HECToR Annual Report 2008

01 January – 31 December 2008

Issue: 1.0

Contents

1 Introduction

This report covers the period from 1 Jan 2008 at 0800 to 1 Jan 2009 at 0800.

The next section of this report contains an executive summary for the year.

Section 3 summarises service availability, performance and utilisation statistics for the year. Section 4 covers the Helpdesk statistics. Systems support is covered in Section 5, with Science Support in Section 6.

The Appendices define some of the terminology and incident severity levels and list the current HECToR projects together with their overall utilisation profile to date.

This report and the additional SAFE report are available to view online at http://www.hector.ac.uk/about-us/reports/annual/2008.php

2 Executive Summary

Following the early start of the HECToR high performance computing system on the Cray XT4 towards the end of 2007, this year has seen the consolidation of the service and the migration of some users from HPCx to HECToR. Many users have taken advantage of the thousands of processing cores on HECToR and scaled their applications to enable deeper research and better science.

The highlights of the service over the year included:

- Helpdesk statistics for the year were excellent. A total of 3566 queries were handled with targets exceeded in most areas. A handful of technical assessments missed SLA in Q1/Q2 but this has since been addressed.
- Feedback received on science support has been positive. Take-up on CSE training courses has been mixed, with HECToR specific and basic HPC courses proving the most popular. The CSE team has attended a number of national and international meetings to promote HECToR and to share experiences with other HPC users and support services.
- A large part of the Cray CoE effort was focused on ensuring that EPSRC's capability projects were able to fully and efficiently use HECToR in a timely fashion. In addition to those projects, the Centre worked on a wide range of applications coming into contact with a total of twenty-seven individual applications over a broad range of scientific disciplines. In other notable areas, the Centre over-delivered compared to its requirements in providing training to the service users and consortia partners.
- The Cray X2 Vector went live in July 2008 as planned. There have been no major issues to date. Initial utilisation has been relatively low, at an average of 22%.
- The HECToR and HPCx services are designed to operate in a complementary manner. Throughout the year a number of consortia have taken the opportunity to transfer AUs from HPCx to HECToR. Users have also been encouraged to transfer AUs from the XT to the X2, although as yet there has been no take up on this.

- Average utilisation for the year was 64%. The Capability Challenge projects ended in September which led to a reduction in utilisation. A number of new projects have recently been approved which should help to build the utilisation back up again.
- In 2008 there were 46 technology attributed service failures. 70% of these were software failures, with the remaining 30% hardware fails. Fixes have been provided and applied for all but one software failures.
- Testing of the proof of concept archiving solution is ongoing on the test and development server. A new version of the job scheduling tool PBS is also under test. PBS V10 is required in order to support reservations and interactive queues on HECToR.
- Options for the HECToR Phase II upgrade, including a combination of quad core followed by the next generation Cray MPP supercomputer, are currently under review.

3 Quantitative Metrics

3.1 Reliability

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

	1Q08	*2Q08*	*3Q08*	*4Q08*	*2008*
Incidents	84	104	119	83	407
Failures	17	13	10	11	51

The incidents above are primarily related to single node failures. Details on both the service failures and single node fails in 2008 can be found in Section 5.

3.1.1 Performance Statistics

- MTBF = (732)/(number of failures in a month)

