

HPCx Quarterly Report

April - June 2004

1 Introduction

This report covers the period from 1 April 2004 at 0800 to 1 July 2004 at 0800.

The next section summarises the main points of the service for this quarter. Section 3 gives details of the usage of the service, including failures, serviceability, CPU usage, helpdesk statistics and service quality tokens. A summary table of the key performance metrics is given in the final section. The Appendices define the incident severity levels and list the current HPCx projects.

2 Executive Summary

- This has been a very busy quarter with the upgrade to Phase 2, high usage of the system and a significant number of events organised.
- The upgrade to Phase 2 was completed on schedule with the Implementation Certificate issued by EPSRC in mid-June. This upgrade involved faster processors, additional frames, the new High Performance Switch, migration of all data and major changes to the system software environment. This was expected to be the major technical challenge of the whole HPCx project and the fact that it was completed on time and with only modest disruption to users is a major source of satisfaction and pride for all involved.
- Utilisation remained around 75% throughout the upgrade and there was significant capability usage. There was minimal lag to the usage as the system increased in performance and we expect the system to remain busy throughout the rest of this year.
- There are currently 34 research groups, which is in excess of the maximum agreed for 2004. HPCx have now submitted a revised options paper to EPSRC.
- The second HPCx Annual Seminar was successfully held in Edinburgh on 9 July. There were more than 60 attendees at the wide range of

interesting talks. The user group and new terascaling course, which were held the previous day, were also very well received.

- We also held a user workshop in London discussing initial experiences with Phase 2. This was preceded by a workshop of NERC users so that we could better understand their activities.
- This has also been an active quarter for on Outreach. The Life Sciences projects are beginning to make progress and we held an Industry Day at Daresbury in April.
- The updated User Guide was ready in advance of the start of service on the Phase 2 system; the value of this was demonstrated by the lack of additional queries on the Phase 2 system. We also produced 6 technical reports this quarter on topics including the performance of the new switch and the tape archive.
- Most user applications have been ported successfully to Phase 2 and many of these have already received initial performance tuning. Significant improvements have been made to a number of codes including LUDWIG, GAMESS-UK and CASTEP. As shown by the user benchmarks, many users are indeed seeing a factor-of-two performance improvement over Phase 1.

3 Usage Statistics

3.1 Availability

3.1.1 Failures

The monthly numbers of incidents and failures (SEV 1 incidents) are shown in the table below:

	<i>April</i>	<i>May</i>	<i>June</i>
Incidents	9	24	31
Failures	1	5	4

The following tables give more details on the attribution of the failures:

April

<i>Failure</i>	<i>Site</i>	<i>IBM</i>	<i>External</i>	<i>Reason</i>
04.072	0%	0%	100%	Network failure at Manchester

May

<i>Failure</i>	<i>Site</i>	<i>IBM</i>	<i>External</i>	<i>Reason</i>
04.092	100%	0%	0%	LL job classes not drained
04.095	0%	100%	0%	GPFS failure
04.108	0%	100%	0%	Recoverable virtual shared disk server fails to start
04.109	0%	100%	0%	Recoverable virtual shared disk server stopped after switch problem
04.118	0%	0%	100%	Loss of external network

June

<i>Failure</i>	<i>Site</i>	<i>IBM</i>	<i>External</i>	<i>Reason</i>
04.128	0%	0%	100%	External network failure
04.131	100%	0%	0%	Power failure
04.143	0%	100%	0%	Full switch failure
04.146	0%	100%	0%	Full switch failure

3.1.2 Performance Statistics

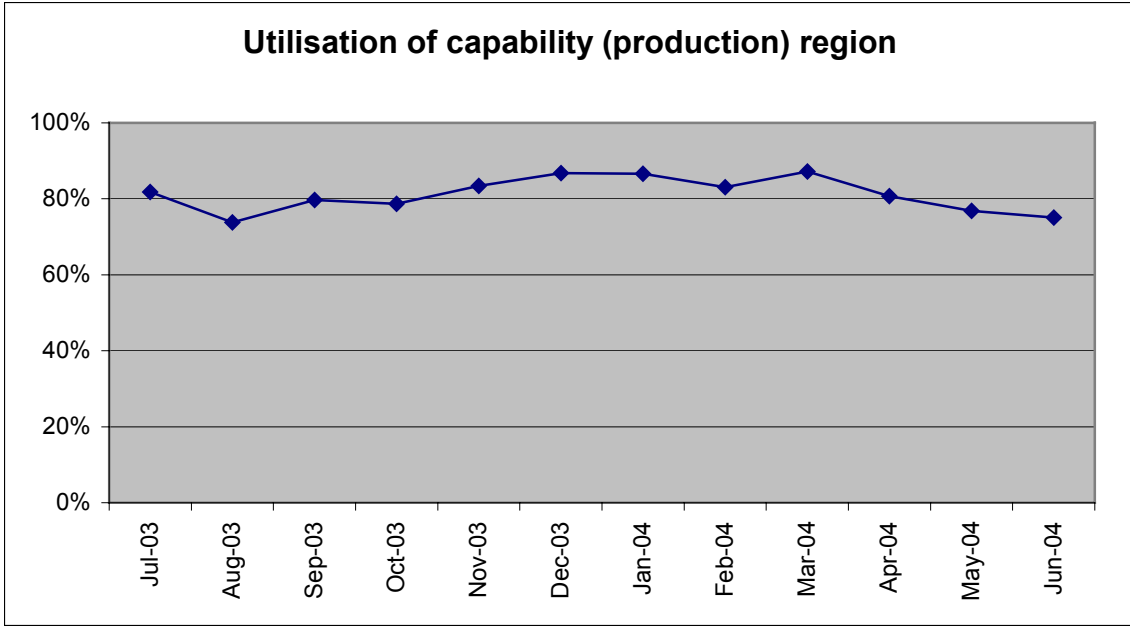
This section uses the definitions agreed in Schedule 7, ie,

- $MTBF = (24 \times 30.5) / (\text{number of failures in month})$
- $\text{Serviceability (\%)} = 100 \times (\text{WCT} - \text{SDT} - \text{UDT}) / (\text{WCT} - \text{SDT})$

<i>Attribution</i>	<i>Metric</i>	<i>April</i>	<i>May</i>	<i>June</i>	<i>Quarterly</i>
IBM	Failures	0	3	2	5
	MTBF	∞	244	366	439
	Serviceability	100.0%	98.4%	98.0%	98.8%
Site	Failures	0	1	1	2
	MTBF	∞	732	732	1098
	Serviceability	100.0%	100.0%	99.3%	99.8%
External	Failures	1	1	1	3
	MTBF	732	732	732	732
	Serviceability	99.8%	94.3%	96.1%	97.0%
Total	Failures	1	5	4	10
	MTBF	732	146	183	220
	Serviceability	99.8%	92.7%	93.4%	95.6%

3.2 Capability Utilisation

The monthly utilisation for the 1024-processor capability region is shown in the following graph. During the transition period there was no division between the development and capability regions, so for April and May the figure for overall utilisation is used instead.

