

HPCx Quarterly Report

July-September 2003

1 Introduction

This report covers the period from 1 July 2003 at 0800 to 1 October 2003 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. Section 4 includes reports on progress against the Annual Plan from each of the functional teams. A summary table of the key performance metrics is in Section 5. The Appendices define the incident severity levels and list the current HPCx projects.

2 Executive Summary

- During this quarter, utilisation of the service has again averaged more than 75% on the capability region.
- The modal job size remains 128 CPUs and but the fraction of utilisation from jobs using at least 256 CPUs increased significantly.
- However, looking ahead, it appears that demand may outstrip supply until the Phase 2 upgrade. We have submitted an options paper to the Oversight Committee on controlling demand.
- We have agreed to increase the maximum number of concurrent research groups to 35, on the assumption that EPSRC will be able to fund 2 additional posts from the start of next year.
- Due to a cluster of incidents in August, the number of failures is somewhat higher than last quarter. However, the MTBF and serviceability figures are at or close to the Full Service Levels.
- The first part of the 'Early Access' Phase 2 system has been successfully installed, with the next part due in 3Q03. The planning process for Phase 2 is now well under way.

- The Phase 1 development system has allowed some progress with HSM under TSM. However, we are still discussing with IBM how to ensure that this has appropriate functionality for users. We will make additional on-line disc available to users who are adversely affected by the delays in HSM.
- HPCx is now coordinating the UK part of the collaboration with ETF and we are developing the technology required for a computational steering experiment at SC2003.
- Interactions with the users have been good:
 - the helpdesk is exceeding all its targets;
 - the HPCx mini-workshop at All Hands attracted an audience of around 50;
 - we ran a course at Cambridge;
 - the HPCx annual seminar is being arranged at Daresbury in December and there will be an associated user group meeting;
 - there are now 7 technical reports available on the web site, which is ahead of target.
- This quarter has seen good progress on a number of key outreach activities:
 - we believe that the VAT issues with the IBM Lifesciences funding have now been resolved;
 - initial plans for the HPCx Industry Day have been drawn up, although this may now be run early in 2004.
- We have made good progress in investigating the performance characteristics of the system:
 - porting T3E MPI to HPCx;
 - studies and technical reports on the performance of single-sided communications, boundary exchanges in MPI, Java benchmarks and mixed mode programming.
- The Terascaling team has begun work on a good range of applications codes and has had a number of successes:
 - the CRYSTAL results were well received at the All Hands mini-workshop;
 - the following capability incentive awards have been made:
 - bronze: NAMD
 - silver: CRYSTAL
 - gold: LB3D, PDNS3D
 - Gaussian is now available on HPCx and an initial study has been planned;
 - enhanced versions of a number of key tools have been installed.

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:

	July	August	September
Incidents	10	12	8
Failures	2	6	2

The number of incidents and failures are both higher than in the previous quarter.

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

July

<i>Failure</i>	<i>Site</i>	<i>IBM</i>	<i>External</i>	<i>Reason</i>
03.103		100%		GPFS failure on I0f01
03.100		100%		I3f01 off switch; GPFS hung

August

<i>Failure</i>	<i>Site</i>	<i>IBM</i>	<i>External</i>	<i>Reason</i>
03.111		100%		Plane 1 crashed, taking out GPFS
03.113		100%		VSD servers sharing VG's, idled system
03.115		100%		GPFS loss on I1f01
03.116	100%			Loss of link to external network
03.117		100%		I3f02 crashed, taking out GPFS
03.122			100%	JANET problems at Manchester

September

<i>Failure</i>	<i>Site</i>	<i>IBM</i>	<i>External</i>	<i>Reason</i>
03.125	25%	75%		LL home directory inaccessible
03.129		100%		GPFS failure, cluster wide

3.1.2 Performance Statistics

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

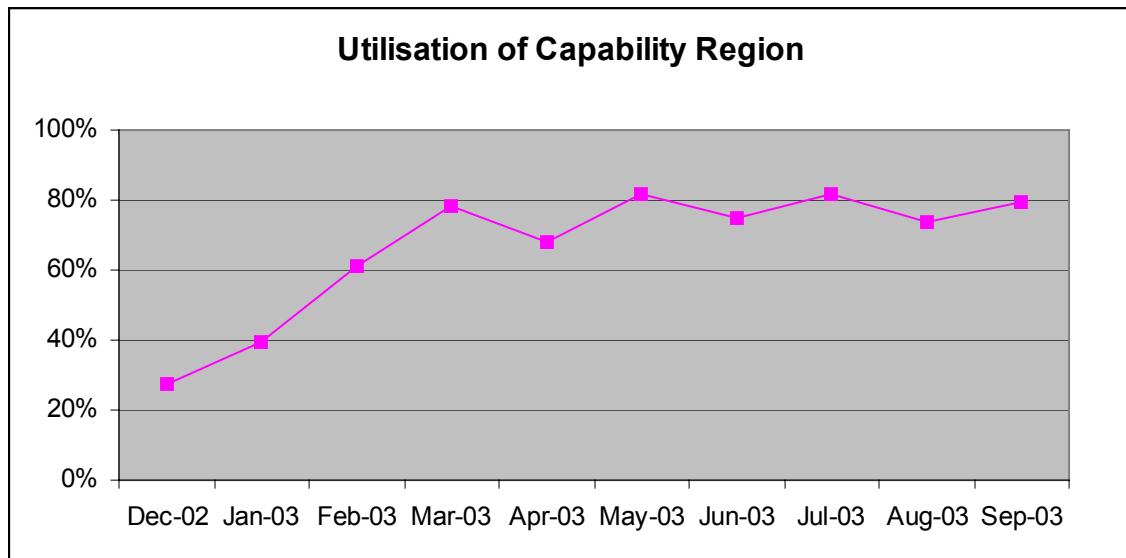
- $MTBF = (24 \times 30.5) / (\text{number of failures in month})$

