPUs / job.



Support and Development with Research Communities and Projects

The following communities have received particular attention in this period.

Chemistry

Researchers from across the nation required La Trobe, UTas, Swinburne, CSIRO, Monash, ANU, UNSW, USyd, Adelaide and UQ). In-depth assistance was given to Drew Parsons (x21, ANU) regarding calculations using GAMESS-US and Gaussian. Ian Dance (k32, MAS) regarding problems with DMol3.

Through a user request, Rika Kobayashi delivered a half-day "Introduction to Gaussian" course at Macquarie University.

A number of chemistry packages were updated.

- ADF program was upgraded to 2007.01.
- The latest revision of Gaussian 03 Rev E.01 was installed on AC.
- Installed wien2k v7
- Installed gromacs v3.3.2
- Installed Abinit 4.44
- Added vasptools to vasp v4.6.31 (vasp4.6.31nf1)
- CASTEP license server issues were addressed. We note that the CASTEP usage continues to be low and we will consider arrangements in the new year.

The process of porting packages to Intel-64 has begun, using the Altix XE system as a platform.

In this period there was also considerable ongoing work towards organising the WATOC2008 conference to be held in Sydney, September 14-19 2008.

Climate Modelling

A large amount of work has been done in getting UK Met Office Unified Model running on the system. Dr Larson made several visits to the Met Bureau to ensure that code installed on the National Facility was the same as that at the Bureau. Several improvements have been made to the configuration to increase the robustness of the job submission system and the automatic check-out and compilation of source code. SGI and ARCNeSS (ARC Network for Earth Systems Sciences) have appointed Mike Rezny to drive the porting effort.

The CAM (Community Atmosphere Model) software has been further investigated and tuned for a series of paleo-climate models (You, k33 MAS). John You's student, Nicholas Herold was also getting ancilliary Fortran code running.

The ARCNess workshop on using GEOS_CHEM was assisted. Guergana Guerova (m19, MAS) was the conference organizer and Margaret Kahn (NF) gave an introduction to the National Facility at this conference.

The following packages were also installed:

- MCT 2.30
- CSIRO Mk3L 1.0
- Ncarg 5.0-0

Engineering add Fluid Mechanics

Sebastian Azagari (h66, MAS) was assisted with the use of fluent.
The ABAQUS software package was updated to 6.7-1

Plate Tectonics

The latest release of Underworld (version 302-p1) is just about to be released. Underworld was previously supported through the APAC CT&T program.

Medicine

The project by Cherben (ANU, j66) has progressed into a substantial project. To support the MRI scanning component, a new version 4.0.0 of the freesurfer software has been installed, and its memory usage has been characterised.

Linear Optimisation Problems

Ting Yu (Project ba0 MAS) was provided assistance with the evaluation of parallel CPLEX. This project will eventually involve running a large optimization problem, too large for the 1 processor CPLEX. A trial parallel CPLEX license but it was discovered that it did not suit the problem. The parallel optimization packages such as TAO, and OPT++ were investigated to see if they would assist. The work was continuing after the end of this period.

General tools and Other packages

The Intel compilers have been updated to latest 10.1 releases. Boost was updated to 1.34.1 for both the intel 9 and 10 compilers. The Intel Trace Collector and Totalview have all been updated. Octave and Ferret have also been updated.

The parallel performance of Intel MKL LAPACK routines was investigated as part of the support for Igor Bray (d35, MAS).

The Photonics package Fullwave was updated to the latest release, 8.0.rc4.

A range of software on the system is regularly installed and updated as new versions of packages, compilers and libraries become available. The full set of software available can be found on the APAC software registry http://nf.apac.edu.au/facilities/software.

APAC Computational Tools and Techniques Sub-Program

The activities of the Computational Tools and Technique sub-program within the National Facility program are no longer funded. However, some of the activities under the program have been reported within the items above as some of the projects have provided support for the relevant communities.

Helpdesk

In the third quarter of 2007 there were 564 emails received by the helpdesk. 18.4% were concerned with minor matters (eg. Passwords, requests for variations to job limits, minor compilation problems etc.) and the remainder being more substantial requests for help. The median response time was 30 minutes and 16% had a response time greater than four hours (this includes responses arriving on weekends and out-of-hours) and 4% longer than 1 day. In the forth quarter there were 636 emails to help, 17.0% relating the minor matters. The median response time was 24 minutes and 22% had a response time greater than four hours, but only 8% longer than 1 day.

A new FAQ was started and placed on the web site. A newsletter was published in Decemeber.

System Modifications

SGI System AC – ac.apac.edu.au

The SLES10 upgrade took place in late November. In general the transition from a user perspective was transparent. There have been new issues with "Out Of Memory" (OOM) issues within the kernel. These issues have been raised with SGI however, given that the linux source code is available to the NF staff, a modification to the kernel has been coded that addressed the problem. This has been sent to SGI for comment.

A new image has been configured so that the node size is now 64 processor compute nodes instead of the 32 processors. This has allowed more flexible handling of large memory jobs, and allowed larger parallel jobs for software that does not support the SGI MPI software stack. This configuration change required a reconfiguration of the systems I/O subsystems. Through this process some redundant hardware was removed and the system configuration simplified. With a larger system image size the number of CXFS clients was also reduced, which will allow development work for data grid transfers to proceed so that they have direct access to the global filesystem.

Dell Linux Cluster LC - lc.apac.edu.au

No changes have been made to the configuration of the linux cluster. The system is out of maintenance and is being tended to on a time and materials basis. A replacement system has been planned as part of the NF upgrade strategy and is waiting on NCRIS processes.

SGI Altix XE Linux Cluster

Lustre has been upgraded to 1.6.4.2 and the system patches for standard RHEL5.1/Centos5.1 kernels for 2.6.23. The OpenMPI was upgraded from 1.2 to 1.2.5.

Now that the system is closer to a production-ready environment, some test groups have been approached and are now beginning to use the system. Since the system has limited cycles the groups approached have been limited to those who were in deparate need of cycles, and appear to be ready to use this system and provide feedback.

Management systems

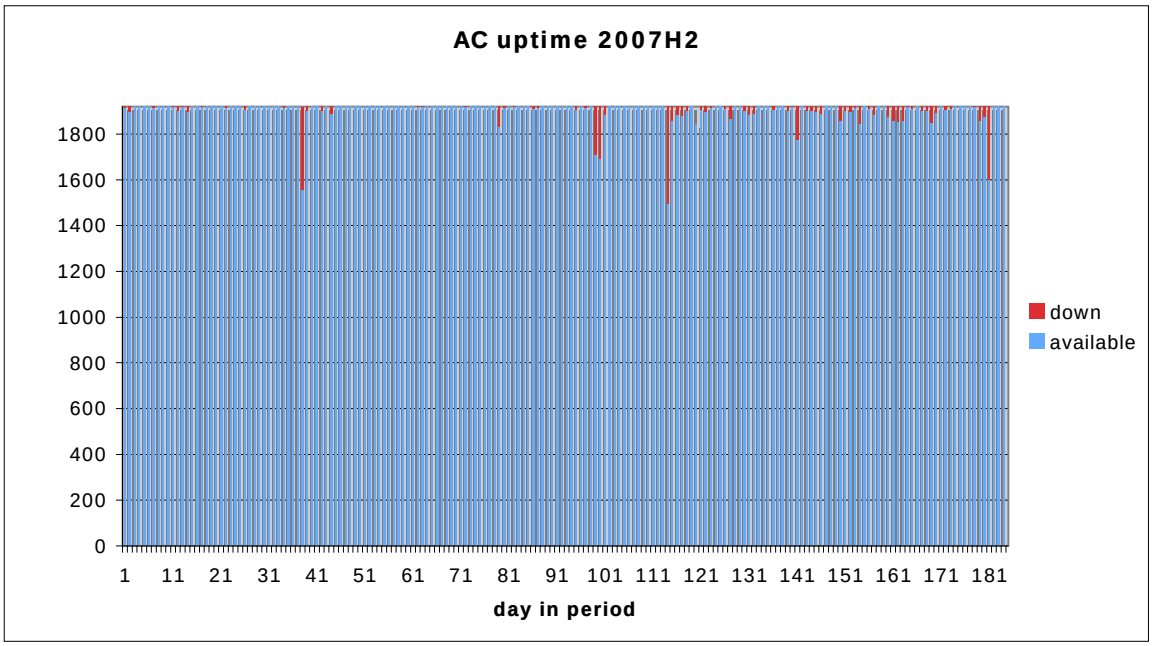
The Nagios software monitoring system has been rolled out to provide both low level and services monitoring to take place across all National Facility systems.

Grid Integration

The relief on the CXFS software licenses through the larger sized nodes has meant work can now commence on installing grid software that directly connects to the global filesystem. SGI are providing details on the restrictions for nodes to attach to CXFS.

Operational Data

The AC system was available for 99.06% of the total time theoretically available. A graph of the system uptime is shown below. The theoretical maximum assumes the full system was 100% reliable and that no downtime was required for system management during the entire period.



The linux cluster LC was available for 98.94% of the total time theoretically available.

Data Projects and Access to Datasets

There were 16 MAS-D granted projects that continued in the second half of the year. These projects were led by 13 Principal Investigators. These projects are supported by a total of 12 ARC grants, 3 are supported through other granting bodies. There are 92 researchers with login access to the data, and a further 15 researchers who play an associated role. Many of these projects have a much larger community who access the data on the system through other authentication mechanisms so the total community services are much larger.