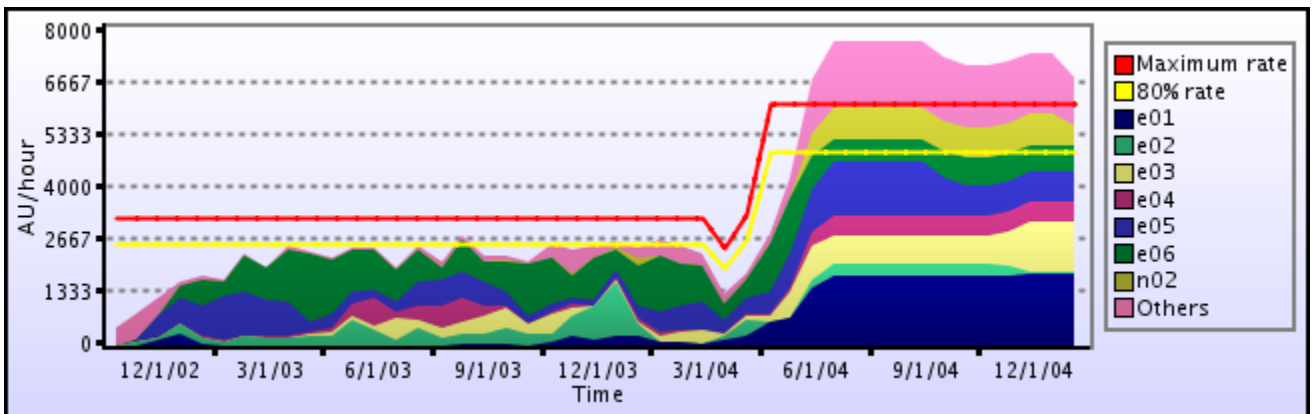


3.3 Capacity Planning

Predicted Utilisation

The following graph shows the utilisation since the start of the project and the projected utilisation until February, 2005. The scale on the y-axis is AUs per hour, where the peak that HPCx Phase 1 could currently deliver is around 3240 AUs per hour, and Phase 2 around twice that (the red line in the graph). The yellow line corresponds to the more practicable 80% level.

The graph assumes that each project will use its remaining allocation pro rata with its usage profile from the SAF, which is often simply that on the original application form.

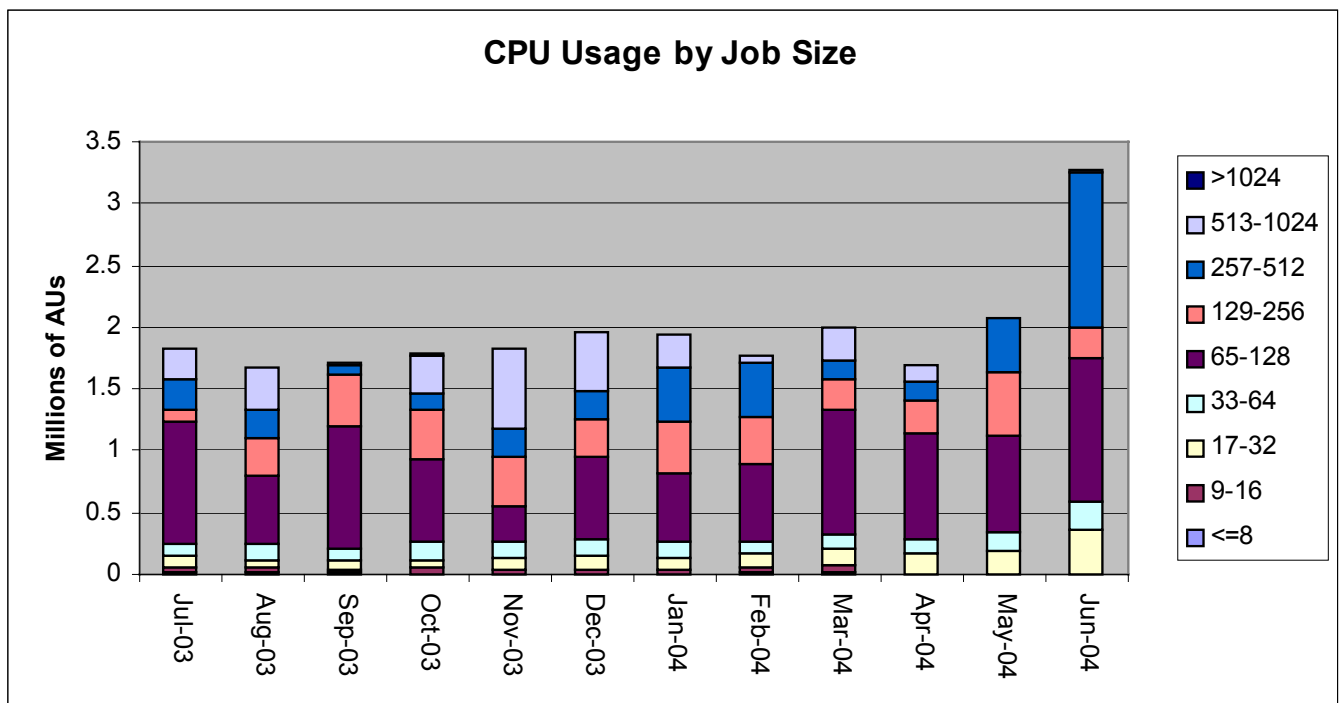


The graph suggests that the service will be substantially overloaded through the rest of this year.

Numbers of Research Consortia

There are currently 34 research consortia using the HPCx system. Two of these, however, are Class 2 projects which will be closed on July 16.

3.4 CPU Usage by Job Size



The above graph shows that in June there was substantial change towards larger job sizes, with jobs in the 257-512 band now using the most time. Capability usage, that is usage by jobs of 512 processors or more, was affected by the reduction of the system to 640 processors during the changeover period

3.5 AU Usage by Consortium

The PIs and titles for the various consortia are listed in Appendix B.

A few of the figures for May are slightly different from those shown in the monthly report for that month. This is because data for a number of jobs were not added to the database until after that report was compiled, as a result of problems during the upgrade process.

<i>Consortium</i>	<i>April</i>	<i>May</i>	<i>June</i>	<i>Quarterly</i>	<i>%age</i>
e01	65145	186756	515129	740178	13.5%
e02	1	244965	2	244968	4.5%
e03	183511	53559	231882	468952	8.6%
e04	18553	21838	41715	82106	1.5%
e05	459060	299751	388146	1146957	21.0%
e06	569702	358021	923047	1847877	33.8%
e07	1322	1666	576	3564	0.1%
e10	192			192	0.0%
e11	10037		4742	14779	0.3%
e12	474	13958	6529	20961	0.4%
e15	0	415	540	955	0.0%
e18	0	2929	6961	9890	0.2%
e20	0	0	3666	3666	0.1%
<i>EPSRC Total</i>	1307997	1183858	2122935	4585045	83.9%

n01	80054	64170	127746	271970	5.0%
n02	3432	677	4042	8151	0.1%
n03	87404	26986	3154	117544	2.2%
n04	12488	965	9661	23114	0.4%
<i>NERC Total</i>	183378	92798	144603	420779	7.7%