- Serviceability (%) = 100 x (WCT – SDT – UDT) / (WCT– SDT)

<i>Attribution</i>	<i>Metric</i>	<i>July</i>	<i>August</i>	<i>September</i>	<i>Quarterly</i>
IBM	Failures	2	4	1.75	7.75
	MTBF	366	183	418	283
	Serviceability	99.5	98.6	99.6	99.2
Site	Failures	0	1	0.25	1.25
	MTBF	∞	732	2928	1757
	Serviceability	100.0	100.0	99.9	99.9
External	Failures	0	1	0	1
	MTBF	∞	732	∞	2196
	Serviceability	100.0	99.9	100.0	99.9
Total	Failures	2	6	2	10
	MTBF	366	122	366	220
	Serviceability	99.5	98.5	99.6	99.2

3.2 Capability Utilisation

The monthly utilisation for the 1024-processor capability region is shown in the graph below. This has averaged more than 75% for the last 4 months.

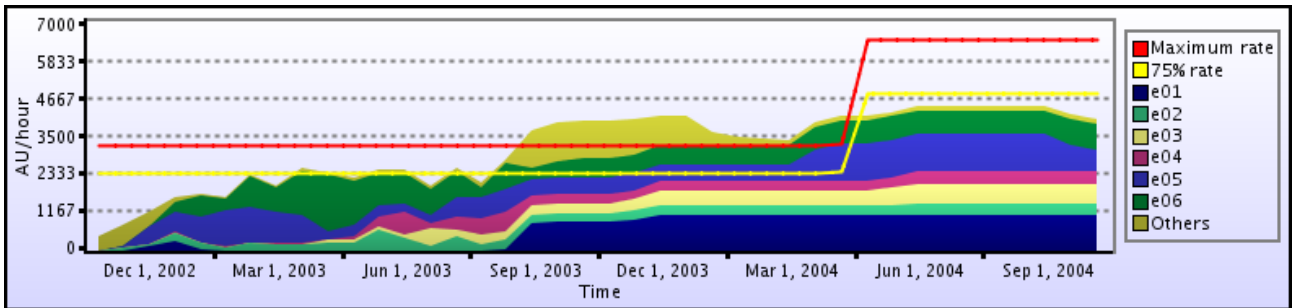


3.3 Capacity Planning

Predicted Utilisation

The following graph shows the utilisation since the start of the project and the projected utilisation until November 2004. 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 (the red line in the graph). However, the practical maximum is probably around 75% of this, i.e., 2430 AUs per hour, which is shown as a yellow line.

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.



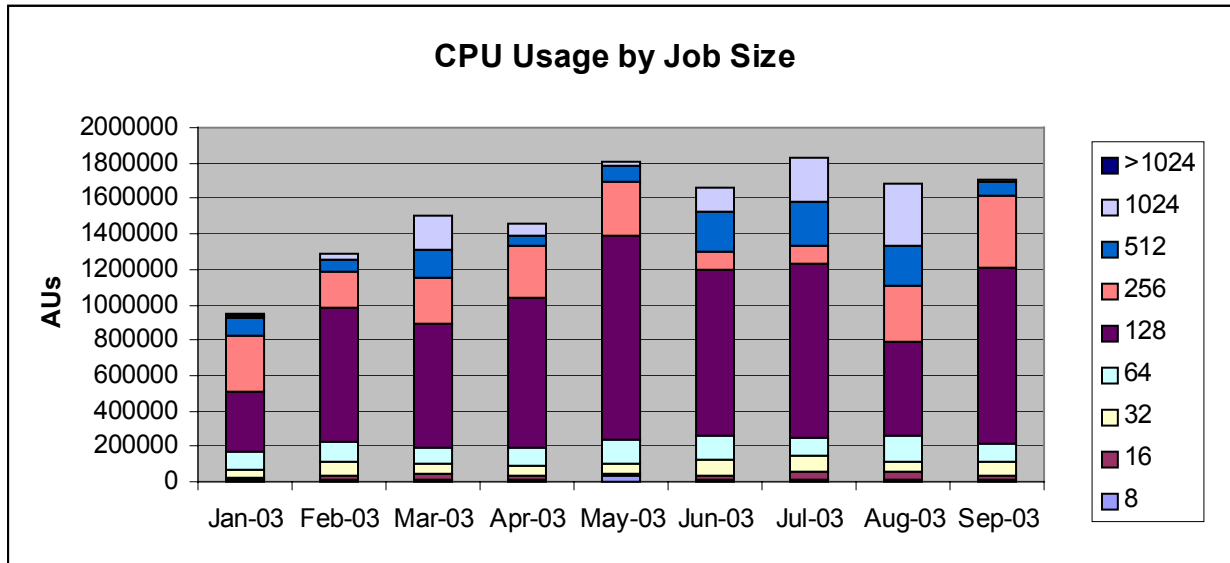
This analysis suggests that demand until the start of Phase 2 appears to exceed the practical maximum by more than 40%. It appears likely that, even without new grants coming on to the system, demand on the Phase 1 system will exceed supply.

In discussions with EPSRC, we have agreed that investigating methods of controlling demand over this period may well be the best option. A paper on this has been submitted to the Oversight Committee.