In addition there were a further 9 start-up projects, including 20 researchers with login access and a further 6 researchers with some associated role. It was the intention to assess these projects in the June MAS call.

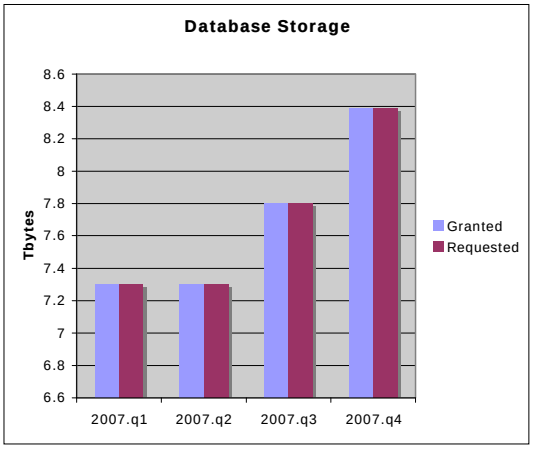
There are also 4 other reference datasets being held.

Data Allocations

The following storage allocations were made for data projects. A * indicates that the accounting processes have not been finalised.

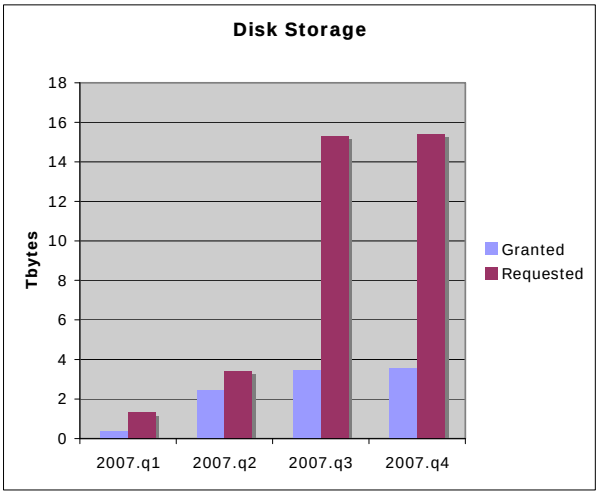
Database storage

Projects	Requests(GB)	Granted (GB)	Maximum Usage(GB)
MAS-D (Data Projects)	8458	8458	*
MAS-D (Data Start-up projects)	130	130	*
Dataset hosting	0	0	0
Totals	8588	8588	*



Online Data Storage (guaranteed disk resident)

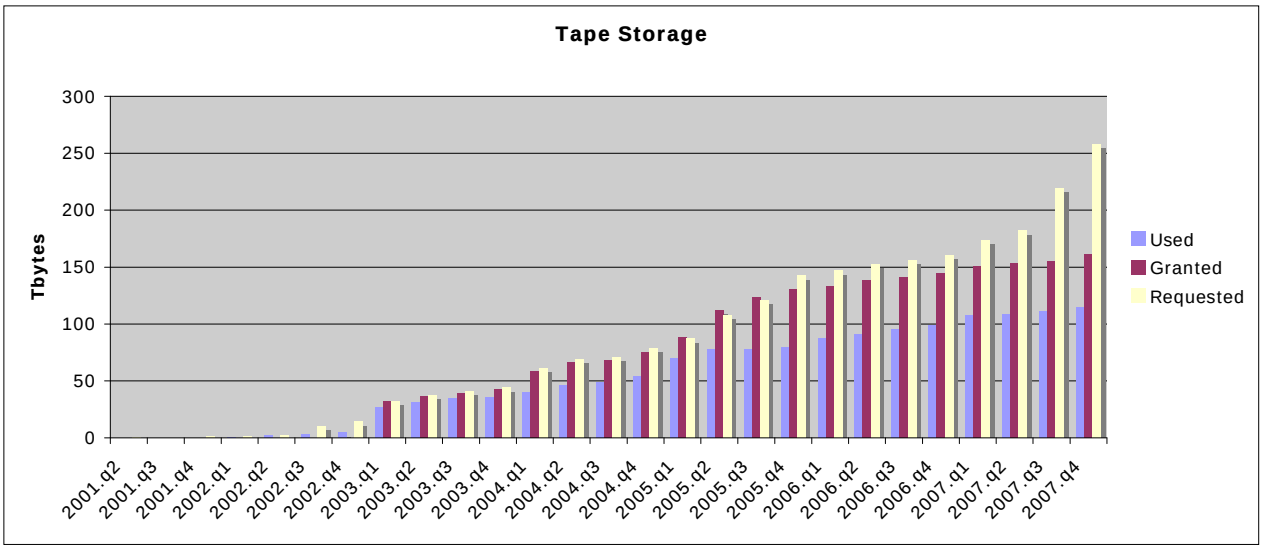
Projects	Requests(GB)	Granted (GB)	Maximum Usage(GB)
MAS-D (Data Projects)	11134	1060	*
MAS-D (Data Start-up projects)	4652	2604	*
Dataset hosting	0	0	0
Totals	15786	3664	*



Data System HSM (tape) storage

The following table and graph shows the requests and use of tape storage space by the MAS-D, MAS-D startup projects, data hosting and also the storage required for computational projects from MAS and shares. The figures represent the first copy only (two separate copies are kept of all HSM data, so that total amount stored on the system is approximately double that shown).

Projects	Requests(TB)	Granted (TB)	Maximum Usage(TB)
MAS-D (Data Projects)	157.88	66.26	46.82
MAS-D (Data Start-up projects)	2.05	2.05	0.94
Dataset hosting	0.13	0.13	0.09
MAS_and Other Computational Shares	97.77	92.45	66.39
Totals	257.83	160.89	114.24



Support for Data Communities and other Data Projects

The following communities have received particular attention in this period.

Astronomy

The SLOAN Digital Sky Survey Data (Hopkins, MAS-D k27) is now available and provides an Australian-based copy of the data that is considered to be a valuable resource for a wide range of astronomers.

There is a wide interest in making astronomy available using Astronomy specific VO tools and technologies. The Australian based MACHO project data (Schmidt, MAS-D d80) contains internationally significant datasets that are being enabled with VO tools. To enable this to table place, a large amount of the data needs to be re-ingest data through relational databases to be accessible in the VO format (Schmidt, MAS-D x60).

- Initial deployment of web services for the MACHO data-set includes:
  - Observation metadata IVOA standard ConeSearch service
  - Observation metadata IVOA standard SIAP service
  - Observation image data FITS file IVOA standard download service
  - Plus corresponding search form interfaces
- Fully compliant MACHO observation metadata ConeSearch v1.0 service developed and deployed.
- Registration of MACHO observation metadata ConeSearch with NVO; and
- ESAVO service registries.

The VO work included completion and deployment of MACHO SIAP service v1.0, including required extensions to the MACHO image FITS file download service. The service is now registered with the NVO service registry. The version 1.0 of an IVOA standard ConeSearch is now available for the MACHO Star databases, including the variable star catalogue. Investigation is underway to assess the performance issues of this large data size of approximately 75 million stars.

Social Sciences

Support has continued for the Australian Social Science Data Archive (Mitchell, MAS-D m26). In particular, a research project in Historical Census has developed and facility staff are involved in the quality assessment of the data markup and portal. The Koori Health Research Database project are also considering technologies that may make better use of the National Facility.

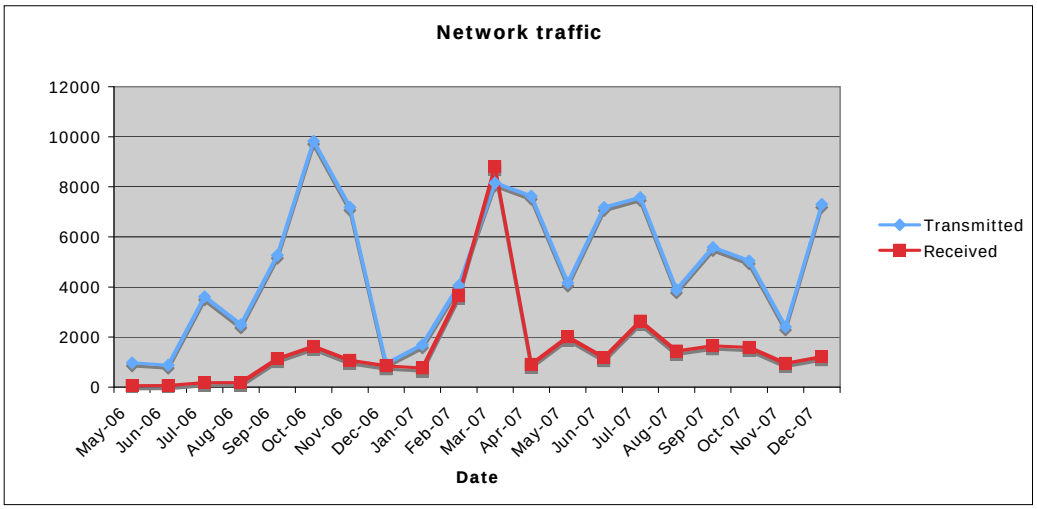
A social network analysis tool called VOSON (Ackland, MAS-D start-up u24) is being made available on the data cluster. This project has a potential to become a broader service of interest to social scientists.

Other Research Group Support

Special assistance was also provided to the new start-up projects on the facility, and also in the transition of the service from the MDSS facility to the new Data Cluster architecture.

Network Traffic

There has been a consistent requirement for data movement both in and out of the facility. Nearly all data is on-net, with data transmitted mostly domestic on-net, while the large data movement in March 2007 was due to the ingest of a large international on-net site. The system is shown to be a data source rather than a data sink.