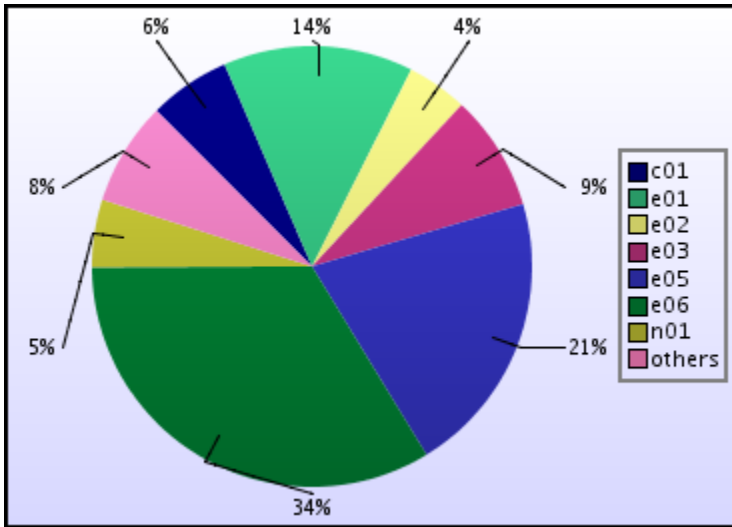
p01	0	1743	11339	13082	0.2%
<i>PPARC Total</i>	0	1743	11339	13082	0.2%

c01	167646	66137	95579	324089	5.9%
<i>CCLRC Total</i>	167646	66137	95579	324089	5.9%

b01		53	272	325	0.0%
b02	11		2354	2365	0.0%
<i>BBSRC Total</i>	11	53	2626	2690	0.0%

z001	8239	10218	54506	72963	1.3%
z002	19	18530	22983	41532	0.8%
z004			679	679	0.0%
z06	31	1811	87	1929	0.0%
<i>HPCx Total</i>	8289	30559	78255	117103	2.1%

x01		27		27	0.0%
<i>External Total</i>		27	0	27	0.0%



3.5.1 Discounts

There are now a number of user codes that have qualified for capability discounts. The following table shows the discounts that were awarded during the last quarter.

<i>Consortium</i>	<i>AUs Used</i>	<i>AUs Charged</i>	<i>Discount</i>
e01	1072895	767030	123352

3.6 Helpdesk

3.6.1 Classifications

<i>Category</i>	<i>Number</i>	<i>% of all</i>
Administrative	117	32.5
Technical	227	63.1
In-depth	14	3.9
PMR	2	0.6
TOTAL	360	100.0

<i>Service Area</i>	<i>Number</i>	<i>% of all</i>
Phase 1/2 platforms	316	87.8
Website	26	7.2
Other/general	18	5.0
TOTAL	360	100.0

3.6.2 Performance

<i>All non-indepth queries</i>	<i>Number</i>	<i>%</i>	<i>Target</i>
Finished within 24 Hours	291	84.6	75%
Finished within 72 Hours	344	100.0	97%
Finished after 72 Hours	0	0.0	

<i>Administrative queries</i>	<i>Number</i>	<i>%</i>	<i>Target</i>
Finished within 48 Hours	117	100.0	97%
Finished after 48 Hours	0	0.0	

3.6.3 Experts Handling Queries

<i>Expert</i>	<i>Admin</i>	<i>Technical</i>	<i>In-Depth</i>	<i>PMR</i>
epcc.ed.ac.uk	83	88	5	1
dl.ac.uk	4	24	3	0
Sysadm	30	113	6	1
Other people	0	2	0	0

3.7 Service Quality Tokens

There are no current quality tokens.

4 Support

4.1 Applications Support (*Dr David Henty*)

The key targets for this quarter, in addition to the ongoing programme of training and technical reports, were to provide a smooth transition for users from Phase 1 onto Phase 2 and to finalise arrangements for this year's Annual Seminar. Both of these were accomplished successfully. Although the Seminar actually took place in July, shortly after the end of Q2, it seems appropriate to report on it (and the associated User Group) here while still fresh in our minds.

4.1.1 Documentation

Before the switch-over to the Phase 2 machine happened we ensured that all the changes that would affect users were detailed in a document called *HPCx: Phase 2 Differences* which was available on the WWW in time for early user access. At the time of the changeover, this was fully incorporated into an updated User Guide for Phase 2. In addition we provided a new set of template LoadLeveler scripts for serial and parallel jobs, and updated various web pages and FAQ entries to reflect the new hardware setup, changes to the charging mechanism, the new AU rate etc.

Overall, we experienced very few issues during the changeover. Despite the complete replacement of hardware and system software, users saw an almost identical user interface on the new service. The provision of complete Phase 2 documentation right from the start of service meant that we did not see any significant increase in queries over the transition period.

4.1.2 Technical Reports

A total of four reports were due this quarter in the following areas:

- (a) Overview of the HPCx tape archive system
- (b) Low-level benchmarks on initial Phase 2 system
- (c) Practical guide to user tools
- (d) An ongoing assessment of serial efficiency of applications

We have actually produced six reports this quarter, all available at <http://www.hpcx.ac.uk/research/hpc>, with the following titles:

- **HPCxTR0404:** *Initial Experiences Porting Hydra_MPI, which requires MPI-2 remote memory access calls, to HPCx*, Paul Walsh and Gavin Pringle
- **HPCxTR0405:** *HPCx Archiving User Guide*, Elena Breitmoser and Ian Shore
- **HPCxTR0406:** *A Performance Study of the PLAPACK and ScaLAPACK Eigensolvers on HPCx for the Standard Problem*, Elena Breitmoser and Andy Sunderland
- **HPCxTR0407:** *LAPI on HPS*, Adrian Jackson
- **HPCxTR0408:** *Using TotalView on HPCx*, Adam Carter
- **HPCxTR0409:** *Planned AlltoAllv: A Cluster Approach*, Adrian Jackson and Stephen Booth

Reports 05, 07 and 08 correspond directly to the titles (a), (b) and (c) above. We decided to move the production of report (d) to later in the year (when we will have had more opportunities to evaluate the Power4+ processor) and have replaced it with report 06. This new report was originally due in Q4 in the area of *Scientific library performance on Phase 2*. Report 04 is an additional piece of work done in collaboration with one of our MSc in HPC students. The results of this report are of general interest to users as the program under examination uses advanced features of MPI that, although not yet widely used, are crucially important for certain codes that require a parallelisation model akin to SHMEM on the Cray T3E. Report 09 is a write-up of work done under the Software Engineering team which is a general, portable implementation of a terascaling technique previously reported in HPCxTR0401 (which focused on a particular application).

In the first half of this year we have produced a total of nine reports, three more than the target. Of the six reports listed for Q1 and Q2 in the 2004 Annual Plan, five have been produced as planned and one swapped with a title originally scheduled for Q4. Overall, the production of technical reports is progressing extremely well.

4.1.3 Training

In Q2 of 2004 we ran the following five courses:

- **Edinburgh, 13-15 April:** *Exploiting the Computational Grid.*
- **Edinburgh, 28-29 April:** *Message-Passing Programming.*
- **Edinburgh, 4-5 May:** *Shared-Memory Programming.*
- **Edinburgh, 25-27 May:** *Scientific Visualisation.*
- **Daresbury, 4 June:** *Using the HPCx Service.*

The statistics are summarised below alongside the annual targets (where appropriate).