Numbers of Research Consortia

There are currently 19 research consortia using the HPCx system. The HPCx support activity is sized on a maximum of 25 concurrent research consortia (1.9.3 in Schedule 3). With the Life Sciences projects starting imminently, we will probably soon be at this limit. We have agreed with EPSRC to raise the limit to 35 on the understanding that EPSRC will fund 2 additional posts from the start of next year.

3.4 CPU Usage by Job Size



The above graph shows that the modal job size is still 128 CPUs. However, during this quarter, utilisation by jobs of at least 256 CPUs has increased to 37.8% of the total.

3.5 CPU Usage by Consortium

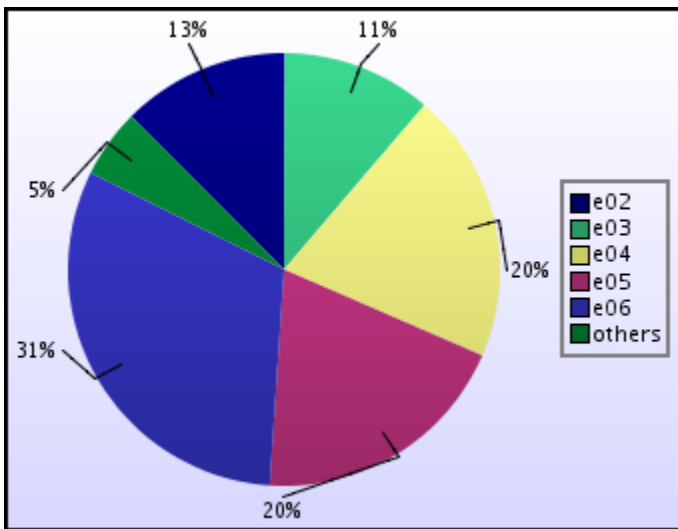
The PIs and titles for the various consortia are listed in Appendix B.

<i>Consortium</i>	<i>July</i>	<i>August</i>	<i>September</i>	<i>Quarterly</i>	<i>%age</i>
e01	2565	6182	40012	48759	0.93
e02	271653	262441	122716	656810	12.56
e03	119650	276385	190066	586101	11.21
e04	444946	188674	425258	1058878	20.25
e05	184603	340997	496567	1022167	19.54
e06	773041	535360	329909	1638310	31.33
e07	0	0	0	0	0.00
e08	907	0	0	907	0.02
e09	1032	0	0	1032	0.02
e10	0	25421	0	25421	0.49
e11	0	5	1	6	0.00
EPSRC Total	1798397	1635465	1604529	5038391	96.34

n01	263	7902	74372	82537	1.58
n02	186	4422	76	4684	0.09
n03	12064	14299	14158	40521	0.77
n04	6252	6506	1122	13880	0.27
n05	0	0	11484	11484	0.22
NERC Total	18765	33129	101212	153106	2.93

p01	577	148	262	987	0.02
PPARC Total	577	148	262	987	0.02

z001	14259	8244	3959	26462	0.51
z002	577	8	12	597	0.01
z004	4375	4029	590	8994	0.17
z06	362	729	250	1341	0.03
HPCx Total	19573	13010	4811	37394	0.72



3.6 Helpdesk

3.6.1 Classifications

Category	Number	% of all
Administrative	82	35.2
Technical	138	59.2
In-depth	11	4.7
PMR	2	0.9
TOTAL	233	100.0

<i>Service Area</i>	<i>Number</i>	<i>% of all</i>
Phase 1 platform	181	77.7
Website	28	12.0
Other/general	24	10.3
TOTAL	233	100.0

3.6.2 Performance

<i>All non-indepth queries</i>	<i>Number</i>	<i>%</i>	<i>Target</i>
Finished within 24 Hours	174	79.1	75%
Finished within 72 Hours	219	99.5	97%
Finished after 72 Hours	1	0.5	

<i>Administrative queries</i>	<i>Number</i>	<i>%</i>	<i>Target</i>
Finished within 48 Hours	80	97.6	97%
Finished after 48 Hours	2	2.4	

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	48	60	7	0
dl.ac.uk	8	21	4	1
Sysadm	26	52	0	1
Other people	0	5	0	0

3.7 Service Quality Tokens

<i>Date</i>	<i>Person</i>	<i>Value</i>	<i>Comment</i>	<i>Status</i>
Aug 29, 2003 8:26:59 AM	Mr Paul Burton	••	flexlm license manager for totalview not restarted after system reboot. Second time this has happened recently (Lois Steenman-Clark had this problem a couple of weeks back) - had to wait good few hours for flexlm to be restarted.	Operating procedures have been enhanced to ensure that this takes place on every reboot
Aug 29, 2003 8:24:24 AM	Mr Paul Burton	••••	No communication/progress report on Q3515 - holding up my work	Very complex problem with a mass of evidence, and user away! Finally referred to IBM who provided a work-round

Jul 24, 2003 6:49:35 PM	Mr Andrew D Walkingshaw	***		
Jul 7, 2003 5:35:49 PM	Dr Fernando Bresme	...		

4 Support

4.1 Applications Support (*Dr David Henty*)

The HPCx Applications Support effort continues to go smoothly with frontline queries being dealt with as stated in the SLA. This quarter we completed three in-depth technical reports, ran a course and a workshop offsite, gave six presentations at four external events and installed five new packages for users.

4.1.1 Helpdesk

The helpdesk has continued to operate extremely smoothly, and the statistics above show that we are meeting our targets for answering user queries.

4.1.2 Documentation

The HPCx User Guide is now fairly complete and requires little or no revision, so is still at version 1.2. However, the online FAQ has expanded significantly to provide answers to a wide variety of technical questions that are too detailed for inclusion in the main User Guide.