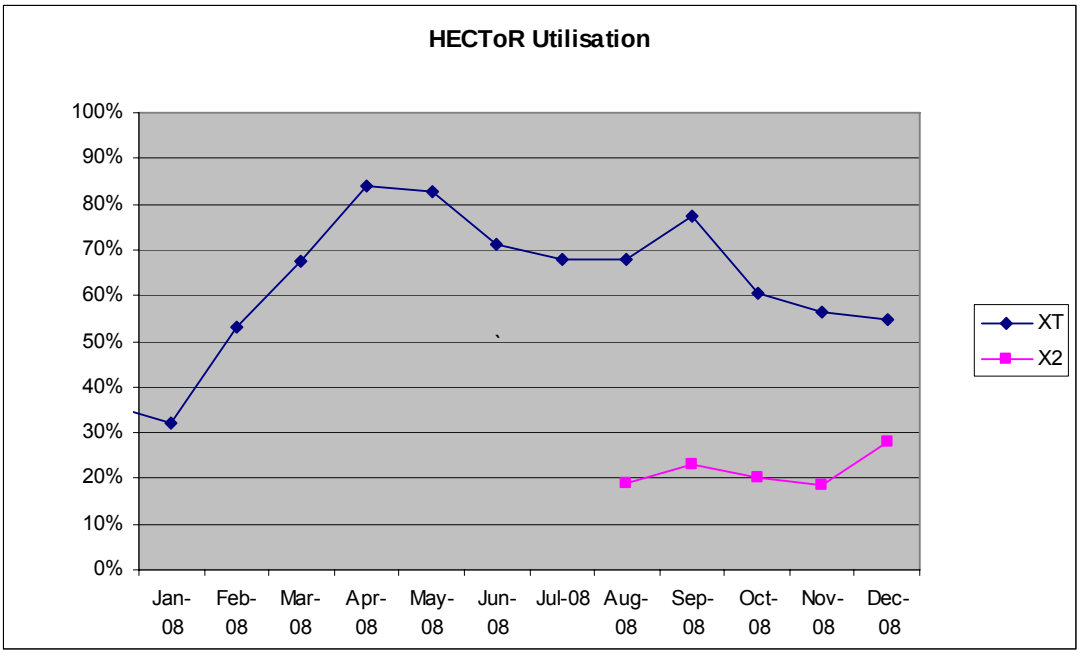
  Quarterly MTBF = (3x732)/(number of failures in a quarter)

  Annual MTBF = (12*732)/(number of failures in a year)

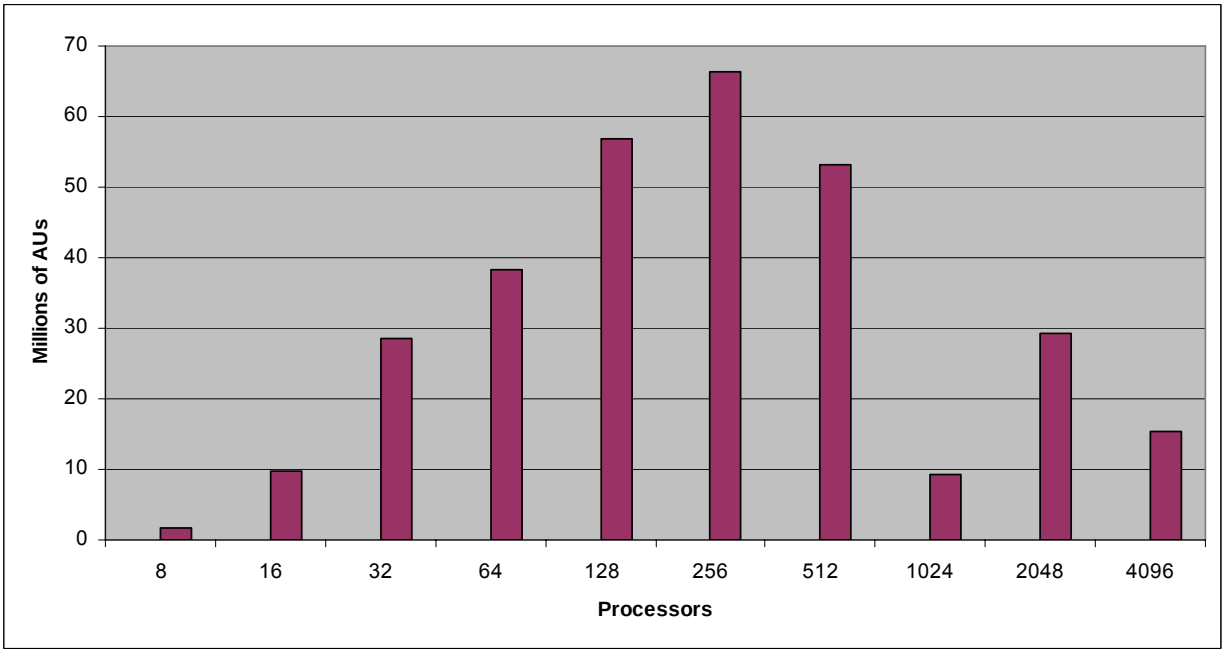
Attribution	*Metric*	*1Q08*	*2Q08*	*3Q08*	*4Q08*	*2008*
Technology	Failures	15	13	8	10	46
	MTBF	146	169	275	220	191
Service Provision	Failures	1	0	2	1	4
	MTBF	2196	∞	1098	2196	2196
External	Failures	1	0	0	0	1
	MTBF	2196	∞	∞	∞	8784
Overall	Failures	17	13	10	11	51
	MTBF	129	169	220	200	172