<i>Metric</i>	<i>Total to date</i>	<i>Target for year</i>
Course days	22	30
Number of courses	9	
Different courses	8	12
Different locations	3	4
Student-days for HPCx users	200	
Student-days for HPCx staff	14	
Student-days available for HPCx	417	600

Note that the last three courses were run specifically for HPCx and did not form part of any other training programme. Registered HPCx users are naturally given top priority, but we additionally used these courses as an outreach opportunity to promote capability computing in general, and HPCx in particular, to the UK academic community. For such courses, all academic attendees are counted in the figures for student-days delivered.

A new course on *Improved Performance Scaling on HPCx* has been developed by Mark Bull, Adrian Jackson, Lorna Smith and Stephen Booth, and will run for the first time on 8 July, the day before this year's Annual Seminar. As mentioned in the Q1 report, we have also investigated porting the Visualisation Tool Kit (VTK) library to HPCx. This was possible due to recent upgrades to the graphics infrastructure (eg the availability of a native OpenGL implementation), and we are therefore considering running the subset of our Scientific Visualisation course that concentrates on the use of this library as part of the HPCx training activities.

4.1.4 Workshops and Conferences

HPCx Phase 2 Workshop

An "HPCx Phase 2 Workshop" was held at University College, London on 26 May 2004, preceded by a special meeting to discuss the specific requirements of NERC consortia (as requested by STAC). HPCx was represented by David Henty, Kevin Stratford, Mike Ashworth and Martin Plummer.

The NERC meeting was attended by five key HPCx users and was extremely valuable in understanding their future computational requirements. There were presentations from Andrew Coward and Paul Burton. The major issue that arose is that the UK HIGEM climate-modelling project (PI: Lois Steenman-Clark) will require significant computational resources using relatively modest numbers of processors, rarely more than 256. Scalability of the Unified Model, which uses a lat-long grid, is severely restricted by the coordinate singularities at the poles especially when running very high-resolution models. We need to consider carefully how and if we can reconcile this with the capability focus of HPCx. Although this will present something of a challenge, the meeting was vital in ensuring that we were fully aware of this issue well in advance.

The Workshop itself was attended by twelve users and Jonathan Follows from IBM. It began with three brief talks from HPCx staff outlining the key points of the Phase 2 system and presenting a number of case studies to illustrate typical performance. This was followed by open discussion, at which the overriding impression was that users were generally very happy with the Phase 2 service. The two major issues were lack of turnaround due to long queues and the fact that some users were experiencing problems in allocating all the memory that should be available to them. The former problem has gone away as the size of the system has increased to the full 50 frames; the latter was subsequently investigated in some detail and fully resolved.

Overall, this was an extremely successful meeting. The format of having an applications-focused session followed by general discussion worked very well and is something we will consider doing again in the future with a different user community, eg for the Life Sciences.

Second HPCx Annual Seminar

The theme of the seminar was *New Science from Capability Computing* and was held on Friday 9 July at the National e-Science Centre in Edinburgh. Although it followed on very closely to the major Phase 2 HPCx upgrade, we wanted to focus mainly on the science being enabled by very high-end supercomputing as opposed to concentrating on the technology. Funding for the meeting was provided by EPCC, IBM and NeSC, as well as charging a small registration fee for non-academic attendees. This allowed us to hold a reception at the Talbot Rice gallery the evening before which was attended by more than 50 people.

The Seminar attracted in excess of 60 registrations and attendees, and was an extremely successful event. We invited two keynote talks from international speakers: Dr Thomas Lippert, Director of Computational Science at the John von Neumann Institute for Computing; Dr Wanda Andreoni, Manager of Computational Biochemistry & Materials Science at IBM Research, Zurich. There were also talks from leading members of four major EPSRC and NERC user groups (Neil Sandham, Patrick Briddon, Andrew Coward and Paul Selwood), as well as two presentations from HPCx staff (Joachim Hein and Martyn Plummer / Andy Sunderland). All the talks are being published on the HPCx website.

The event was a success and we received very positive feedback from both speakers and delegates. As an example, Dr Lippert contacted us afterwards to say that “the meeting was very interesting and has demonstrated the high standard HPC has in UK”.

4.1.5 User Group

The third HPCx User Group meeting took place on Thursday 8 July, immediately before the Annual Seminar evening reception, and was attended by around 20 people. There were presentations from Steve Andrews covering the upgrade process from Phase 1 to Phase 2, and from David Henty pointing out the major issues for users. The actions from the previous meeting were reviewed to ensure that there were no outstanding issues on Phase 2. This was followed by a discussion session at which the major topic was testing new compiler releases. This may be problematic in the future if the Training and Development machine is decommissioned, but we committed to keeping users fully informed of compiler updates to help them prepare for and deal with any problems that might arise.

4.1.6 Newsletter

The third issue of Capability Computing was produced on time and distributed with the delegate pack at the Annual Seminar. We are currently organising a mailshot of some 3500 copies to the UK, Europe and Overseas.

4.1.7 Packages

All packages and libraries were successfully ported to Phase 2 without any major problems. We are currently in negotiation regarding full licenses for FLUENT and CFX, although they can currently still be run using temporary licences.

4.2 Outreach Activities (*Dr Richard Blake*)

Over the past quarter the outreach activities have progressed in the following areas.

4.2.1 Lifesciences Projects

The Lifesciences projects at Bristol University, Oxford University and the John Innes Centre are now well underway. Support requirements for the other two projects will be developed during July. In these quarterly reports we will present major updates on various of the projects in turn.

Towards a Virtual Outer Membrane: Prof Mark S.P. Sansom & Dr. Jorge Pikunic, Dept. of Biochemistry, University of Oxford

Our work is aimed at progressing molecular dynamics simulations of biological systems from the single molecule to the sub-cellular level. We see this as an essential first step towards biomolecular systems biology (i.e. integration between different levels of description of complex biomolecular systems) [1].

Capability computing is essential in achieving our goals for this project. Therefore, our first step is to select the molecular dynamics simulation package with best scalability for the size and type of systems that we are studying. We are collaborating with Dr. Joachim Hein, from EPCC, in measuring the performance of the available simulation packages (Gromacs, NAMD, CHARMM, Amber, and DL-POLY) on HPCx. We have selected the outer membrane protein BtuB, the vitamin B₁₂ transporter in *E. Coli*, as a benchmark system. The atomic interactions involved in this system resemble those that we expect to have in our model outer membranes, making it a suitable benchmark. Aside from this, the function of BtuB and other TonB-dependent transporters is not well understood. These proteins are of biomedical importance as potential targets for new antibiotics, due to their ubiquity. To set up the system, we started from the crystal structure of BtuB [2], modelled the missing residues, estimated the charges of all ionisable groups, inserted the protein in a lipid bilayer of DMPC, and solvated with water. The size of the resulting system (see Figure 1) is ca. 100,000 atoms. We have performed ca. 10 ns of molecular dynamics simulations of this system starting from the crystal structure of the protein in presence and absence of vitamin B₁₂. We are currently analysing the trajectories to investigate the mechanism of trans-membrane signalling upon substrate binding.