4.1.3 Technical Reports

Of the eight technical reports identified in the Annual Plan, six were due by the end of Q3. Four reports were already available at the end of Q2, although the single-sided report had to be substituted by a different report, HPCxTR0302, due to unforeseen complications with the LAPI installation. The planned single-sided report is now available on the WWW.

Two new reports were due in Q3, one on performance tools and another on mixed-mode programming. The tools report focuses on the HPM Toolkit, an IBM-specific utility provided by the Advanced Computing Technology Center. A separate document describing the use of more generic tools on HPCx (eg Totalview and VAMPIR) is also in preparation. The mixed-mode report has benefited significantly from the work of a number of our MSc students who did their research dissertations on HPCx. However, since these were only submitted in September this has delayed the production of the associated report until Q4.

In summary, the following reports have been published this quarter, bringing the total to seven (one more than the target).

- **HPCxTR0305:** *Single sided communications on HPCx*, Adrian Jackson.
- **HPCxTR0306:** *File IO from Multi-Processor Jobs*, Joachim Hein.
- **HPCxTR0307:** *Using the Hardware Performance Monitor Toolkit on HPCx*, Joachim Hein.

4.1.4 Training

The summer is always a quiet time for training, and there were no courses scheduled at EPCC this quarter. However, we will be offering 16 course-days of training in Q4. It was stated in the Q2 report that we planned to offer courses at various UK locations. The MPI course was indeed run at Cambridge University on the 29th and 30th of September, with a number of places being reserved specifically for HPCx. It attracted around 40 participants, some registering via the HPCx WWW site and the rest via the Cambridge HPCF. Plans for running the introductory course “Using the HPCx Service” at All-Hands were altered due to the limited time allocated at the meeting; we ran a short introductory workshop instead (see below). The quarterly training statistics were as follows:

Course days	2
Number of courses	1
Different courses	1
Student-days for HPCx users	18
Student-days for HPCx staff	0
Student-days available for HPCx	18

4.1.5 Conferences and Workshops

As mentioned above, we presented an “HPC Mini Workshop” at the All Hands Meeting in Nottingham on September 3rd. This comprised three talks:

- **Lorna Smith:** *Using the HPCx Service*
- **Gavin Pringle:** *Achieving Terascale Performance From Capability Applications*
- **Ian Bush:** *Large Scale Biomolecular Simulations: Current Status and Future Prospects*

Organisation of the HPCx Annual Seminar on 10th December is well underway, and we are currently taking registrations via the HPCx WWW pages. It is almost certain that we will be holding the international ScicomP meeting at EPCC in mid-2005; detailed discussions with the organising committee are currently underway.

In addition to the above meetings organised by HPCx, we have made presentations at the following four external events:

- IBM System Scientific User Group, ScicomP8, Minneapolis, 5-8 August.
Joachim Hein: *File I/O and Optimised MPI communication on a Large Power4 System.*
- Posters at the All Hands Meeting, Nottingham, 2-4 September.
Adrian Jackson: *3D FFT's on HPCx (IBM vs FFTW)*
Lorna Smith: *Mixed Mode Programming on clustered SMP systems*
Gavin Pringle: *Achieving Scalability to over 1000 Processors on the HPCx System*
- British Association's Festival of Science, Salford, 10 September.
Marty Guest & David Henty: *HPCx - Enabling Technology for Research and Discovery*

- PLATO workshop, Loughborough, July.
Gavin Pringle: *Parallelising PLATO's Matrix Diagonalisation on HPCx*

4.1.6 User Group

The second HPCx User Group meeting for 2003 will take place in Daresbury on December 10th, immediately after the Annual Seminar.

4.1.7 Newsletter

The second issue of Capability Computing is in production, with printed copies available in time for publicity at Supercomputing 2003, 15-21 November.

4.1.8 Packages

This quarter, four new package accounts have been set up: ddt, globus, chemsh and gaussian. The GAUSSIAN application has been built from source, tested and made available to HPCx users (subject to various licence agreements). We have also compiled the HDF5 IO library and placed it in a central location.

4.2 Outreach (*Dr Richard Blake*)

Over the last quarter there has been significant progress in a number of the outreach activities:

- IBM has finally agreed a form or wording for the transfer of the funds to CCLRC that will ensure that the Lifesciences programme lies outside the scope of VAT. The programme will be deemed to have started on the 1st of October 2003. The announcements of time for the projects still need to be confirmed with BBSRC. Daresbury Laboratory is recruiting a computational scientist with a biological background with interviews due to take place in mid-October. Dr Lorna Smith, who has a strong background in life-sciences applications, will coordinate the EPCC effort. Once agreements have been signed by the Universities of Oxford and Bath then we will then develop workplans for the HPCx Added Value staff to support the projects. This will allow us to identify the residual resource available for supporting other life-science and medical simulation applications.
- HPCx is planning to follow up its initial programme of supporting a number of BBSRC projects through a meeting to develop a 'Proposal to form a Biological Consortium on the HPCx Facility' during September. The following invitation was issued on 3rd October 2003 and was sent to a significant number of people, including the participants in CCP1, CCP4 and CCP5:

Discussion Meeting: Opportunities for Biological Consortia on the HPCx Facility

Background

High performance computing facilities have, to date, been primarily the domain of the physical scientists. However, over the past few years there has been a dramatic increase in capability of such machines coupled to an increase in the sophistication of parallel coding algorithms. These advances now mean that such facilities are suitable for use within the biological simulation community.