3.2 HECToR Utilisation

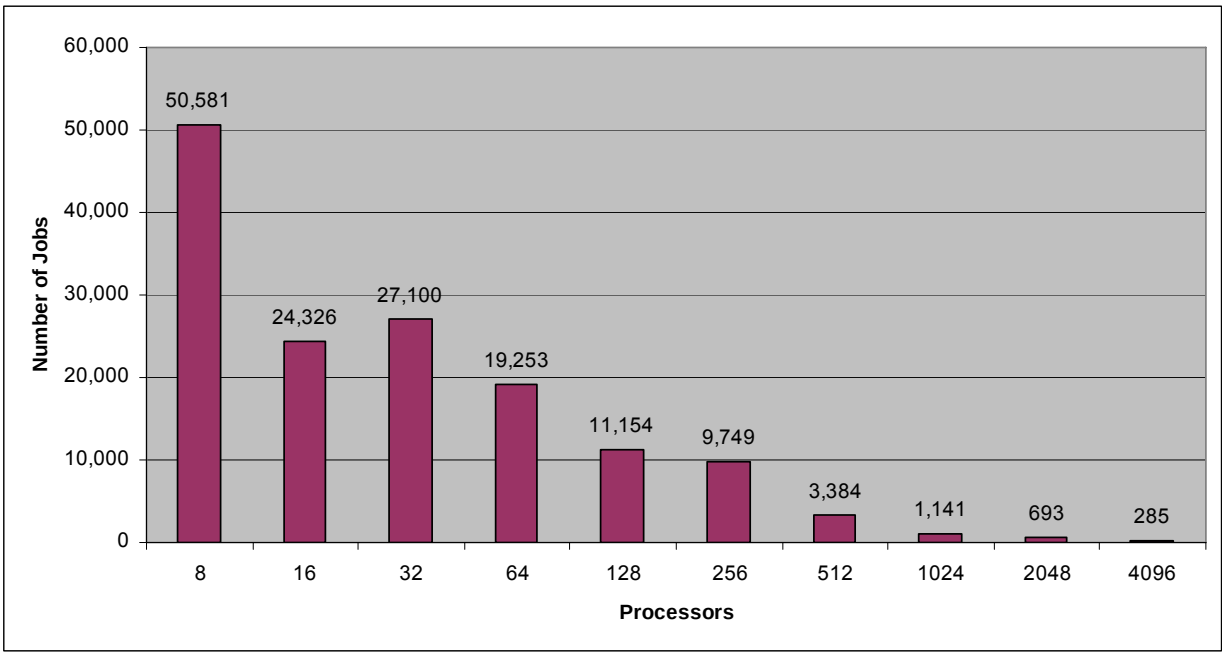
3.2.1 Overall Utilisation



3.2.2 XT Utilisation by Job Size



3.2.3 XT Number of jobs per queue



3.2.4 Usage by Consortium

3.2.4.1 XT Usage by Consortium

A total of 480,028,032AUs were available during this period

Project	AUs	Number of Jobs	%age of Use	Utilisation
y01	0	1848	0.00%	0.00%
y02	901,000	4970	0.30%	0.20%
y03	2,805	1482	0.00%	0.00%
y04	1,595	62	0.00%	0.00%
y05	18,800	1035	0.00%	0.00%
y06	29	6090	0.00%	0.00%
y07	0	19	0.00%	0.00%
z01	481,293	4012	0.20%	0.10%
z02	50,054	1610	0.00%	0.00%
z03	329,748	8464	0.10%	0.10%
Internal Total	**1,785,324**	**29592**	**0.60%**	**0.40%**
c01	6,485,295	5962	2.10%	1.40%
e01	1,298,377	336	0.40%	0.30%
e05	5,892,541	6099	1.90%	1.20%
e10	4,512,119	2712	1.50%	0.90%
e100	10	5	0.00%	0.00%
e102	15,165	30	0.00%	0.00%
e109	0	1	0.00%	0.00%
e24	8,289,416	4943	2.70%	1.70%
e34	14,902	85	0.00%	0.00%
e35	29,939	138	0.00%	0.00%
e42	2,196,869	2264	0.70%	0.50%
e59	16,321	222	0.00%	0.00%
e63	3,546,013	1045	1.10%	0.70%

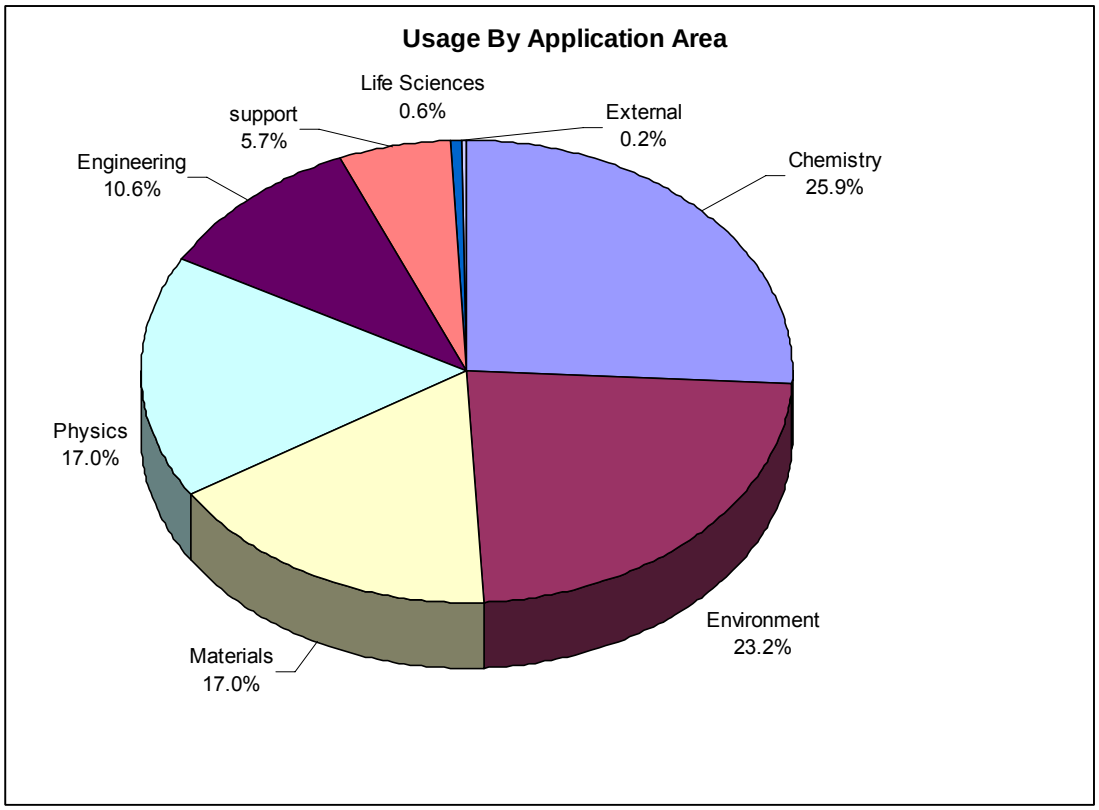
e68	19,350,961	11649	6.30%	4.00%
e69	71,136	746	0.00%	0.00%
e70	132,667	388	0.00%	0.00%
e71	39,866,282	4986	12.90%	8.30%
e72	8,459,145	6178	2.70%	1.80%
e73	11,080,774	1156	3.60%	2.30%
e74	32,410,752	4756	10.50%	6.80%
e75	24,659,485	1466	8.00%	5.10%
e76	1,682,903	152	0.50%	0.40%
e77	40,993	352	0.00%	0.00%