System Modificiations

Data Cluster – dc.apac.edu.au

Four projects were migrated over to use the new data system. A number of enhancements where made to the data transfer process, to make this transition as smooth as possible.

A project has commenced to implement virtulization work on nodes that attach to the SAM-QFS filesystem. This includes investigating full support for Linux clients.

Two STK T10K 500Gbyte tape drives have been put into production to provide improved capacity and throughput for HSM bound projects. The SAM-QFS filsystem software was upgraded to 4.6.25, and patches applied to fix NFS serving issues. Orders have been placed for 10gbit networking and associated cards for fast transfers to all nodes of the system. The firmware was also updated on the 6140 disk arrays.

Mass Data System MDSS – store.apac.edu.au

No major changes have taken place.

System Management

The Nagios systems monitoring software continues to be customised to have components for monitoring the data services.

Grid Integration

A trial installation of the File Transfer Service (FTS) has commenced on the data system. A fast gridftp transfer path has also been initiated between the dc and ac system. This is being trailed and will eventually replace the underlying transfer mechanism used within the NF specific tools: mdss and netcp/netmv.

National Facility User Workshops, Training, Education and Visits

Several of the National Facility staff attended and presented at the APAC07 conference in Perth. The meeting was focused on grid services and community support to use these services.

Dr Larson made several trips to the Bureau of Meteorology in preparation for the Hadley Unified Model.

Dr Evans travelled to numerous NCRIS workshops for Platforms for Collaboration. In particular, these meetings required presentations on the National Facility.

Other travel and visits are detailed below.

Conference and Meeting Attendance

Staff associated with the National Facility attended the following conferences in the period:

Who	Where	Why	When
Kevin Pulo	Switzerland	11th International Conference on Information Visualisation	04-06jul07
Stuart Ramsden	Lyon	CECAM Workshop 'Theoretical Aspects of Design & Periodic Materials'	16-31jul07
Ajay Lamaye	Perth	Drishti Workshop	18-21jul07
Ben Evans	Melbourne	CSIRO Advanced Scientific Computing Workshop	14aug07
Andrey Bliznyuk	London/Boston	ACS National Meeting & Exposition	15-26aug07
Vladislav Vassiliev	Sydney	Seminar / Discussion with ANU Chemistry Portal Users	22aug07
Ben Evans	Sydney	NCRIS Meeting	24aug07
Jay Larson	Melbourne	CABLE Workshop & Collaboration with Bureau of Meteorology	29aug07
Ben Evans	Melbourne	Discussion with Monash users	17sep07
Robert Davy Jason Ozolins	San Diego	IEEE Mass Data Conference	23-28sep07
David Singleton Kevin Pulo	Adelaide	Presentation of MPI Courses at SAPAC	25-26sep07
Jay Larson	Melbourne	Access University Meeting with UKMO Reps / BOM	25-26sep07
Drew Whitehouse Stuart Ramsden Ajay Limaye Rhys Hawkins	Adelaide	OZVIZ 2007	28sep07
Ben Evans David Singleton Robin Humble	United States	Supercomputing 2007 Conference / Texas Advanced Computing Centre	9-13oct07
Ben Evans	United States	NCSA	14-15Oct07
Ben Evans	Melbourne	NCRIS BOM Meeting	18oct07
Margaret Kahn	Brisbane	Bioformatics Australia Conference	22-24oct07
Jay Larson	Melbourne	CSIRO / BOM CAWCR Modelling Workshop	27-30Ooct07
Ajay Limaye	India	Training Drishti Software	11-16nov07
Rika Kobayashi Ivan Rostov Andrey Bliznyuk Vladislav Vassiliev	Melbourne	Molecular Modelling 2007	27-30nov07
Ajay Limaye		Presenting Drishti Course at Graphite 2007	30nov07-2dec07

There was a large number of meetings that took place over the Access Grid during this period, particularly relating to NCRIS arrangements.

Courses during July-December 2007

Programming course at SAPAC, Uni of Adelaide, Pulo & Singleton Presenters – 25 & 26 Sept
Chemistry Intro Course, Sydney Rika Kobayashi, 8-9 Oct 07
Drishti Software, India, Ajay Limaye, 11-16 Nov 07
Drishti at Graphite 2007, Ajay Limaye, 30 Nov-2 Dec 07

Visitors during July - December 2007
Amanda Doolan & Neil Hurst, Dell Presentation – 5 July 07
James Kelly & Kent Winchel, IBM Presentation – 6 July 07
Platform Computing HPCC Visit, Dell – 9 Aug 07
Blue Gene, IBM Presentation – 17 Aug 07
Mike Rezy, SGI/TPAC Visitor - 4 & 5 Sept 07
Raj Das, SGI Storage Visit – 19 Sept 07
DEST Visit – 24 Sept 07
Australian Astronomy VO Services Meeting -17 Oct 07
ANU Media and Communications Presentations – 3 Dec 07
MAC Meeting – 5 Dec 07
Andrew Tridgell, IBM – 12 Dec 07

Staff Activity in the Period

The major focus of activity in the period was a continuation of assisting users with new codes and better use of user codes on the system.

	indicative breakdown of overall EFT by activity
• Assisting users with optimization or consulting on codes	40%
• Helping users with general problems and use of software	10%
• Investigation of and familiarisation with third party software	5%
• Developing scheduling software and system optimization	10%
• SGI related, including systems, management and application packages	5%
• System administration	10%
• Developing and delivering courses, visiting partners	5%
• Development of projects on the mass data storage system	5%
• Investigation on new acquisition of systems	5%
• Negotiating software licenses, installing software	
• Work on National Facility documentation, communications	
• MAC and Partner Allocations servicing	
Total in these items	5%

Table of Staff Involvement for the Period

Staff Member	EFT over the Period
Antony	0.5
Amos	0.2
Barlow	0.2
Bliznyuk	0.8
Brown	1
Davy	1
Evans	0.8
Gohar	1
Holder	0.8
Humble	1
Hungerford	0.2
Jenkinson	1
Larson	1
Kobayashi	1
McCabe	1
Ozolins	0.5
Pulo	0.5
Rostov	1
Simpkins	1
Singleton D	1
Total EFT persons providing service	15.5

Appendix 4: National Facility Operations Report – January 2008 to June 2008

This Appendix outlines the systems in the National Facility and the systems and user support provided by ANU staff working in the National Facility during the period January 2008 to June 2008.

Merit and time Allocations

Merit Allocation Scheme (MAS)

The Merit Allocation Committee met in Canberra on December 11, 2007 to finalise the assessments and to grant resources to projects for 2008. The arrangements for 2008 considered the reallocation of partner shares for the period Jan-June, and increase of MAS time to 74% of the total resources. The Merit Allocation Process for Data projects (MAS-D), which included support for Data Collections of National Significance, Data grid projects, and other data projects of significant complexity and research merit that required the resources of the National Facility. These projects are reported under the section on data activities later in the document.

Committee Time Allocation Summary

The Merit Committee Chairman reported to the Steering Committee. In summary, the time granted for 2008 as a percentage of the overall MAS share was:

	2006	2007	2008
Top 10 projects	42%	40%	45%
Top 20 projects	60%	58%	62%
Top 40 projects	78%	80%	82%

The MAC strongly recommended that applicants should combine research teams (eg academic staff, postdocs and PhD students studying the same project in the same research group) into single projects. The committee recognised that sometimes there are arguments for facilitating career development of a postdoc by encouraging them to submit their own application, but it made for a complex evaluation process in terms of identifying hidden overlaps. The National Facility accounts process now includes a mechanism to help such projects organize themselves within a super project. This will facilitate research teams make these organisation changes.

Reports on Individual Projects

Principal Investigators of Merit Allocation Scheme projects are required to submit a report on the project each year. A call for Annual reports for the 2008 period is about to be made and due in May. These reports are available on-line at http://nf.apac.edu.au/annual_reports/.

A list of all MAS projects including brief descriptions can also be found at http://nf.apac.edu.au/accounts/grants/

Grant and Usage Date

In this period the National Facility serviced a total of 224 computational projects through MAS and other shares with 935 researchers. There were an additional 77 researchers associated with projects but who did not have a login-names, bringing the total number of researchers involved with the projects to 1012. There were 32 new MAS projects awarded time starting in 2008.H1 and 11 new start-up projects.

The tables below present the overall number of projects and researchers and their grants and usage during the period broken down by projects under the MAS and Partner shares. More details on the individual projects are available on the NF website. It should be noted that some researchers have both MAS and partner share project. Projects may also have more than one source of grant.

Of these projects, 157 had support from ARC or external grants. (Note that CSIRO projects are not eligible for ARC grants).

The MAS was 1.52 times over-requested at the start of the period. The committee granted 1.1 times the formal MAS share. Additional MAS Data (MAS-D) projects requested some additional resources on the peak facility.