Access to the national computational facilities is usually organised via a consortium model. It appears likely that bids for computational resources on HPCx from a consortium focusing on biological applications would be well received by the BBSRC, and also welcomed by HPCx. We would therefore like to invite you to a meeting at the Royal Institution of Great Britain (Albemarle St, London) on the 6th November, 2003 to discuss the scientific and technical opportunities afforded to the computational biology community by HPCx and to consider the possibilities for the formation of one or more consortia to bid for resources.

HPCx is a capability computing service and it is well known that many of the codes in use in biochemical research lack the scalability to run on large processor counts. The objective of a consortium bid in the early stages is to provide CPU and support resources towards biological applications to allow those interested in the high-end computing field to

explore the value of the current codes and establish the potential for future developments.

It is likely that there will be interest from a number of different areas of research, and that this portfolio of applications will evolve as work on scalable algorithms proceeds. Based on the current state of codes we judge that productive work could be performed in the following areas:

1. Classical MD simulations of very large systems over long time periods.
2. Large scale electrostatic calculations (e.g. ABPS)
3. Mesoscale methods (e.g. DPD)
4. QM/MM calculations with large QM subsystems
5. Large-scale pure QM calculations (small proteins and protein models)
6. Sampling approaches that require extensive statistics (e.g. replica exchange MC and replica path methods).
7. Large-scale search algorithms (e.g. ligand docking, molecular replacement with large numbers of trial models)

We would particularly welcome involvement from researchers in these subjects, but would also welcome interest from other potential application areas.

Meeting Format

The meeting will be held in the Conversation room and will start at 11:00, aiming to finish around 4.00 pm.

We hope to have a presentation from IBM on the Life Sciences activities, in particular those of their Life Sciences Group based in Zurich, and an introduction to the HPCx service.

We will then have a series of short (ca 10 minute) presentations from participating groups introducing their particular area of interest and the types of application they would wish to run on the HPCx facility. This will serve to bring everyone 'up to speed'. Members from the HPCx team and CCP support staff will provide an update on the current performance characteristics of codes that have been benchmarked on HPCx.

Following the presentations the meeting will consist of a 'round-table' discussion focussing on the potential overlaps expressed by participants and seeking to establish if there is sufficient interest for a consortium bid to be made. If the meeting feels that such a bid should go forward, the meeting will also need to address the issue of PI for the grant request and details of the co-ordination of the grant preparation. The meeting will also need to consider the likely level of support that would be required to achieve porting of codes to the HPCx facility and the effort needed to achieve scaling of those codes. A consortium bid should include support from the HPCx centre, but there may also be a case for inclusion of one or more PDRAs.

We would wish the meeting, and any subsequently formed consortium, to be as inclusive as possible so please circulate this message to any other interested parties.

Response

If you wish to attend this meeting we would be grateful if you could reply by email d.jones@dl.ac.uk, including full name and address details.

Please indicate:

i) if you would like to give a presentation on the work you would like perform on HPCx

ii) if you have work you would like to migrate to HPCx please also highlighting the software you are currently using on serial and parallel machines.

If you are unable to attend, but would like to kept informed of developments, please let us know, including if possible a response to question ii).

- The UK – ETF proposal has been developed further with an emphasis on seeking to exploit the integrated capability of the systems in high-profile scientific demonstrations. There have been a series of meetings and numerous email interactions:
 - Wednesday 3rd of September 2003: HSC, EPSRC, Manchester, EPCC, Reality Grid and Daresbury Laboratory met with Teragrid representatives Rick Stevens and Pete Beckman to discuss the potential collaborative demonstrations coupling together HPC facilities.
 - Wednesday 10th September 2003: DL and Manchester Access Grid meeting to discuss Reality Grid software stack and implementation on HPCx.
 - Richard Blake was asked to coordinate UK involvement.
 - Monday 22nd September 2003: Access Grid meeting involving UK and Teragrid sites to discuss scope of demonstration and establish peer connections.
 - Wednesday 1st October 2003: UCL, Daresbury and Manchester Access Grid discussion of specific simulations to be undertaken. Plan to be forwarded to EPSRC for specific compute and financial resources to support the project.
- Plans are still being developed for a “HPCx Industry Day” but the date will probably shift into early next year because of the additional effort required to coordinate the UK-Teragrid demonstration project and the need to give plenty of notice if we are to attract high quality speakers. The current plan is:

HPCx Industry Day

Purpose of the Day

The objectives of the day are:

- to introduce, raise awareness of, and demonstrate how Terascale class High Performance Computing systems such as HPCx and successive generation of facilities can add value to industrial organisations.
- to promote the skill-base available in HPCx for efficiently and effectively exploiting high performance computing systems, developing new scientific functionality and simulation technologies
- to explore the quality of service required by industrial users of national High Performance Computing Facilities.
- to explore/ validate the scope of potential commercial interest in the HPCx and subsequent generations of facilities.

Target Audience

The target audience is senior level business and technical industry decision makers representative of established and target market sectors that exploit high performance computing and simulation in R&D with other opinion formers from academia and local and regional government.

Programme

The event proposed is a one-day workshop structured around exemplar applications in selected sectors together with discussions around the existing and likely future computational challenges facing the various sectors. A possible programme could contain the following:

Time	Description	Speaker
10.00	Welcome and object of the day	Paul Durham
10.15	Keynote Speaker 'The future of HPC'	IBM
10.45	Overview of the HPCx Service	Alan Simpson
11.15	Coffee Break	
11.30	HPCx Scientific Support	Martyn Guest
	HPCx Applications in:	
11.45	Engineering/ Energy/ Aero	tba
12.15	Life Sciences/Biotechnology	tba
12.45	Lunch and Tour of Facility	
13.45	Environmental Technologies	tba
14.15	Chemicals	tba
14.45	Materials	tba
15.15	Coffee Break	
15.45	Discussion: 'What can HPCx do to support your computing	
16.30	Summing up and close	

Timing

The timing could be as mid-December but this is getting very tight with diaries. Mid January may be more realistic. The constraints include likely availability of senior personnel from industry and other target organisations and the need to develop content for several of the proposed programme components. Attendance by invitation only.