e78	91,184	114	0.00%	0.00%
e79	20	35	0.00%	0.00%
e80	30,347,190	394	9.80%	6.30%
e81	100	249	0.00%	0.00%
e82	90,727	290	0.00%	0.00%
e83	100,589	56	0.00%	0.00%
e84	16	11	0.00%	0.00%
e86	103,105	21	0.00%	0.00%
e87	91,363	51	0.00%	0.00%
e88	9,839,902	1970	3.20%	2.00%
e89	17,208,271	6148	5.60%	3.60%
e90	286	83	0.00%	0.00%
e91	106,693	68	0.00%	0.00%
e92	50,856	113	0.00%	0.00%
e93	508,680	567	0.20%	0.10%
e94	5,570	44	0.00%	0.00%
e95	102,440	151	0.00%	0.00%
e98	25,118	93	0.00%	0.00%
e99	104,662	341	0.00%	0.00%
u02	147,917	639	0.00%	0.00%
u03	357,175	121	0.10%	0.10%
u10	685,231	1506	0.20%	0.10%
u11	64	1	0.00%	0.00%
EPSRC Total	**230,019,224**	**68737**	**74.40%**	**47.90%**
n01	5,161,009	5412	1.70%	1.10%
n02	25,125,895	28037	8.10%	5.20%
n03	41,386,806	14274	13.40%	8.60%
n04	1,060,192	1634	0.30%	0.20%
u07	13,332	54	0.00%	0.00%
NERC Total	**72,747,234**	**49411**	**23.50%**	**15.20%**
b01	1	11	0.00%	0.00%
u04	1,714,168	549	0.60%	0.40%
BBSRC Total	**1,714,169**	**560**	**0.60%**	**0.40%**
T01	6,178	115	0.00%	0.00%
x01	483,338	1476	0.20%	0.10%
External Total	**489,517**	**1591**	**0.20%**	**0.10%**
d01	861,902	119	0.30%	0.20%
d02	1,138,659	39	0.40%	0.20%
d03	13,467	684	0.00%	0.00%
d04	26,018	889	0.00%	0.00%
d05	30,583	26	0.00%	0.00%
d07	263,587	46	0.10%	0.10%
DirectorsTime Total	**2,334,217**	**1803**	**0.80%**	**0.50%**
Total	**309,089,684**	**151694**	**100.00%**	**64.40%**