Overall Usage of Computational Time

Division	No. of PIs	No. of Projects	No. of Users with log-in Names	No. of Researchers (total)	No. with ARC Support	No. with NHMRC Support	No. with NCRIS Support	No. with Industry Grant	No. with Other Funding	Grant as % of Total Resource Available	Usage as % of total Resource Used
MAS	147	147	731	788	85	4	1	9	36	80.9	73.3
Other Shares	72	77	330	356	31	3	1	4	12	30.8	26.6
Data Projs	4	5	18	18	3	0	0	0	1	0.1	0.1
Start-up	12	12	20	25	0	0	0	0	0	0	0
TOTALS	219	241	983	1065	119	7	2	13	49	111.8	100

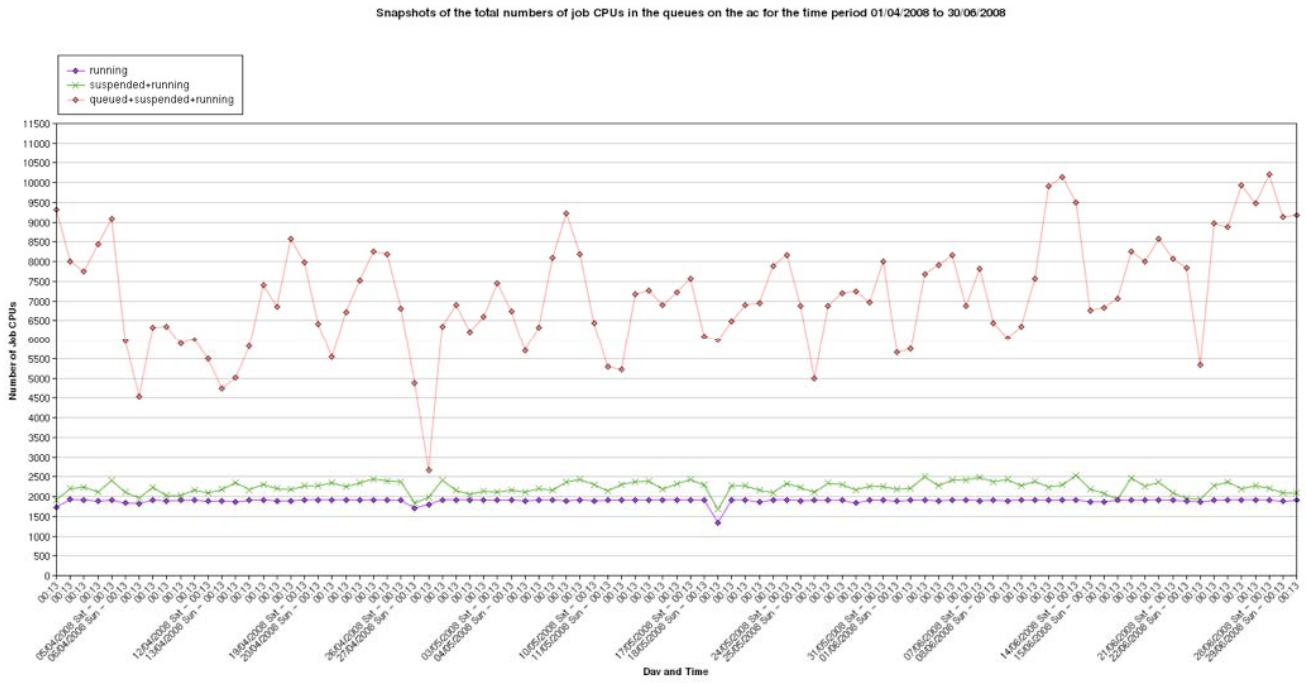
Other Shares

Division	No. of PIs	No. of Project	No. of Users with log-in Names	No. of Researcher (total)	No. with ARC Support	No. with NHMRC Support	No. with NCRIS Support	No. with Industry Grant	No. with Other Funding	Grant as % of Total Resource Available	Usage as % of total Resource Used	Format % Share of the System
ANU	42	47	253	265	25	3	1	3	7	28.2	24.5	25.00
CSIRO	13	13	39	43	1	0	0	0	3	1.2	1.3	1.15
IVEC	8	8	25	31	1	0	0	1	1	0.5	0.1	0.46
QCIF	7	7	9	13	4	0	0	0	0	0.4	0.3	0.46
Other Shares	2	2	5	5	0	0	0	0	1	0.5	0.4	0.00

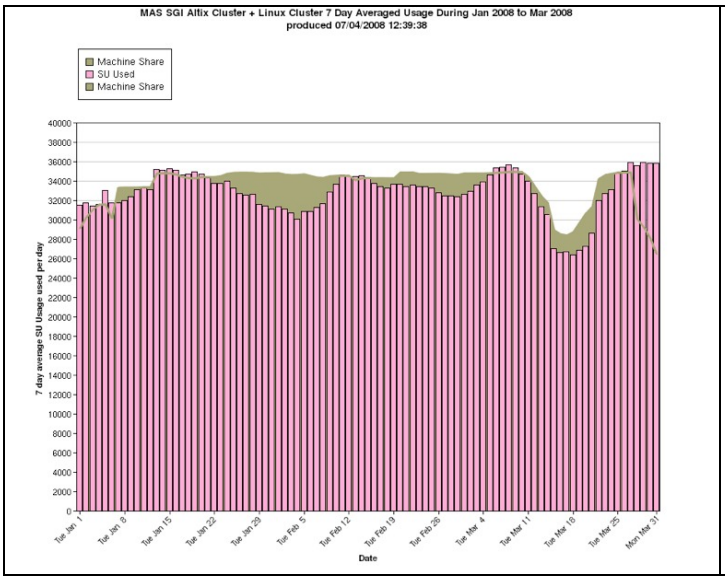
Note: Totals represent unique PI, researchers, and users.

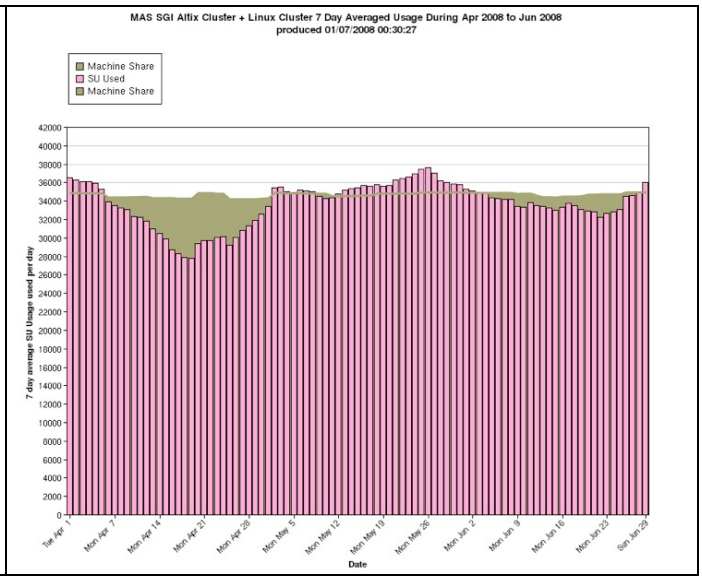
Overall Usage and demand for the System

There is now a shortage of cycles and consequently there are long queues. The following graph shows the amount of queued work on the system in the second quarter of 2008. As is shown in the graph, there has been approximately 3.7 times the number of CPUs queued on the system than available, rising to a peak of 5 times oversubscribed at the end of the quarter.

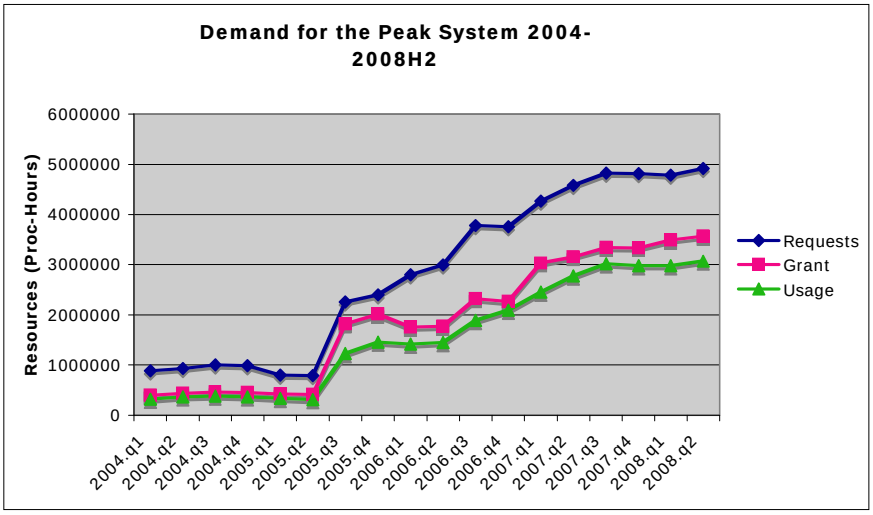


The MAS share has been fully used on the system over the whole period as shown in the graphs on the next page. Overall, system utilisation was an average of 97.4% utilisation for the period, which is extremely high, and is nearing to capacity of the system for the current job mix. The remaining cycles could potentially be used with more small single CPU jobs.