Speakers

Need to generate a shortlist of speakers in the 5 target applications areas. Speakers should be invited to review what impact 1 TF, 10 TF and 100TF sustained performance would have on industrial R&D applications in their sector. The relative importance of capability versus capacity in terms of R&D impact and the quality of service and support

programmes that would enable them to exploit such facilities. In order of priority we should approach:

- World class UK industrialists using HPC
- World class UK academics using HPC with proven track record of working with industry
- Overseas industrialists using HPC
- Overseas academics with proven track record of working with industry

Next Steps

- Agree this proposal between EPCC, CLRC and EPSRC.
 - Invite Group Leaders at both sites to suggest names for speakers.
 - Secure shortlist of speakers and invite – mid October
 - Discuss sponsorship with NWDA + Scottish Enterprise – end of October
 - Book Venue around DL (preferably Lecture Theatre plus Science Centre) – mid November
 - Programme – mid November
 - Invitation list – mid November
 - Requirements/ QoS questionnaire – December
-
- Work has progressed on reviewing the international portfolio of HPC applications and a report provided to EPSRC titled ‘ HPC and Scientific Opportunity’

4.3 Terascaling Applications (*Dr Martyn Guest*)

The work described below covers the period July-September 2003, 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 Computational Materials

AIMPRO

- Patrick Briddon has developed a new version of AIMPRO, which shows substantial performance improvement on HPCx. We are currently looking to gain a copy of this code for further investigation.

Castep

- We have made a major new optimization (for metal and metal-like substances) by re-casting the density-mixing eigenvalue minimization algorithm. This saves up to 30% of job time. The re-casting has been agreed with the Castep Developers' Group (CDG).
- Martin Plummer attended a full UKCP meeting representing HPCx, and gave a presentation, demonstrated optimizations, and participated in a discussion on whether any users should transfer resources to the CSAR Newton service for jobs with smaller numbers of processors. A couple were keen to transfer, however most people wanted to remain with HPCx.
- We are looking at the implementation of a Gamma-point version which takes full advantage of real/complex FFTs for single k-point calculations.
- Work will start with Barbara Montanari and Matt Probert on developing the internal Castep task-farming routines, which are used for bringing together and analysis of data after main runs, according to specific scientific requirements.

Crystal

- More of the analysis code has been parallelized, so that results can be generated more readily from large scale runs.
- Implemented restarts off eigenvectors. This also makes it possible to avoid diagonalization in the analysis stage, which is a major benefit.
- We have shown that the capability demonstrator calculation on Rusticyanin is indeed possible. This feasibility study has involved
 - a) Generating a suitable structure for the calculation, through collaborations with S. Hasnain's group (Synchrotron Radiation Department). The main problem here is the positions of the hydrogen atoms, especially those in the water molecules.

b) Checking that the basis set is of sufficient quality for a sensible calculation.

c) Running the calculation for a very small number of cycles on HPCx.

This calculation only just fits on HPCx and can not be run using 8 processors per LPAR due to its memory requirements, but running with 6 processors per LPAR was successful.

At present none of the Householder diagonalization methods are suitable for large-scale calculations in CRYSTAL due to memory constraints.

- An article has been written for the HPCx newsletter.
- CRYSTAL has been used to generate diagonalization test cases for the work on eigensolvers.

VASP

- Work has commenced on the *ab initio* quantum-mechanical molecular dynamics VASP code (recently received from Price and Alfe) with a view to its parallel optimization on HPCx. The current scalability is extremely poor.

4.3.2 Molecular Simulation

PLATO

- Gavin Pringle presented a talk at the PLATO workshop. Feedback from this workshop suggested there was considerable interest in an extensive benchmark activity on HPCx. Hence effort within this quarter has focussed on this activity and the results made available to the interested users. The remaining effort has focussed on integrating the chosen diagonaliser into the code.

NAMD

- This code scales well to at around 256 processors and has been awarded a bronze star. Significant effort has been spent benchmarking and profiling the different versions of this code to understand the performance bottlenecks.

AMBER

- We have been collaborating with Robert Duke, who has obtained significant performance improvement on amber 7. These optimisations are likely to be incorporated into amber 8. Scaling is however still poor: profiling suggests this is primarily due to collective communications and that the code would benefit from developing a mixed MPI / OpenMP implementation. Work is focused on developing such an implementation.

DL-POLY

- Work has started on the introduction of Fortran 90 features into the code. This is to help maintainability.

4.3.3 Atomic and Molecular

PRMAT

- We have benchmarked a new eigensolver stage of PRMAT (using the new ScaLapack diagonalizer). The resulting code runs around 50% faster on 128 processors. The impact on the overall speed-up of the code is around 10-30% depending on the dataset.
- We are currently developing a new version of the eigensolver stage that takes advantage of BLACS sub-grids to distribute sector diagonalizations among groups of processors. This approach is expected to scale very well.

Armour code

- We have commenced analysis of an electron/positron scattering code from Prof. E.A.G. Armour, Chair of CCP2, which is used to generate potential surfaces for helium anti-hydrogen scattering, with a view to efficient parallel implementation on HPCx.