3.2.4.2 X2 Usage by Consortium

A total of 7,915,622AUs were available during this period.

Project	AUs	Number of Jobs	%age of Use	Utilisation
y01	0	2	0%	0.0%
y02	3,180	332	0.20%	0.0%
y03	0	67	0%	0.0%
y05	29	4	0%	0.0%
z01	180,836	572	10.50%	2.3%
z02	25,243	101	1.50%	0.3%
z03	50,550	1515	2.90%	0.6%
z06	233	123	0%	0.0%
Internal Total	**260,071**	**2716**	**15.10%**	**3.3%**
e01	372,613	569	21.70%	4.7%
e10	829,092	365	48.20%	10.5%
e24	1,859	151	0.10%	0.0%
e75	212,842	786	12.40%	2.7%
e88	10,578	9	0.60%	0.1%
EPSRC Total	**1,426,986**	**1880**	**82.90%**	**18.0%**
n01	16,066	161	0.90%	0.2%
n02	14,400	552	0.80%	0.2%
n03	1,370	53	0.10%	0.0%
NERC Total	**31,837**	**766**	**1.90%**	**0.4%**
d04	1,427	20	0.10%	0.0%
DirectorsTime Total	**1,427**	**20**	**0.10%**	**0.0%**
Total	**1,720,320**	**5382**	**100%**	**21.7%**