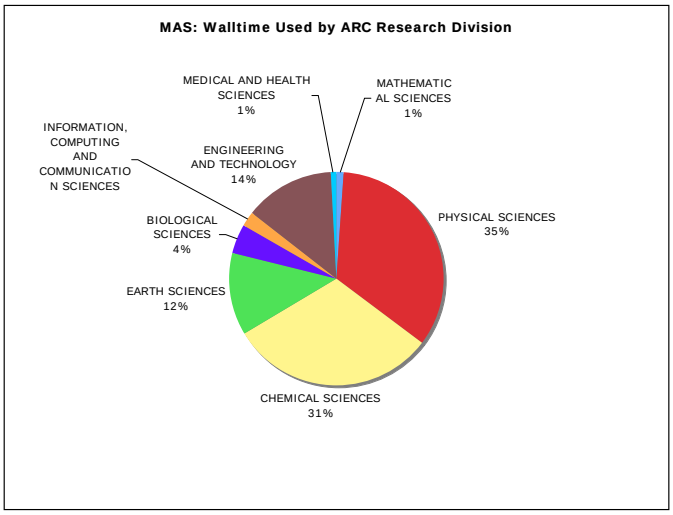


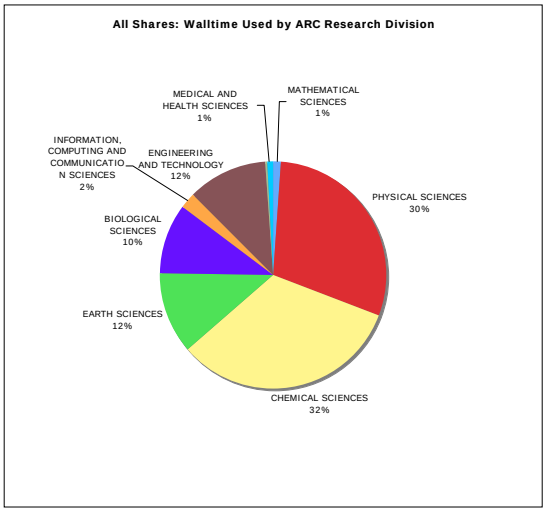
The following graphs show the demand for MAS share over a longer period; 2004 to present.



Usage by Research Field

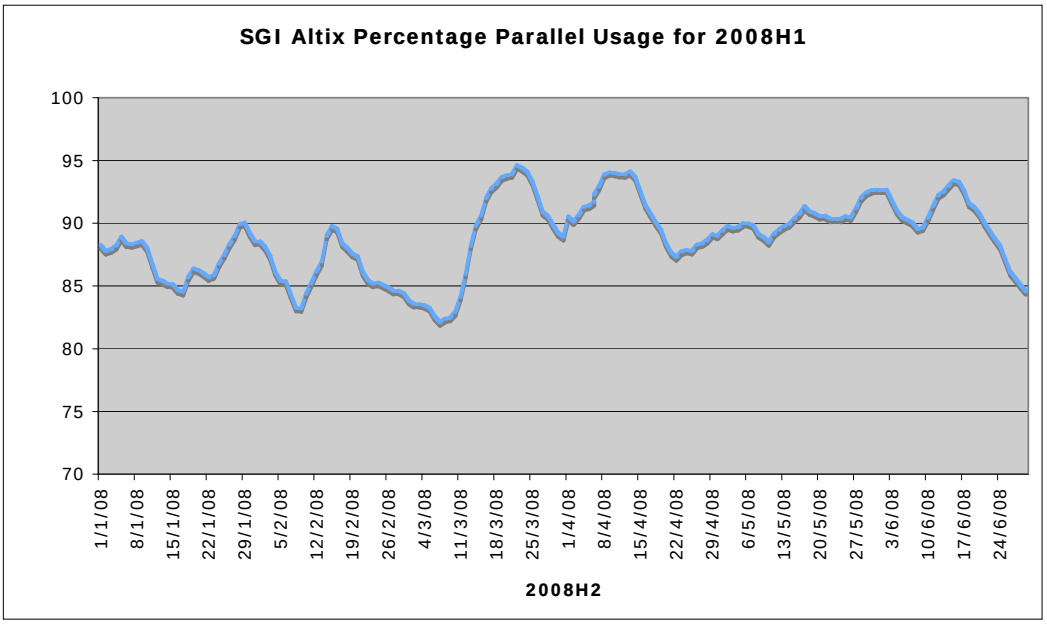
A summary of the usage by research disciplines during the period is given in the two pie-charts below. The first is restricted to MAS computational projects only while the second shows usage over the entire user base.





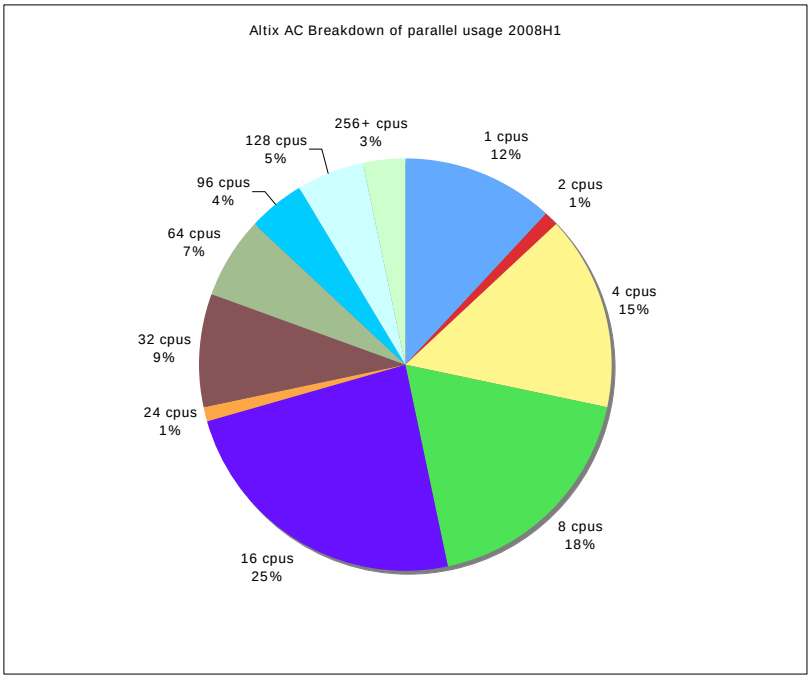
Measure of Parallel Use of the AC

The following graph indicates the percentage of the ac system resources used by parallel jobs. The NF encourages users who might otherwise run an ensemble of tasks as a parallel task, to run the jobs serially to allow the scheduler to make maximum use of the system. The percentage of parallel usage has fallen slightly but this is an artefact of the increase in the overall system utilisation. The increase in utilisation is particularly driven by new large scale projects, some of which require fast single CPU performance for analysis along with the fast performing global filesystems and well managed large memory management.



Breakdown of Number of Processors Used on the AC

The following pie-chart shows the fraction of the total usage consumed by jobs making use of the number of processors. The number of CPUs or processors per job increases clockwise from the top, generally in multiples of 2, from 1 CPU / job to 256+ CPUs / job.



Support and Development with Research Communities and Projects

The following communities have received particular attention in this period.

Biology and Bioinformatics

Through support to Karin Kassahn (Ragan, d85 MAS) a problem with the 64-bit version of one of the routines for was found. The final tree summary is now being completed using the 32 bit version on the ac-x86 node of the ac.
Beast (Bayesian Evolutionary Analysis Sampling Trees) 1.4.6 was installed. Blast was updated to 2.2.17, mpiBlast to version 1.4.0.

Chemistry

JMolEditor: The JMolEditor program that was developed under APAC has continued under funding from ARCS until mid-year pending the status of the NCI CT&T subprogram. After discussions with key computational chemistry users of the National Facility (in particular, Prof. Julian Gale, Dr. Michelle Spencer, Prof. Jill Gready, Dr. Karl Jalkanen, Dr. Rob Stranger, and Dr. Joachim Mai), a project plan has been developed that is focused on the advantages and target areas of the program with the lead developer, Dr Vlad Vassiliev of ANUSF.

Consistent with that program the following changes have been made:

- Support for visualizing the input and output file for popular Quantum Chemistry program, ADF.
- Interactive Charting to analyze Gaussian, ADF and Vasprun.xml output files
- Animation of vibrations for frequencies calculations for Gaussian and ADF
- Visualising displacement vectors for frequencies calculations for Gaussian and ADF
- Visualising plots of vibration spectra for Gaussian output files
- Export and import of data between a program and system clipboard
- Support for visualizing the SIESTA input files as well as output of coordinates in SIESTA format, needs some more testing with a wider array of data sets
- Updating POV-Ray functionality
- Multiple bug fixing, program improvements and updating help pages
- Initial implementation of support for GULP input files and animation of GULP MD trajectories
- Interactive charting is implemented for analyzing calculated results for Gaussian, ADF output files and Vasprun.xml files (VASP).

There are plans for adding further support for GULP, SIESTA and VASP (as requested by Prof. Julian Gale, Curtin Uni, Dr. Michelle Spencer, RMIT); visualization of vibration spectra and animation of vibration for Gaussian and ADF calculations (Dr. Karl Jalkanen, Curtin Uni, Dr. Rob Stranger, ANU); and new functionality for creating high-quality graphics for publications interfacing with Pov-ray (Prof. Jill Gready, Babu Kannappan, ANU).

JMolEditor has been down loaded 1332 times from 57 countries in April-June (3 months).

Other chemistry support:
The Quantum Chemistry package GAMESSPLUS was installed on ac upon request of Junming Ho. It is based on GAMESS-US and has a number of extra features, especially in area of solvation and DFT calculations.

Gaussian 03 Revision E01 (g03e01) was ported on the Intel EM64t platform (Altix Xe machine) using the latest Intel compilers and MKL libraries. As well as passing the Gaussian quality assurance acceptance tests, this installation using the Intel software has been demonstrated to provide a substantial performance improvement of the alternative to the standard supported installation using PGI compilers. As part of this porting exercise, a bug was isolated in the Intel ifort optimsation and submitted to Intel. Once a fix is in place the code will be recompiled with the higher optimised routine.

General chemistry helpdesk support was provided to users from across all university research institutions and CSIRO. In-depth help was provided to to Ekaterina Izgorodina (Monash), Parthapratim Munshi (UWA), Brian Salter-Duke (Monash), Leaf Lin (ANU) and Peter Cummins (ANU). This was particularly related to the best use of chemistry methods and their impact on the system resources.