We are also working on developing a model of an extended 'patch' of outer membrane, of up to 0.5 million atoms. This involves selecting an array of proteins with biological meaning. Our next steps are to complete the performance study on HPCx, and carry on molecular dynamics simulation of our resulting outer membrane model.

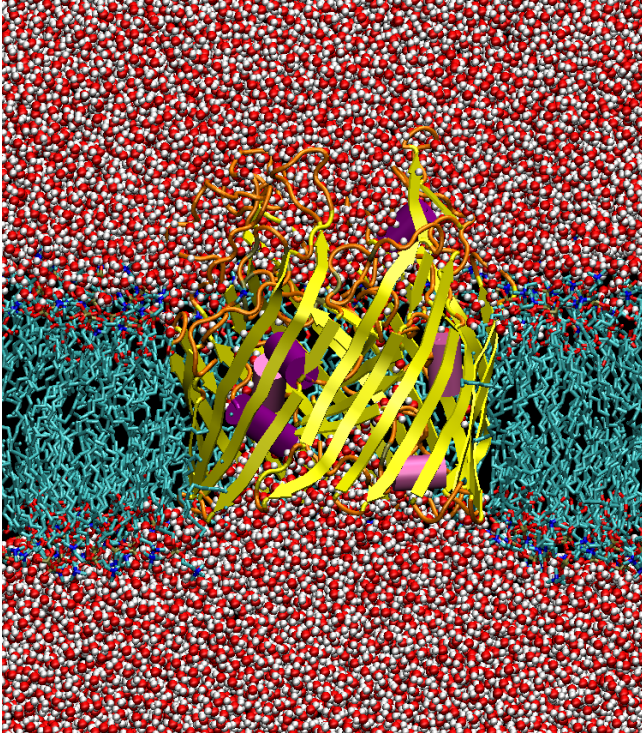


Figure 1. Benchmark system: BtuB (shown in cartoon representation) embedded in a lipid bilayer, solvated with water.

References.

- [1] Arinaminpathy, Y., Biggin, P.C., Bond, P.J., Domene, C., Pang, A. and Sansom, M.S.P. (2003) Large scale biomolecular simulations: current status and future prospects. *Proc. UK e-Science All Hands Meeting 2003 (ISBN 1-904425-11-9)* pp. 901-907
- [2] D.P. Chimento, A.K. Mohanty, R.J. Kadner, M.C. Wiener, *Nature Structural Biology* **10**, 394 (2003)

Quantum Directed Virtual Evolution: Marcus Durrant, John Innes Institute

The software modules required to perform the calculations for the project are:

- Master complex table. This records the results of all calculations performed to ensure that only truly new calculations are setup for future generations. Responsibility: JI
- Input generator. This processes a "nanogene" and generates the corresponding GAMESS-UK input. Responsibility: JI (of course DL help wherever needed)
- Job submission and execution. Performance of the quantum chemistry calculation, tests for potential problems, etc. The energy expression requires the use of the LANL2DZ basis sets and effective core potentials and the B3LYP density functional. Responsibility: DL
- Output parser. Takes the output from the quantum chemistry calculation and extracts relevant results like energies, geometries, etc. Responsibility: DL
- Selection routine. Selects survivors of the nanogenes based on the computed results. Responsibility: JI
- Breeding and mutation. Create a set of nanogenes for the next generation of calculations by breeding and mutation. Responsibility: JI

For Daresbury Laboratory the biggest task is to provide the quantum chemistry engine and to ensure that this is robust enough to deal with the large numbers of calculations to be performed without significant human intervention. For this purpose we have broken the provision of the third component listed above down into the following steps:

- Initial investigation of the code performance using the traditional SCF drivers. This flagged up a number of problems that had to be resolved in that the built in basis set was not identical to the one in GAUSSIAN. This has been fixed. The traditional convergence approach failed in about 24% of the cases tried. Extrapolating this to the full size batches would mean that 100 jobs per generation would need human attention. Clearly this is undesirable and is addressed with a new convergence scheme described below.
- Extension of the newscf driver to handle harmonic basis functions as required by the LANL2DZ basis. This required building in some new linear algebra and has been completed.
- Adaptation of the convergence control logic to the challenges posed by transition metal complexes. The original choices of when to store the intermediate results and how to restore them to back out of divergent pathways resulted in superfluous iterations being performed. A few changes to the logic fixed this.
- Optimisation of the convergence scheme for the type of calculations associated with this project. Various schemes were tried to find an approach that converges as aggressively as possible while at the same time being robust. The best scheme found to date fails to converge in 16% of the cases. This is a marked improvement assuming from past experience which indicated that about 5% of the molecules are so unphysical that they cannot be expected to converge.
- Development of a task farm harness to exploit HPCx most effectively for the large number of calculations involved. A few potential approaches were considered to enable task farming a large number of jobs with GAMESS-UK including:
 - 1) Programming poe in such a way as to run GAMESS-UK on sub sets of the nodes allocated to the whole task. The system administrators advised us however that this would require opening up various bits of poe to users at large. Effectively this would make the whole machine interactively accessible to anyone and is clearly undesirable.
 - 2) MPH4: the distributed multi-component environment. However this environment is much more complex than we require as it allows completely independent programs to be connected up through sharing MPI communicators. Furthermore we would have to license this tool from Lawrence Berkeley National Laboratory.
 - 3) Setup our own top-level program with appropriate MPI communicators that divides all of the processors up into master, slave-master and slaves. The harness simply links in GAMESS-UK as a subroutine and provides it with an MPI communicator. As a result the application code will operate as desired with very little modification. As this could easily be done we chose to implement this approach.

We are now ready to start building the first generation of molecules.

4.2.2 HPCx Industry Day

CCLRC Daresbury Laboratory hosted the HPCx Industry Day on 5 April 2004. The main objectives of the meeting were:

1. to introduce, raise awareness of, and demonstrate how Terascale class High Performance Computing systems such as HPCx and successive generation of facilities can meet the challenges of industrial R&D;
2. to promote the skill-base available in HPCx for efficiently and effectively exploiting high performance computing systems, developing new scientific functionality and simulation technologies;

3. to explore the scale and scope of potential commercial interest in the HPCx service and subsequent generations of facilities; and,
4. to explore the quality of service required by industrial users of academic research High Performance Computing services.

The audience was made up of potential users of the HPCx service from the industrial sector, a selection of software vendors and academic researchers in addition to Research Council officials with high-performance computing interests. Talks covered the areas of computational engineering, life sciences, environment, materials and chemistry simulations. We were particularly interested in over-viewing the impact that systems with sustained performances of 1 Teraflop, 10 Teraflops and 100 Teraflops would have on industrial R&D applications.