3.2.5 HECToR Usage by Application Area



3.3 Performance Metrics

Metric	TSL	FSL	January	February	March	April	May	June	July	August	September	October	November	December	Annual Average
Technology reliability (%)	85.0%	98.5%	98.6%	97.8%	98.7%	94.5%	99.3%	96.7%	96.2%	99.8%	100.0%	97.8%	99.2%	98.5%	98.1%
Technology MTBF (hours)	100	126.4	183.0	104.6	183.0	122.0	366.0	146.4	104.6	732.0	∞	244.0	244.0	183.0	191.0
Technology Throughput, hours/year	7000	8367	8468	8347	8232	8205	8551	7750	8311	8356	8476	8378	8575	8257	8325
Capability jobs completion rate	70%	90%	93.5%	97.0%	98.0%	95.2%	97.9%	93.3%	91.3%	100.0%	100.0%	98.4%	92.0%	100.0%	96.5%
Non in-depth queries resolved within 1 day (%)	85%	97%	99%	100%	99%	100%	100%	100%	100%	98%	100%	99%	100%	98%	99%
Number of SP FTEs	7.3	8.0	9.5	7.9	8.0	8.0	8.2	8.4	8.0	8.2	9.0	8.7	8.0	8.0	8.3
SP Serviceability (%)	80.0%	99.0%	99.9%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	97.8%	98.1%	100.0%	100.0%	99.4%	99.7%

Colour coding:

Exceeds FSL	
Between TSL and FSL	
Below TSL	

4 Helpdesk

A total of 3566 queries with a specified service metric were completed in this period.

4.1.1 Helpdesk Targets

Metric	Pass	Total	Fraction	Target
All queries finished in 1 day	2949	2966	99.4%	97%
Admin queries finished in 1 day	2617	2632	99.4%	97%
Queries assigned in 30 min	3536	3566	99.2%	97%
Technical assessments in 10 days	90	95	94.7%	97%

4.1.2 Queries by Service Metric

Service Metric	Queries	Percentage
Automatic	1872	52.5%
Admin	760	21.3%
In-depth	505	14.2%
Technical	334	9.4%
Technical assessment class-1	70	2.0%
Technical assessment class-2	25	0.7%

4.1.3 Queries by Category

Query Category	Queries	Percentage
New User	538	14.8%
Set user quotas	409	11.2%
Set group quotas	355	9.8%
Disk, tapes, resources	247	6.8%
Access to HECToR	220	6.1%
None	213	5.9%
New Password	210	5.8%
Add to group	207	5.7%
3rd Party Software	165	4.5%
Compilers and system software	163	4.5%
Batch system and queues	157	4.3%
User behaviour	134	3.7%
Other	120	3.3%
New Group	97	2.7%
User programs	72	2.0%
Login, passwords and ssh	67	1.8%
SAFE	58	1.6%
Join Project	47	1.3%
Node Failure	38	1.0%
Performance and scaling	29	0.8%
Courses	23	0.6%
Static website	15	0.4%
Delete from group	13	0.4%

4.1.4 Queries by Handler Category

Handlers	Total	In-depth	Automatic	Technical	Admin	Technical assessment class-2	Technical assessment class-1	Percentage
CSE	287	198		4	8		70	8.1%
OSG	2147	65	1871	135	76			60.2%
USL	1001	158	1	171	653	18		28.1%
Cray Systems	94	74		17	3			2.6%
Cray CoE	14	9		4	1			0.4%
Other	23	1		3	19			0.6%

4.1.5 Helpdesk General

Regular communications have been sent to users regarding changes to the working environment, upcoming courses and general HECToR news throughout the year. These mailings are archived in the SAFE and can be viewed there by all users.

The HECToR 'Code of Conduct' has been updated as the service has matured. This will continue to be a living document. The user FAQs and the HECToR User Guide are also updated on an ongoing basis.

A working team was setup in 4Q08 with members of both the Helpdesk team and the CSE team to review and update the layout of the main HECToR website, in particular links to the dCSE section. This work is in progress and will continue in 2009.

5 Systems Hardware

5.1 HECToR Technology Changes

Phase 1b of HECToR's technology provision arrived in June. This consisted of a storage upgrade and two cabinets of Cray's latest Vector processing system, the X2.
One X2 cabinet was integrated into the main HECToR XT4 system and the other was attached to the TDS XT4 for test and development activities.