A number of chemistry packages were updated.

- LAMMPS version 21 was installed.
- Amber-10 was ported on AC, LC, and Xe
- Abinit was updated to version 5.4.4
- Xplor-NIH package was installed on the AC.
- Towhee was upgraded to version 5.2.11 on AC
- GAMESS-US was updated on AC and LC
- Molpro was ported to the Altix Xe.
- Fixes were applied to ADF to resolve a license issue
- CASTEP licensing was resolved.

In this period there was also considerable ongoing work towards organising the WATOC2008 conference to be held in Sydney, September 14-19 2008.

Climate Modelling

Stephen Jeffrey (n71 MAS) is now running a variety of climate models concerned with developing climate change simulations for regional Australia. We have provided assistance with profiling the performance of the CSIRO Mk3L model, and with the OpenMP verison of POM. This is not complete yet and there is more to be done on porting the threaded version.

MAS project e14 - Matthew England's coupled climate model studies. There has been ongoing support for Matthew England's project (e14) with better tuning of CCSM3. Project e14 uses coupled climate models such as CCSM3 to investigate long term trends in the earth's climate and particularly the links between ocean, atmosphere etc in the Southern hemisphere. CCSM was also ported to the Xe system in preparation for the new development cluster.

Engineering and Fluid Mechanics

An analysis of performance was undertaken to resolve issues for Seyed Mostafavi (MAS f77).
There was also work to understand better configuration options to force more use "out of core" methods for Abaqus as opposed to standard method of using virtual memory. This has meant that new larger calculations can be performed more efficiently. LSDyna was updated to the current version.

Plate Tectonics

There were no new updates this quarter but we are expecting an update to underworld in the near future.

Medicine

The project by Cherbui (ANU, j66) is now performing comparisons of magnetic resonance brain images. To undertake this required the installation of a new version of FSL for Nic Cherbuin. There was a modification made to the code's use of temporary files to run more effectively on the system.

Linear Optimisation Problems

The project by Ting Yu has now become a MAS grant (ba0). This financial modeling project involves solving an extremely large optimization problem that needs to be solved a few times per year; in sync with the release of the relevant data products. There has been a detailed investigation on appropriate methods in parallel optimization to address this problem, which included an investigation of some new developments in the research domain. Rob Wormersley, a mathematician at UNSW with expertise in this area has been investigating some of this work. The problem is concerned with modelling the global flow between industries and this was discovered that it could be reduced to a very large constrained optimization problem requiring a large amount of memory.

Social Sciences

Assistance was provided to Mai Pham (ANU, p09) to understand the use of Stata to perform large GLLAMM analyses of hospital admission data. The project is to investigate differences in the quality of care between private and public hospitals in Australia. The input data set is very large and the regression analysis carried out using the Stata module, gllamm can take a long time to converge.

General tools and Other packages

Software support included reporting of compiler bugs to Intel which resulted in corrections in the next versions of the Intel compilers.

A range of software on the system is regularly installed and updated as new versions of packages, compilers and libraries become available. The full set of software available can be found on the APAC software registry http://nf.apac.edu.au/facilities/software.

System Modifications

SGI System AC – ac.apac.edu.au

The SLES10 roll-out of operating systems has been successfully completed and all porting issues resolved. This included the National Facility building its own kernel to overcome "out-of-memory" problems with the standard SLES10 kernel under the NF's demanding production load and scheduling regime. The job migration facility that was developed as part of the SGI-ANU collaborative agreement is now being used for manual scheduling. Automatic use of this facility is currently being implemented in the NF scheduler. The reconfiguration of the nodes from 32cpus to 64cpus, including reconfiguring the I/O connectivity, has also been completed. The increase in node size has meant storage dedicated to a node for local scratch could be used more effectively. Some AC nodes have had their local scratch doubled while some of the released storage has been retasked to the XE system.

Unfortunately system stability continues to be an issue. There were 5 instances of a single node failure resulting in a large fraction (usually 80%) of the system either crashing or requiring a reboot with the resultant loss of work. The underlying cause of these problems is intrinsic to the design of the high-speed interconnect of the Altix system and is unlikely to be completely resolved. On the positive side, virtually all long-term issues with the cluster filesystem, CXFS, were resolved by the upgrade to SLES10. A workaround has been provided for the one outstanding issue.

The number of very large memory single cpu jobs requested by users has increased noticeably. Only a limited number of such jobs can be cost-effectively satisfied and while users are being encouraged to look at more scalable use of the system, there is a need to reconfigure the memory. After analysing the usage characteristics of the system the plan is now to increase the memory on a small number of nodes by moving memory from some of the other large nodes and returning them to a default memory configuration of 2Gbytes/CPU. The NF scheduler will also be further modified to try to better schedule large memory jobs.

Dell Linux Cluster LC - lc.apac.edu.au
No changes.

SGI Altix XE Linux Cluster
The XE system is now in production use and available for some users to begin testing applications. Not all applications are currently available, though there has been steady progress in porting the main applications and resolving issues as they arise.

Gaussian has been ported and tuned on the system and runs significantly faster than on the AC. VASP has been installed, and some issues with performance with OpenMPI were diagnosed and resolved. A number of other issues with computational chemistry packages were debugged including Molpro, and ADF.

Two additional nodes were purchased from Xenon systems for login and metadata servers to complete the XE hardware configuration for production use. The system has used disk storage released from the Altix system through the reconfiguration from 32 to 64cpu nodes, by reusing the storage that was dedicated to local storage.

Lustre has been upgraded to 1.6.5.1, which has resolved some outstanding performance and functionality issues. This latest release simplified OS support for modules in the RHEL5 Lustre kernels. NF staff are now well versed with the Lustre configuration and are ready for its deployment in a larger production system. The first system of this nature is the development cluster which is currently in preparation.

The next phase of Lustre testing will involve using diskless nodes rather than using dedicated local scratch. Already some issues have been resolved, such as resolving Out of Memory (OOM) jobfs deadlock scenarios. This will release storage to be used more efficiently within the system. It requires additional changes to the ANU PBS queuing system to manage this configuration, which are being developed and tested on the lc.

Lustre is also being tested for its possibility to support a swap file directly on the cluster filesystem. Lustre quotas have also been switched on to understand their impact on system management and performance and to prepare for wider production usage.

We have developed a monitoring daemon for the Infiniband traffic and integrated it with the bobMon cluster monitoring application. Some investigation has taken place to improve the hardware environmental sensors of the system.

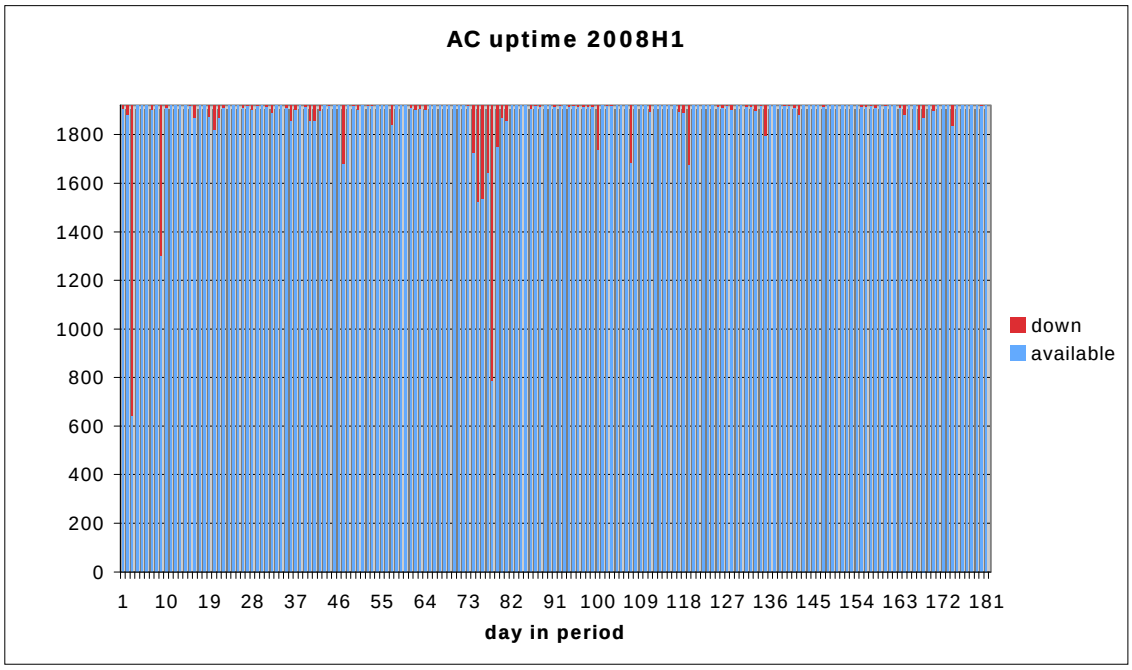
A significant amount of work has been done to update the system administration processes, with a view to making the development system easy to install and make available to users.

Grid Integration
Additional data transfer nodes have been installed and tested to improve the performance of fast parallel transfers from the AC to the DC using gridftp. The APAC/ARCS Grid gateway was updated in the first quarter as part of ANU involvement with ARCS. Only 6 minutes of computational jobs have been logged for 2008H1.

A CXFS native x86_64 client has been installed to set up gridftp software to support data transfers to the data cluster. This is discussed in more detail in the Data section of this report for system changes to the data cluster.