The day began with an official welcome from Prof Paul Durham, Director of the Computational Science & Engineering Department at Daresbury Laboratory. Alan Simpson (HPCx) provided an *Overview of the HPCx Service* and Martyn Guest (HPCx) described some of the challenges of *HPC - Scaling to 1000's of Processors*. Adrian Mulholland (University of Bristol) overviewed *Computational enzymology: modelling enzyme catalysis with HPC*. Mike Payne (University of Cambridge) described *Next Generation Technologies for First Principles Atomistic Simulation* and Mark Sansom (University of Oxford) reported on *Large Scale Simulations of Biological Membranes*. Ben Slater (The Royal Institution) discussed *HPC in Atomic Simulations of Materials*. Jonathon Chin (University College London) overviewed *HPC in Modelling Complex Fluids* and Phil Tattersall (Qinetiq) described *Uses of High Performance Computing in Aerodynamics and Aeroacoustics*. The day ended with a discussion of the quality of service and software environment that would need to be provided to make the HPCx service attractive to industry.

A total of sixty people attended the Industry Day – twenty from software and hardware vendors and resellers, eight from Universities, twenty-seven from research organisations and five from industrial organisations. Ideally, we would have liked to attract more people from different industrial organisations to the event but some excellent presentations made the exercise very worthwhile. Most of the talks are now available on the web site. The HPCx service will progress its engagement with industry on a number of fronts, in particular:

- holding one-on-one meetings with the industrial groups that attended the meeting, and those that expressed interest, on the quality of service that they would require;
- discussions with commercial software vendors on how to tempt their high-end users onto the facility, one key issue here is licensing costs;
- discussions with key academics on how to support them bringing their industrial collaborators onto the service.

4.2.3 Commercial Codes

Both Fluent and Abaqus have been ported to HPCX Phase 2 and initial test results reported to industrial customers alongside initial discussions on the required Quality of Service. The ports of the commercial codes took longer than expected and some issues remain in running the applications under LoadLeveller. We plan to return to the ports of these codes in July following the Phase 2 acceptance tests.

4.2.4 Teragyroid Experiment

A report on lessons learnt from the Teragyroid experiment undertaken last November was submitted in April 2004. Richard Blake has been invited to sit on a Steering Committee, representing HPCx, which will overview a Call for Proposals for future experiments.

4.2.5 JCSR Working Group

Richard Blake has been invited to sit on the JCSR Working Group chaired by Prof J Annett to explore the academic community's needs for visualisation. The Working Group will arrange a 'user requirements meeting' in September with a report to be submitted to JCSR in November. The outputs from the user consultation will be used to inform CCLRC's programme of visualisation activities.

4.3 Terascaling Applications (*Dr Martyn Guest*)

The work described below covers the period April-June 2004, and details evaluation and development terascaling activities around application codes, libraries and tools, plus details of staff training, and attendance at Consortium meetings and associated events, including presentations by members of the Terascaling Team.

4.3.1 Phase 2 Acceptance Testing

Benchmarking of AIMPRO, CASTEP, DL-POLY, H2MOL and PCHAN for the Phase 1 to Phase 2 transition has been completed. Results have been reported to the ATWG via the Phase 2 Acceptance Test Report. Tests were prepared in advance and run during the Acceptance Tests. This work included assessment of memory affinity and processor affinity (via the VSRAC tool) which became relevant with the transition from 8-way to 32-way LPARs, and which have been disseminated in User Documentation on the Web Site.

4.3.2 Computational Materials

Castep

- Testing of processor binding using the VSRAC has proved highly beneficial and has been employed in setting up new executables and documentation for Phase 2. Work has begun on addressing particular problems associated with the modelling of organic liquids. This is applicable to large systems in general and has raised the issue of whether real-space or reciprocal space potentials are more efficient. As the system gets bigger, real-space potentials should be better but they are not yet coded as efficiently as reciprocal space potentials and also have more reliance on 3D-FFTs.
- Investigations are proceeding of the HPCx general split-communicator task-farming package, which has been made available using MPH. This is operating successfully for stand-alone codes. Some additional work is required for codes using ScaLAPACK/BLACS as these tools require knowledge of the global communicators, which needs to be cleverly matched to the MPH communicators.

SIESTA

- Following an extensive study, we identified the eigenvalue solver as the limiting factor for performance in this code. Comparing the performance with different libraries (PESSL and ScaLAPACK) demonstrated that ScaLAPACK gives the best performance. Further investigation revealed that the eigenvalue solver uses a 1D processor array and by modifying the code to use a 2D array of processors we have achieved a performance improvement

of around 20% on 32 processors. Joachim Hein is currently working on an HPCx report summarising these findings.

VASP

- We have spent some time trying to profile this code. However, as it caused a system crash on the Phase 2 system, effort focussed on porting it to the TND machine, to allow users to utilise this code. The bug was also reported to IBM who have now fixed the problem. Further work will investigate this issue further, and then refocus on code profiling. Gavin Pringle is also looking at creating a task farm harness for this code, following discussions with one of the users at the HPCx Annual Seminar.

4.3.3 Molecular Simulation

CHARMM

- We have started work on this code, building and testing both 32-bit and 64-bit executables (the latter not being supported by the distribution version of the code and requiring significant code changes) as well as the parallel QM/MM code incorporating GAMESS-UK. Initial performance results are disappointing (as expected from earlier work on the code) and investigations are continuing in consultation with the CHARMM development community.

Gromacs

- Gromacs 3.1.4 has been installed on HPCx with single-precision and double-precision versions, both in serial and parallel. A user has confirmed that this release is functioning correctly. Problems were encountered when installing Gromacs 3.2.1 on the system. However, following a visit from Angelo Rossi (IBM), Fiona Reid has identified the problem as relating to the use of the MASS library, and has successfully installed Gromacs 3.2.1 without this. She is currently investigating how this will influence performance.

MD Code Intercomparison

- We have initiated a comparative study of the main molecular dynamics codes: CHARMM, AMBER, NAMD, DL_POLY, Gromacs and Lammmps. Benchmarks have been obtained from the major user groups and we already have some preliminary results on a few codes. Work will now focus on the non-trivial task of converting input files for different codes.

NAMD

- We have been comparing the performance of HPCx and the Altix system for this code. The code scales well on both systems, but runs faster on the Altix.

4.3.4 Molecular Electronic Structure

NWChem

- Work here has not proceeded due to fundamental performance problems with the Global Array tools on the system. Communication with the GA developers on the origin on the problems continues and some improvement is expected when the revised microcode is available for the HPS, at which point further benchmarking activity will be justified. Additional performance optimisation of the GAs is the subject of an ongoing dialog between the GA developers and IBM (Poughkeepsie) but it will probably require additional programming interfaces (over and above the LAPI libraries) to be made available for significant progress to be made.

GAMESS-UK

- Scalability improvements to the SCF/DFT kernel have continued. This activity focuses on the use of MPI-based distributed matrix algebra including ScaLAPACK/BLACS. The latest step has been the implementation of symmetry adaption and projection of components from the basis set. These two operations involve transformation of distributed operator matrices and molecular orbitals resulting in rectangular (rather than square) matrices, leading to practical changes to the distributed matrix library on which the kernel is built. The timing analysis has been updated and benchmark data is being collected. It is anticipated that GAMESS-UK will satisfy the requirements for (at least) the Bronze star Capability Incentive classification.
- The development of a massively-parallel task-farmed implementation, based on split MPI communicators, allows GAMESS-UK to accept as input a list of input files (held on a GPFS file system) and to process them in a round-robin fashion, using a task group of typically 8 processors for each job. The jobs run asynchronously (they do not have to be the same size) and failures (such as input errors) will abort the current calculation allowing the job to proceed with the next one.