4.3.4 Molecular Electronic Structure

Global Arrays and LAPI

- We are still awaiting feedback from IBM on the GA/LAPI and ScaLAPACK/MPI matrix multiply benchmarks. A revised version of the GA-based matrix-multiply routine shows some improvement in performance, but remains hindered by the underlying LAPI implementation. A similar matrix diagonalisation benchmark based on MPI and LAPI (see "libraries" below) has also been forwarded to IBM.
- A recent release of LAPI has cured the problem with segmentation violations that hindered previous attempts to develop a revised matrix-matrix multiply algorithm within the GAs. This exploits non-blocking algorithms and LAPI vector functionality (e.g. LAPI-GETV); the work is still in progress.

GAMESS-UK

- The *newscf* module has been implemented in Fortran 90. This work involved devising suitable objects that can be handled opaquely by a driver program, in the sense that the driver need not know whether these are distributed objects or not. Suitable data types have been devised and a set of routines to operate on these data types has been written. The serial implementation is all but finished and work is starting on the parallel

implementation. This also relates to CRYSTAL k-point parallelism, as the data structures and the library of operations are sufficiently flexible to deal with that code as well. It is intended that eventually both codes will use this library, thus simplifying code maintenance.

- Implementation of the parallel DFT analytic 2nd is now complete, following major enhancements to memory utilisation. Performance and scalability tests are currently in progress.
- An initial implementation of the QM/MM functionality involving the hybrid GAMESS-UK/CHARMM code is now complete.

Performance Evaluation of QC Codes

- CPMD: Following the impressive performance figures achieved by Alessandro Curioni (IBM Zurich Research Laboratory) on HPCx (1 Tflop sustained on 1280 CPUs on a 1000 atom system), we have now implemented and tested the latest MPI release of the code (3.7.2). Performance figures on smaller systems (C₁₂₀ and Si₅₁₂) suggest that the hybrid OPENMP/MPI implementation (as used by Curioni) will be required to circumvent the communication bottleneck associated with MPI_AllToAll.
- QM/MM: As part of this implementation, the CHARMM code has been implemented on HPCx. In common with the Alphaserver SC and SG Origin, scalability of this replicated data molecular dynamics code is poor.

Gaussian

- The various licensing issues with this program have been resolved and the code is now available on HPCx. A reasonable number of users have already been given access to the code. Future work will involve carrying out a feasibility and scaling study for a potential HPCx user, these results should also be of general use to more users.

4.3.5 Computational Engineering

Turbulence Code

- A scaling and feasibility study was carried out on this code for a potential HPCx user. This work is now complete and the report available on the HPCx web site.

4.3.6 Environmental Science

POLCOMS

- The parallel coupled POLCOMS/WAM code is currently being tested by users at the Proudman Oceanographic Laboratory in collaboration with the UK Met Office.

Ensemble Modelling

- Following a request from Lois Steenman-Clark, the Terascaling team investigated and developed an ensemble modelling capability on HPCx. MPH (a program which allows multiple climate models to be task-farmed) has been installed and tested. The I/O performance on HPCx has been investigated when several tasks are accessing the disks at the same time, as we would expect from a task-farming scenario. This work is also highly relevant to other users, and has been written up as a Technical Report, which is available on the HPCx web site. This work is now complete.

4.3.7 Libraries

- Analysed timings of several PeIGS implementations that are currently available.
- Our parallel diagonalizer test code, with calls to PeIGS (MPI and LAPI) and ScaLapack, has been sent to IBM to make them aware of our concerns over poor scaling performance.
- Feedback from a specific user suggested a report on the current status of diagonalizers would be beneficial for many codes on HPCx. Hence future work will focus on such a report.
- The predominance of codes on HPCx which require diagonalizer routines has led to the formation of a diagonalizer group, as part of the tera team.

4.3.8 Tools

TotalView

- TotalView has been upgraded to version 6.3 which has support for the Global Arrays Toolkit and expanded IBM support.

DDT

- Streamline have fixed many of the problems with DDT and we are now testing it on some real applications codes. One such test exposed a problem with IBM's implementation of dbx (which underlies DDT).

Vampir

- The Vampir has been upgraded to version 4.0. This has support for the Global Arrays Toolkit and for Java and is also thread-safe.

HPMcount

- HPMcount has proved extremely useful for HPCx staff, when profiling and benchmarking application codes. The quantity of information reported is however extensive. Hence current work is focussed on developing a short report on how to use and interpret the results obtained from HPMcount. This will be made available on the HPCx web site for HPCx users.

4.4 Software Engineering (*Dr Stephen Booth*)

4.4.1 Low Level Communications

Investigation of the low-level communication protocols of HPCx, in particular, MPI, MPI-2 single sided, LAPI and GA-tools.

- *LAPI MPI*: One of the outgoing EPCC MSc students wrote his dissertation on porting the point-to-point communication module of the T3E-MPI library (originally developed at EPCC) to HPCx using LAPI. He successfully completed and submitted this work. This project was very successful. Though not superior to the native IBM library the performance of his code is comparable. In addition this work has exposed a great deal of information about the implementation of the LAPI library. We will compare this implementation with the new federation version of IBM MPI (which is also based on LAPI) when this becomes available.

4.4.2 Grid Integration

- There are now 12 users with certificates registered to use Globus-2 on HPCx. Of these 4 correspond to support staff and 8 to normal users. Globus does not yet seem to be generally required by the HPCx user community although it is significant for particular groups of users.
- We are implementing a feature to allow users to manage their certificates through the SAF rather than by requests to the helpdesk.
- *RealityGrid (e10)*: The RealityGrid consortium uses codes that are grid-enabled to allow them to perform computational steering while the simulation is running. Though these code are running very successfully on HPCx and can be submitted via the Grid, this computational steering functionality requires the program to be able to open and receive network connections to hosts on the wider internet while the program is running. Like many large computational clusters the compute nodes in HPCx are connected to an internal private network; only the login nodes are connected to the external internet. We are currently developing port forwarding technology that will run on the login nodes and allow the RealityGrid consortium to perform computational steering on HPCx. An initial version of this now exists and is being tested on the test and development system. The port forwarder will also required for the proposed Tera-Grid demonstration intended to run at Supercomputing this year so we plan to have this work fully deployed by then.