Operational Data

The AC system was available for 97.40% of the total time theoretically available. A graph of the system uptime is shown below. The theoretical maximum assumes the full system was 100% reliable and that no downtime was required for system management during the entire period.



The major outage in this period occurred in mid March due to problems with the internal system routers. The failed components were finally diagnosed and replaced. All downtimes are reported on http://nf.apac.edu.au/facilities/ac/downtime.php

Data Projects and Access to Datasets

The Merit Committee now allocates resources to data projects under the MAS-D scheme. Data projects applied to the MAS-D through the standard MAS call.

The committee originally approved 25 MAS-D granted projects, which was raised to 31 over this reporting period. 17 of these projects are fully established, while another 14 are considered in start-up phase as their needs are being established. These projects are supported by a total of 14 ARC grants, and 4 additional projects are supported through other granting bodies. There are 120 researchers with login access to the data, and a further 21 researchers who play an associated role. Many of these projects have a much larger community who access the data on the system through other authentication mechanisms so the total community services is much larger.

In addition there are 6 significant dataset clusters that are being hosted as part of the hosting of Nationally Significant datasets, most of which are currently to support computational access.

Data Allocations

The following storage allocations were made for data projects. Projects requiring computational time on the Altix system and the grants and usage have been included in the computational resource usage table.

Database storage

Projects	Requests(GB)	Granted (GB)	Maximum Usage(GB)
MAS-D (Data Projects)	8561	8558	2362
MAS-D (Data Start-up projects)	140	140	1
Totals	8701	8698	2363

The relational databases are typically being used in conjunction with other parts of the storage infrastructure. They may be used for data collection information, metadata housing, fast data methods for indexing into large storage areas, or managing pipelines.

Online Data Storage (guaranteed disk resident)

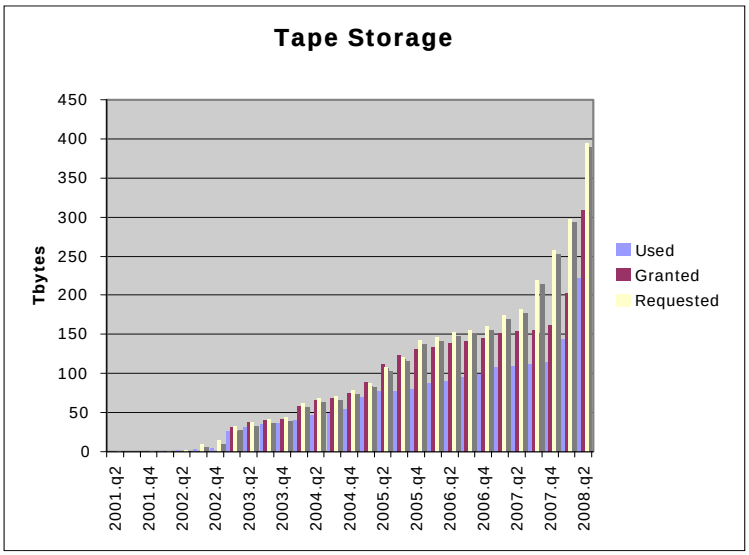
Projects	Requests (GB)	Granted (GB)	Maximum Usage (GB)
MAS-D (Data Projects)	12524	1960	31.28
MAS –D (Data Start-up Projects)	4862	2814	433.73
Totals	17386	4774	465.01

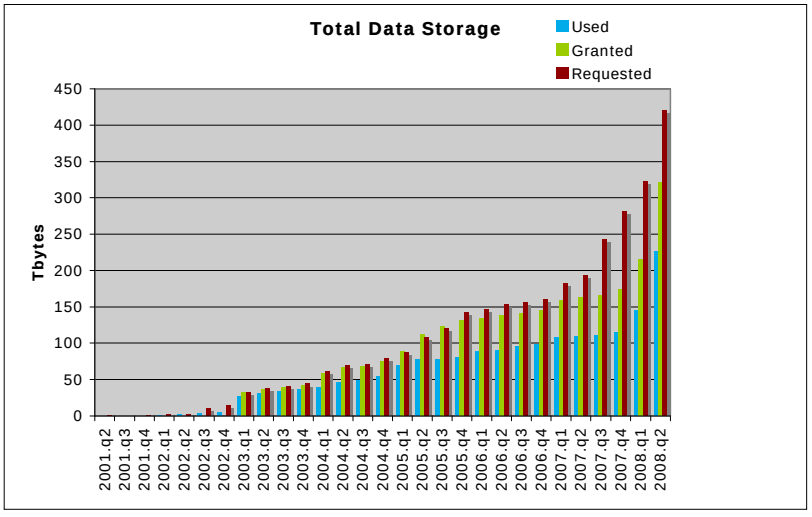
The account processes are still to be finalised and expect these numbers to be revised.

Data System HSM (tape) storage

The following table and graph shows the requests and use of tape storage space by the MAS-D, MAS-D start-up projects, and also the storage required for computational projects from MAS and partner shares. The figures represent the first copy only (two separate copies are kept of all HSM data, so that total amount stored on the system is approximately double that shown).

Projects	Requests (GB)	Granted (GB)	Maximum Usage (GB)
MAS-D (Data Projects)	165.6	77.5	65.6
MAS –D (Data Start-up Projects)	2.5	2.5	1.0
Dataset hosting	15.8	16.2	2.5
MAS and Computational Share Projects	210.7	212.3	153.4
Totals	394.6	308.5	222.5





Support for Data Communities and other Data Projects

The following communities have received particular attention in this period.

Astronomy

Primary activities in this area revolve around the service of IVOA Standard services such as hosting the first Australian implementation of a IVOA ConeSearch and Simple Image Access Protocol (SIAP). The services are now deployed on the data cluster, backed up by data served by the PostgreSQL server on db8, and the massdata archive. Ingest of the data to be served by this service has commenced and is in progress (ultimately there is expected to be up to 14TB, and perhaps 100s of billions of rows involved, and total data ingest will take some months).

The movement of the massdata archive from store to dc, required testing and repair of associated web and ftp services. Reorganisation and rationalisation of the various massdata trees associated with d01, d80, and x60 was also performed with a view to collapsing all data into the d01 project.

Some effort was devoted to investigating debugging issues with the implementation of OpenSkyNode (MAS-D u32), with the help of SkyNode project staff at Johns Hopkins University, however problems remain with the catalogue matching operations.

The Southern Sky Survey (Schmidt, d01 MAS-D) is preparing for operation in 2008. The data analysis pipelines to support the project has involved significant assistance to ensure that the project can make effective use of the batch systems, and the supporting software is in place on both the compute and data systems, and related infrastructure including database support and online/nearline storage is in place on both the data systems. Data grid software has been deployed to become part of the overall workflow management of the project.

Environmental Studies

The large MODIS dataset is now being trialed on the data cluster and is used to investigate some interest within CSIRO and Geosciences for computational projects.

Climate Data

To support the computational project (MAS, m19) we have set up processes for housing primary datasets from high resolution spectrometers housed in Wollongong and Darwin. The data is shipped by disk from Darwin to Wollongong and we have now set up a network transfer from Wollongong to the NF. The bandwidth between Wollongong and the ANU is very limited and it is best to cache the data at the ANU for the computational analysis. The Wollongong data is part of a worldwide network of similar spectrometers that produce as the final data product atmospheric composition of trace species involved in climate change (eg ozone methane, CO2), air quality (eg CO, NO2), and human health (eg formaldehyde). The stored data consists of both the raw data (interferograms), and spectra.

This data will also be replicated to Caltech. This process is similar to the existing support for LIGO data replication to Caltech and we are planning to simply reuse this method.

Other Research Group Support

Special assistance was also provided to projects due to the transition of the service from the MDSS facility to the new Data Cluster architecture.

Dataset support

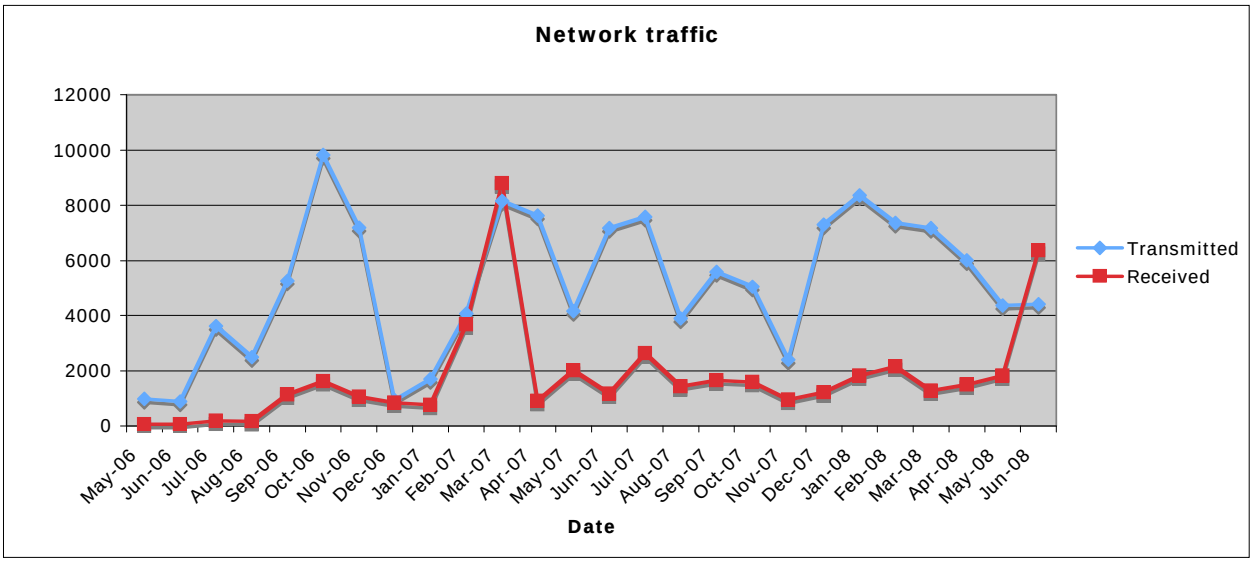
Discussions have taken place with the NCRIS ANDS support people as part of the ORCA project on the nature of datasets hosted at the NF. In particular they would like to harvest dataset metadata hosted at the facility. This will improve on the existing metadata harvesting service that was developed in 2007.