4.3.5 Physics

LUDWIG

- Kevin Stratford has carried out a detailed performance analysis of this code, on both the Phase 1 and Phase 2 systems, which identified the collision/propagation phase as the major bottleneck. Accordingly a substantial reworking of this phase of the code has been carried out. This has resulted in good scaling between 128 and 1024 processors, and around a 25% improvement in performance on 1024 processors. The communication phase has still to be investigated.

4.3.6 Computational Engineering

UKAAC

- A number of staff attended the UK Applied Aerodynamics Consortium (UKAAC) inaugural meeting at DL on 18th June 2004 and Andrew Sunderland and David Emerson gave a presentation entitled *Introduction to the HPCx Service*. We are now assisting Consortium members in dealing with porting and optimisation issues.

Roach/UKAEA

- Lorna Smith has been liaising with this Consortium and discussing their requirements for optimisation support. The CENOTRI code has been compiled on HPCx. Future work will focus on optimising the communications within this code, possibly modifying the 1D decomposition to 2D.

CFX

- CFX has been upgraded to version 5.7. CFX has recently been taken over by Ansys, and the main feature of this latest release is the interoperability it offers with other Ansys Tools and Solvers.

Fluent

- Fluent is running under a temporary license while negotiations proceed for the acquisition of a permanent license. There are still problems with the interaction of the MPI version of Fluent with LoadLeveler, which are under investigation by a Fluent technical consultant.

4.3.7 Libraries

Eigensolvers

- We have investigated the performance of the PLAPACK and ScaLAPACK Eigensolvers on HPCx for the Standard Problem. This has been written up in an HPCx technical report.

4.3.8 Tools

DDT

- We experienced many problems with Streamline's initial implementation, some due to IBM's version of dbx (believed to be cured now in AIX 5.2) and some due to the Mesa open-source OpenGL library which was being used. Streamline now have their own POWER4 development system, which, like HPCx, runs AIX 5.2 and IBM's own OpenGL libraries. Streamline have also developed a partnership with Absoft to allow DDT to run with their fx2 debugger. A brand new release of DDT is promised for July and we shall submit this to testing with a range of applications.

Paraver

- The software has been upgraded to Paraver v3.3 and OMPtrace 1.2. This appears to have cured the stability problems referred to in the previous report.

Vampir

- Vampir and VampirTrace have been upgraded to version 4.0. Vampirtrace 4.0 is now completely thread-safe, thus allowing tracing of multithreaded MPI applications. Support for Java processes is now included.

4.3.9 Consortium Meetings and Presentations

- Mike Ashworth, David Henty, Martin Plummer and Kevin Stratford gave presentations on terascaling applications at the HPCx Scaling Workshop and NERC User Meeting at UCL on 26th May 2004.
- Andrew Sunderland and David Emerson gave a presentation entitled *Introduction to the HPCx Service* at the UK Applied Aerodynamics Consortium (UKAAC) inaugural meeting at DL on 18th June 2004.
- Joachim Hein, Andrew Sunderland and Martin Plummer gave presentations at the Second HPCx Annual Seminar in Edinburgh on the 9th July 2004.

4.4 Software Engineering (*Dr Stephen Booth*)

The major event of this quarter was the upgrade from Phase 1 to Phase 2. Many of the activities of the software engineering team this quarter have been driven by this upgrade, either preparing for the upgrade or evaluating the system afterwards.

4.4.1 Low Level Communications

A good understanding of the low-level communication performance is vital to the efficient use of systems like HPCx.

The following Technical reports have been generated during the reporting period:

- HPCxTR0407 *LAPI on HPS*
http://www.hpcx.ac.uk/research/hpc/technical_reports/HPCxTR0407.pdf. This report investigates the performance of the IBM LAPI communication library on the phase-2 system. On phase-2 the MPI communication library is implemented on top of LAPI so this forms a basis for the understanding of MPI performance on this system.

We will be continuing to investigate how the low-level communication performance of the Phase 2 system changes as the HPS microcode updates become available.

4.4.2 Shared Memory Techniques

The standard environment for developing parallel programs in a shared memory environment is OpenMP. OpenMP is a set of language extensions (compiler directives) and is available for both C and Fortran. We have recently updated the OpenMP micro benchmark suite to support the additional features in OpenMP version 2.0. This benchmark suite has been run to compare the OpenMP implementation on HPCx, the SGI Altix and a Sun E15K. The results are currently being compiled into a report on OpenMP performance which will be published in Q3. The same work also forms the basis for a submission to EWOMP'04, the Sixth European Workshop on OpenMP.

4.4.3 Grid Computing

No major development of the Grid Computing infrastructure was scheduled during this quarter. However, we plan to review the Globus installation in Q3 and update as appropriate.

4.4.4 Data Handling

We produced the following technical report during the reporting period:

- HPCxTR0405 *HPCx Archiving user guide*
http://www.hpcx.ac.uk/research/hpc/technical_reports/HPCxTR0405.pdf. This report gives detailed instructions for users of the HPCx Tape archive system.

User groups are now making significant use of this facility.

4.4.5 General Terascaling Techniques

We have been investigating general terascaling techniques applicable to systems built out of clustered SMP nodes. Much of this work was developed as part of the course *Improved Performance Scaling on HPCx*. However, many of these techniques are generally applicable to other systems with a similar architecture and we will be presenting a tutorial on this material at Supercomputing 2004 in Pittsburgh.

The following technical reports have been generated during this period:

- HPCxTR0409 *Planned AlltoAllv: A cluster approach*
http://www.hpcx.ac.uk/research/hpc/technical_reports/HPCxTR0409.pdf. This report investigates a method of optimizing the important AlltoAllv collective on clustered SMP systems. AlltoAllv is an important communication primitive for many codes most significantly those that require distributed multi-dimensional FFT operations. The current implementation gives significant improvements for operation with small message sizes. We are planning to continue to develop this technique with the hope of developing a library that we can make available to user of HPCx.

4.4.6 Systems Programming

We have made a small number of improvements to the SAF:

- We responded to a request from Dr Adrian Wander (e06) to enhance the consortium reports available to project managers and PIs. The breakdown of usage by consortium member now includes information about which budget was used by each consortium member.
- Support for additional reporting charts.
- Additional SAF functionality for handling the migration from Phase 1 to Phase 2.

We are now investigating some of the additional capabilities in the recent versions of IBM's parallel operating environment to see if they could benefit our users. The features we are currently planning to evaluate include:

- The poe priority co-scheduler
- The MP_TASK_AFFINITY support that will be available in HPS Service pack 7.

4.5 Operations and Systems (*Mr Mike Brown*)

4.5.1 Staffing

There has been no change in staffing levels, although the coverage still remains substantially in excess of the "core hours" contractual requirement.

Additional effort has had to be applied because of the build up of the Phase 2 service migration system

4.5.2 Test & Development System

The Phase 1 test and development system has continued to prove its value until the closure of the Phase 1 system, and the need for a similar system in support of the Phase 2 service remains as strong as ever.