4.4.3 MPI and Mixed Mode Programming

- *Single sided Communications:* A report comparing the performance characteristics of MPI-2 single sided and LAPI communications as been completed and is on the HPCx Web-pages.
- *MPI performance:* A technical report investigating the most efficient way of performing boundary exchanges using MPI has been completed and is on the HPCx Web-pages.
- *Mixed mode:* One of the outgoing EPCC MSc students investigated the performance of Mixed mode codes from the ASCI Purple benchmarks using HPCx. We will produce a technical report summarizing his results.

4.4.4 Java Performance

- One of the outgoing EPCC MSc students investigated the performance of the JAVA-GRANDE benchmark suites using a variety of systems including HPCx. We will extract the relevant section of his results to generate a report on the performance of JAVA on HPCx.

4.4.5 System Administration Functions (SAF)

- The following features have been added to the SAF during the last 3 months:
 - Form for users to request new passwords.
 - Improved capacity planning and project profile code.
 - Opt-out button for user mailing lists.
 - Performance enhancements
 - Improvements to reporting, eg, for capacity planning.
- Ongoing work includes
 - Globus certificate management.
 - Performance improvements when generating reports.
 - Work preparing the SAF for the Phase 2 upgrade.

4.5 Operations and Systems (*Mr Mike Brown*)

4.5.1 Staffing

No change. As noted in the last report, the level of actual on-site and on-call coverage has been the subject of a review, and the cover pattern has been reduced (although it still remains greatly in excess of the "core hours" contractual minimum).

The level of coverage will remain under review, and a more extensive coverage pattern could be restored if stability is adversely affected in the future, or it could be further contracted if it is considered that the applied coverage pattern is not the best use of resources.

There has been no diminution of effort, merely a more efficient re-deployment of resources.

4.5.2 Software Test & Development System

This machine was commissioned in June, and is now in use as a trial and development platform.

It has proved immensely beneficial to have this system available. Currently it is being used to pre-trial the TeraGrid demonstration environment, and to test out HSM. Neither function could have been safely tried out on the phase 1 production service machine.

4.5.3 WWW Server

The WWW server was replaced with a small p615 system, to free up the previous hardware platform so that it could be used as the CWS for the CSM test Phase 2 cluster.

4.5.4 Phase 2 development system

The first phase of the phase 2 (CSM) development system was installed in August. This consists of 2 x 32-way p690 I/O servers, and 2 x 32-way p690+ compute servers. The full complement of phase 2 disc was also installed.

4.5.5 Reliability/Stability

Overall, reliability has stabilised although there was a slight upturn of problems in August, probably due to the disruption in the computer room when the CSM phase 2 system was installed, along with the associated infrastructure works.

The major source of IBM-attributed SEV 1 incidents remains GPFS/switch related, it is believed the complexity of the hardware/software is such that these problems may never be eliminated entirely, but it is hoped that they will reduce to a manageable level.

4.5.6 HSM under TSM

HSM under TSM remains under test and evaluation on the test and development system. There have been a number of issues that have been raised with IBM on this product, the objective remains to provide the users with basic functionality at least as good as was available with DMF on the CRAY services. There is still some way to go, although IBM are receptive to the issues.

Additional on-line disc space was made available to any users who might have been adversely affected by the delays in providing a HSM service, however, to date no access to such space has been requested.

4.5.7 Phase 2 Migration

Continued planning for the phase 2 migration is under way with IBM, who have placed a skilled and competent technical project manager on site to take forward the planning with Edinburgh and CCLRC staff.

4.6 Staffing

<i>AV</i>	<i>July</i>	<i>August</i>	<i>September</i>
DL	4.3	3.4	4.7
EPCC	8.1	6.4	6.2
Total	12.4	9.8	10.9
<i>Systems</i>	6.7	5.1	6.7

5 Summary of Performance Metrics

<i>Metric</i>	<i>TSL</i>	<i>FSL</i>	<i>July</i>	<i>August</i>	<i>September</i>
Technology serviceability	80%	99.2%	99.5%	98.5%	99.6%
Technology MTBF (hours)	200	300	366	183	418
Number of AV FTEs	7.5	10	12.4	9.8	10.9
Number of training days per month	30/12	40/12	33/7	33/8	35/9
Non in-depth queries resolved within 3 days	85%	97%	100.0%	98.5%	100.0%
Number of A&M FTEs	3.75	5.75	6.7	5.1	6.7
A&M serviceability	80%	100%	100.0%	99.9%	99.9%

<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, ie, by the end of September we had run a total of 35 training days over 9 months.

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

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

PPARC Projects

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

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

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

Early User Projects

<i>Code</i>	<i>Title</i>	<i>PI</i>
y001	Materials	Dr Patrick Briddon
y002	DNS of Turbulent Flow	Prof Neil Sandham
y003	Multi-photon and Electron Collision Processes	Prof Ken Taylor
y004	Materials	Prof Jonathon Tennyson
y005	UKAEA	Dr Tim Hender
y006	UK Car-Parrinello Consortium	Prof David Price
y007	Climate Modelling	Ms Lois Steenman-Clark

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