Several minor software updates have been applied to the base Unicos/lc 2.0 Operating system to fix problems and provide new features. We are currently running version 2.0.62.

5.2 Severity-1 Incidents

5.2.1 Technology Failures

Cray technology has been responsible for a total of 46 Severity-1 incidents during 2008. Fourteen of these incidents have been due to hardware failures which are typically Compute-module power related or HSN (High Speed Network) failures. Highly complex, large scale resources such as HECToR will inevitably see hardware failures due to the vast number of components that make up the system, but only a very small percentage of failures result in Severity-1 incidents.

Software and firmware related problems accounted for 32 Severity-1 incidents during 2008. Fifteen of these incidents were caused by three separate software bugs. Each of the bugs occurred five times before a tested and verified fix could be applied to the system. The first of these major bugs was seen in the February-March timeframe and was fixed with the introduction of Unicos/lc 2.0.49 at the beginning of May. The second of the major bugs was triggered by a specific type of HSN error and occurred during the months of June-July - this was fixed with the installation of a patch in early August. The final major bug was related to the lustre filesystem. This was seen mostly in December and has been fixed by applying patches to the XT4 and X2 systems.

Four incidents were caused by Disk controller errors that were fixed with later versions of the manufacturer's firmware and most of the remaining issues were fixed in routine software updates or in a few cases configuration changes. There is one remaining software incident from 2008 that has not been resolved – insufficient evidence exists in the system dump to progress this particular problem.

5.2.2 Service Provision Failures

There were 4 service provision failures during 2008. These were attributed to a network failure, failure of the HECToR web server, and two plant cooling related problems.

5.2.3 External Failures

There was 1 failure in March as a result of an external network failure.

5.3 Single Node Failures

Single node failures can be responsible for the loss of individual user applications. These errors can be broken down into three main categories:

(1) Un-correctable Memory Errors (UME's)