4.5.3 Maintenance Sessions

Regular maintenance sessions were taken up until the closure of the Phase 1 service, and their need continues on the replacement Phase 2 service.

4.5.4 File Archive

The file archiving system under TSM is in full service.

4.5.5 Phase 2 Migration

The full Phase 1 service closed at 26 Apr/0800. Over the next three days the service was migrated over to the initial Phase 2 platform, and this required the complete backup and re-creation of the user filestore upon the new phase 2 discs.

The interim overlap service on a reduced (20 frame) Phase 2 system opened at 29 Apr/1300, and the system was progressively upgraded towards the full 50 frame configuration over the next four weeks.

The interim service closed at 26 May/0800, the system entered final acceptance, and after the performance benchmarks were run on the closed machine, the

service re-opened for full user service (although operating under acceptance conditions) at 28 May/1400. The Implementation Certificate was issued by EPSRC on 17 June.

4.6 Staffing

<i>AV</i>	<i>April</i>	<i>May</i>	<i>June</i>
DL	5.1	5.0	5.3
EPCC	8.8	8.0	10.2
Total	13.9	13.0	15.5

<i>Systems</i>	5.8	5.8	6.0
----------------	-----	-----	-----

5 Summary of Performance Metrics

<i>Metric</i>	<i>TSL</i>	<i>FSL</i>	<i>January</i>	<i>February</i>	<i>March</i>
Technology serviceability	80%	99.2%	100.0%	98.4%	98.0%
Technology MTBF (hours)	200	300	∞	244.0	366.0
Number of AV FTEs	7.5	10	13.9	13.0	15.5
Number of training days per month	30/12	40/12	16/4	21/5	22/6
Non in-depth queries resolved within 3 days	85%	97%	100.0%	100.0%	100.0%
Number of A&M FTEs	3.75	5.75	5.8	5.8	6.0
A&M serviceability	80%	100%	100.0%	100.0%	99.3%

<i>Colour</i>	<i>Meaning</i>
	Exceeds FSL
	Between TSL and FSL
	Below TSL

Note: The number of training days is reported as a running total since the start of the year.

Appendix A: Incident Severity Levels

SEV 1 — anything that comprises a FAILURE as defined in the contract with EPSRC.

SEV 2 — NON-FATAL incidents that typically cause immediate termination of a user application, but not the entire user service.

The service may be so degraded (or liable to collapse completely) that a controlled, but unplanned (and often very short-notice) shutdown is required or unplanned downtime subsequent to the next planned reload is necessary.

This category includes unrecovered disc errors where damage to filesystems may occur if the service was allowed to continue in operation; incidents when although the service can continue in operation in a degraded state until the next reload, downtime at less than 24 hours notice is required to fix or investigate the problem; and incidents whereby the throughput of user work is affected (typically by the unrecovered disabling of a portion of the system) even though no subsequent unplanned downtime results.

SEV 3 — NON-FATAL incidents that typically cause immediate termination of a user application, but the service is able to continue in operation until the next planned reload or re-configuration.

SEV 4 — NON-FATAL recoverable incidents that typically include the loss of a storage device, or a peripheral component, but the service is able to continue in operation largely unaffected, and typically the component may be replaced without any future loss of service.

Appendix B: Current Projects

EPSRC Projects

<i>Code</i>	<i>Class</i>	<i>Title</i>	<i>PI</i>
e01	1	UK Turbulence Consortium	Prof Neil Sandham
e02	1	Ab-initio simulation of covalently bonded materials	Dr Patrick Briddon
e03	1	Multi-photon, electron collisions and BEC HPC consortium	Prof Ken Taylor
e04	1	Chemreact Computing Consortium	Prof Jonathon Tennyson
e05	1	Materials Chemistry using Terascaling Computing	Prof Richard Catlow
e06	1	UK Car-Parrinello Consortium	Prof Paul Madden
e07	2	Turbulent Plasma Transport in Tokamaks	Dr Colin M Roach
e08	2	Organic Solid State	Prof Sarah Price
e09	2	Molecular Properties and their Geometry	Prof Peter Taylor
e10	1	Reality Grid	Prof Peter Coveney
e11	1	Bond making and breaking at surfaces	Prof Sir David A King
e12	1	Parallel programs for the simulation of complex fluids	Dr Mark R Wilson
e13	1	TeraGyroid project	Dr Richard J Blake
e14	1	Blade and Cavity Noise	Prof Neil Sandham
e15	2	CSAR/HPCx Collaboration	Dr Mike Pettipher
e16	1	Cardiac virtual tissues	Prof Arun V Holden
e17	1	Integrative Biology	Dr David Gavaghan
e18	1	DARP: Highly swept leading edge separations	Prof Michael A Leschziner
e19	1	Edinburgh Soft Matter and Statistical Physics Group	Prof Michael E Cates
e20	1	UK Applied Aerodynamics Consortium	Dr Ken Badcock

PPARC Projects

<i>Code</i>	<i>Class</i>	<i>Title</i>	<i>PI</i>
p01	1	Atomic Physics and Astrophysics	Prof Alan Hibbert

NERC Projects

<i>Code</i>	<i>Class</i>	<i>Title</i>	<i>PI</i>
n01	1	Large-Scale Long-Term Ocean Circulation	Dr David Webb
n02	1	NCAS	Prof Alan J Thorpe
n03	1	Computational Mineral Physics Consortium	Dr John Brodholt
n04	1	Shelf Seas Consortium	Dr Roger Proctor
n05	2	Non-linear Wave-particle Instabilities in Plasmas	Dr Mervyn Freeman

BBSRC Projects

<i>Code</i>	<i>Class</i>	<i>Title</i>	<i>PI</i>
b01	2	Quantum Chemistry Studies of the Rusticyanin Protein Crystal	Prof Samar Hasnain
b02	1	Modelling enzyme catalysis	Dr Adrian J Mulholland
b03	1	Towards a virtual outer membrane	Prof Mark S Sansom
b04	1	Life sciences software development	Dr Jo L Dicks
b05	1	Virtual forced evolution of catalytic transition metal complexes	Dr Marcus Durrant
b06	2	Biomolecular computational chemistry	Prof Jonathan D Hirst
b07	1	Simulation of Radioprobing	Dr Charlie Laughton

CCLRC Projects

<i>Code</i>	<i>Class</i>	<i>Title</i>	<i>PI</i>
c01	1	Daresbury Laboratory Facilities Agreement Consortium	Dr Richard J Blake

Externally-funded Projects

<i>Code</i>	<i>Title</i>	<i>PI</i>
x01	HPC-Europa	Dr J-C Desplat

HPCx Projects

<i>Code</i>	<i>Title</i>	<i>PI</i>
z001	HPCx Support	Dr Alan Simpson
z002	Systems and Operations	Mr Mike Brown
z003	Test Project	Dr Denis Nicole
z004	HPCx Training	Dr David Henty
z05	Outreach Projects	Dr Richard Blake
z06	Application Porting	Dr David Henty
z07	Package Installation	Dr Mike Ashworth