Network Traffic

The following graph shows the data transfers to and from the National Facility. The facility has become a national resource for data access. Computational output from the AC appears to mostly be transferred to the Data Cluster and some of these datasets are then accessed through services running on the Data Cluster. An example of this is OpenDAP used for a range of datasets by the climate community.

Nearly all data transfers were on-net, with data transmitted mostly domestic on-net, while the increase in data received has been largely international on-net. The new pilot project with Geoscience Australia has required new accounting methods to be able to directly recover costs for off-net traffic and is the major cause of traffic off-net. Geoscience Australia were invoiced to recover the data transfer costs, which was agreed as part of their trial access to the facility. The maximum data transfers in any one day for the period was 957 Gbytes received on 27th June, and 782 Gbytes transferred out of the facility on the 8th March.

A new 10GigE switch has been installed to increase internal data transfer capability. By the end of the next quarter this new Force10 switch will be linked into the in the main Cisco router. ANU Netcomms has provided a 10GigE uplink into the ANU core router to test and prepare for wider 10Gbit connectivity. Discussions are underway with AARNet to provision a 10Gbit link for on-net traffic, an upgrade over the existing 1Gbit link.



System Modificiations

Data Cluster – dc.apac.edu.au
Nearly all data services, packages and software have been migrated across to the new data cluster. The new architecture allows for better dedication of resources toward services across nodes rather than on a single host as was the case on the MDSS. Only one project has not moved to the new data cluster. The transition has been seamless to users with only a few exceptions due to old accounts not in the LDAP database.

To support the very large databases (~15 Tbytes) required for projects like astronomy data, the database nodes were physically connected to the fibre channel fabric and the OS kernel modules were configured to support multi-path connectivity to the disk arrays.

Virtualisation software is being installed on the dc cluster to support complex data projects more cost effectively. Both Linux and Solaris10 QFS clients are now installed on the system. There have been some issues to resolve for the ESX server, however we expect these will be resolved with a planned firmware update to the 6140 storage controllers.

Special nodes have been dedicated to managing data transfers into and out of the new data cluster. These have been connected to the Force10 switch with 10Gbit ethernet network interfaces. Gridftp, LDR, Globus, and other grid applications have been installed as part of this service. GSIOpenSSH has also been installed for simple scp transfers.

In the interim the transfer of data between the AC and DC is via the single threaded transfer application “fastrcp”. The managed data transfer service FTS, the File Transfer Service from CERN, is also being tested as part of managing large data transfers across the data bandwidth available on these nodes. This deployment is being evaluated as standard way to facilitate data movement between computational and storage nodes, starting with the transfers from AC to DC. This will enable parallel gridftp transfers. Users will then be able to use this transparently without changing their current mode of transfer, which is by mdss and netcp/netmv utilities. However, applications that use gridftp will be able to also use this transfer service.

Three new 16Tbyte 6140 disk trays have been installed and the firmware updated on all of the 6140 trays and controllers, as well as the D280 trays and controllers. Some issues are being resolved to improve the management access to these controllers. The D280 storage array that was used by the store server will now be dedicated as a metadata device for the DC. Once this is put into production on the DC users will see a noticeable improvement of the metadata performance.

A console and management network for the DC cluster is now active and a windows management host for ESX server deployment has been ordered.

Trial installations of searching technologies have been tested on the system including Ferret, Solr, Apache Lucerne and Sphinx to investigate indexing and search performance.

To simplify the support of web hosting projects using Ruby on Rails we have investigated an apache plugin called passenger, which implements a mod_rails style plugin for apache.

Mass Data System MDSS – store.apac.edu.au
All massdata projects (except for MAS-D project f48) moved off store onto the DC cluster. The server store will be upgraded to Solaris 10 and configured as a data node of the cluster. As well as supporting the remaining project, it is still used for some legacy applications that require a big-endian architecture. In time it will be replaced by a much smaller system and itself become a node of the data cluster. Its purpose is to accommodate data projects or tools with big-ended binaries, or other projects that will take a while to transition.

System Management
No major changes have taken place. The National Facility installation of the Nagios systems monitoring software has been customised to have all components for monitoring the data services. This includes the new virtualised systems managed within the nodes of the dc.

Grid Integration
A new GridFTP server was tested and put into production as gridftp-dc.apac.edu.au.

As part of ANU contributions to ARCS, an SRB virtual machine installation has been installed called srb.dc.apac.edu.au. This integrates with the ARCS data fabric. Also as part of that contribution to ARCS, we have tested a new NGdata image (version 1.8.1) on the gateway.

In Q1, VO accounts were installed and made to work between NF accounts and the grid infrastructure. For the next period we plan to configure the GUMS server to also allow VO accounts from a new ARCS VOMRS server. The GUMS server at the NF supports the old and new VOMRS servers and is ready for any new potential VOs that come along. Whenever a new gateway system is installed it will be possible to quickly configure the GUMS server to map VOs to these accounts.

System Downtimes
The move to a clustered architecture with multiple metadata servers and virtualisation has substantially increased the ability to provide non-stop services for the broad range of data services. We are still improving the system configuration to manage potential system outages. There were two brief downtimes in April, one to resolve a hardware fault was isolated on one of the metadata servers and the other by a non-fenced installation process that reconfigured filesystem partition information. Both issues were resolved, and the former was made easier by migrating the metadata serving to a second metadata server host. The metadata servers do not currently automatically failover. A manual process is currently used to a second hot standby metadata server. We will look to implement automatic failover in the second half of this year, though may delay this to use the new method with SAM-QFS 5.

Downtimes are reported on http://nf.apac.edu.au/facilities/mdss/downtime.php

Helpdesk
In the first quarter of 2008 there were 572 emails received by the helpdesk. Around 23% were concerned with minor matters (eg. Passwords, requests for variations to job limits, minor compilation problems etc.) and the remainder being more substantial problems.

The median response time was 22 minutes and 25% had a response time greater than four hours (this includes responses arriving on weekends and out-of-hours), but only 4% took longer than 1 day.

In the second quarter there were 595 emails to help, 20% relating the minor matters. The median response time was 20 minutes and 16% had a response time greater than four hours, but only 2% longer than 1 day.

National Facility User Workshops, Training, Education and Visitis
Dr Ben Evans attended several meetings in Melbourne to discuss CSIRO/BoM interest in peak computing facilities and their management. In March 2008, Drs Evans, Singleton and Humble attended a meeting in Melbourne to start the joint tender process with the Bureau of Meteorology.

Mr Stuart Hungerford travelled to Sydney and Melbourne in relation to expanding the software support for the ASSDA data project.

Conference and Meeting Attendance

Who	Where	Why	When
Ben Evans	Melbourne	NCRIS Discussions	15feb08
Ben Evans	Melbourne	NCI Meeting with CSIRO/BOM	29feb08
Stu Hungerford	Sydney	ASSDA Meeting	18mar08
Ben Evans David Singleton Robin Humble	Melbourne	Joint BoM/NCI Tender Preparation	25mar08
Stu Hungerford	Melbourne	ASSDA Meeting	31mar08
Judy Jenkinson Kashif Gohar	Sydney	MySQL Course	31mar08-4apr08
Ben Evans	Melbourne	Joint BoM/NCI Tender Industry briefing	14apr08
Vladislav Vassiliev	Perth	iVEC & Curtin University	5may-8may08
Ben Evans David Singleton Robin Humble	Melbourne	RFT Tender Presentations	9jun-13jun08
Margaret Kahn Rika Kobayahsi Kevin Pulo	Sydney	Uni Sydney Workshops	11jun-13jun08
Ben Evans David Singleton Robin Humble	Melbourne	RFT Tender Presentations	24jun-27jun08

Table of Staff Involvement for the Period

Staff Member	EFT over the Period
Antony	0.5
Amos	0.2
Bliznyuk	0.5
Brown	1
Davy	1
Davies	0.6
Evans	0.8
Gohar	1
Holder	0.8
Humble	1
Hungerford	0.1
Jenkinson	1
Kobayashi	1
McCabe	1
Ozolins	0.5
Pulo	1
Rostov	1
Simpkins	1
Singleton	1
Watson	0.5
Total EFT persons providing service	15.5

Courses during January - June 2008
David Singleton, Margaret Kahn, Judy Jenkinson, Kevin Pulo, Rika - ANU Workshop on using the NF - May08
Margaret/Rika/Kevin - Uni Sydney Workshop - 12-13jun08

Visitors to the department during January - June 2008
Jason Lohrey - Mediaflux – 22 Feb 08
Kretsch and Liang Chen - Sun – 6 March 08
Peter Kerney - Intel – 6 March 08
Peter Helms & Colin McMurtie, Uni of Canterbury - Blue Gene Workshop - 29-30 May 08
Micah Hamad, UC Bolder – discussion on bioinformatics codes – 30 June 2008.