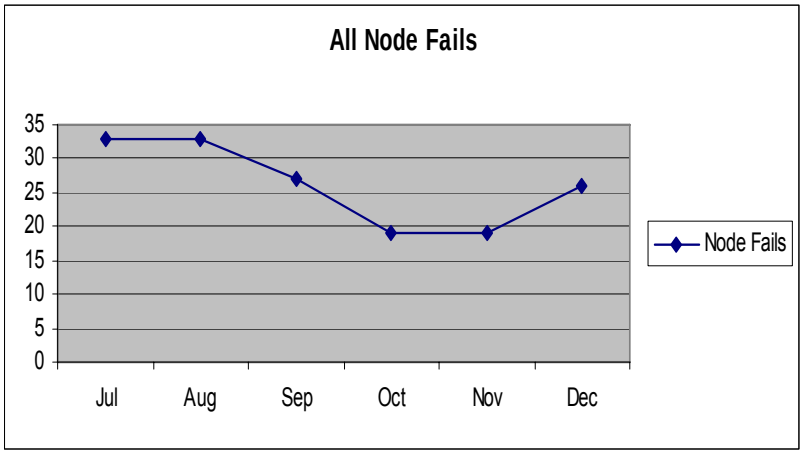
(2) AMD Opteron cache errors

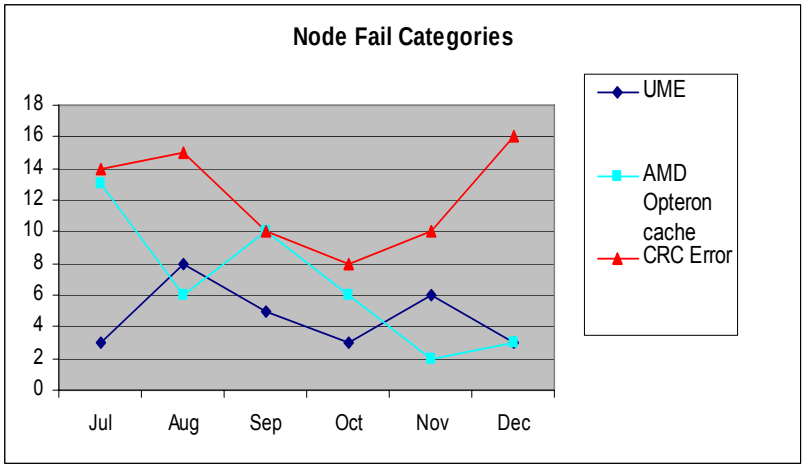
(3) CRC errors from HSN packets

Un-correctable Memory Errors (UME's)
Un-correctable memory errors are usually caused by a failure in one of the memory DIMMs associated with the node. In the HECToR system there are 22,848 DIMMs. Each node has four memory DIMMs arranged in two banks and when a UME occurs it is usually only possible to determine in which bank the error has occurred, but not which of the two DIMMs was responsible for the failure. At HECToR we take an aggressive policy on UME's by replacing both DIMMs in a bank on the first failure, rather than moving DIMMs to determine which of the two components is actually faulty. This reduces the number of UMEs and therefore the number of application failures associated with this type of error.

AMD Opteron Cache Errors
Cache memory errors that occur within the AMD Opteron processors manifest themselves in a similar way to UME's. This particular type of error has declined since the introduction of new software at the beginning of October but is still seen occasionally. Most of these cache errors do not repeat, hence our policy is to only replace the AMD Opteron processor in the event of a subsequent failure – this has been necessary in about 5% of cases for this mode of failure.

CRC Errors from HSN Packets
Cyclic Redundancy Check errors can occur on HSN packets due to three main reasons: a corrupt packet, a corrupt packet header or a packet that has been received apparently out of sequence. In each case the destination node receiving the problematic packet will panic. In some cases a clear hardware reason will be logged for the packet problem and this is diagnosed and repaired at the next scheduled maintenance session, but on other occasions there is no evidence of the root cause of the packet corruption. In these cases, all available information is logged including the source and destination node IDs as well as the route taken by the packet. Any common points in the network that are implicated in several failures will become the subject of diagnostic activity during scheduled maintenance sessions but not all errors of this type are thought to be hardware related. This category of error represents the largest proportion of single node failures


All Node Fails
35
30
25
20
15
10
5
0
Jul
Aug
Sep
Oct
Nov
Dec
Node Fails


Node Fail Categories
18
16
14
12
10
8
6
4
2
0
Jul
Aug
Sep
Oct
Nov
Dec
UME
AMD Opteron cache
CRC Error

6 Cray Centre of Excellence

6.1 Executive Summary

The year 2008 saw a transition for the Cray Centre of Excellence for HECToR from the early use period into full service production mode. The centre saw a shift in efforts from getting people started on the early user service to a targeted delivery of expert support. A large part of the year's effort was focused on ensuring that EPSRC's capability projects were able to fully and efficiently use the Cray supercomputer in a timely fashion. In addition to those projects, the Centre worked on a wide range of applications coming into contact with a total of twenty-seven individual applications over a broad range of scientific disciplines. Highlights of the work performed included: validation and scaling of nine applications to be able to use the entirety of the HECToR Phase 1 service; providing in depth support to some of the more challenging software requirements and ensuring access to the resource for those whose needs are above and beyond the traditional HPC consumer.

In other notable areas, the Centre over-delivered compared to its requirements in providing training to the service users and consortia partners. They also participated in the distributed CSE process and Performance Working Groups; provided guidance and support for postgraduate students studying relevant topics in High Performance Computing; and prepared and delivered assessments on the impact of possible HECToR Phase 2 configurations on the UK's HPC application base. In addition to the Centre's core staff, a further six members of Cray's Application team were involved in HECToR activities where such specialist input was needed.

At the end of the first period of service, Cray and the CoE team are looking forward to building upon the now established operational user base to deliver significant improvements in application scaling which results in new science.

6.2 Applications Support

6.2.1 Capability Challenge Projects

The Cray CoE worked with the five Capability Challenge winning projects to ensure the best possible use of the Cray system and where necessary the required improvements to applications were made to ensure that projects could deliver the science they promised. In some cases significant improvements were made to application codes, most notably DL-Poly, CASINO and Incompact3D.

The application DL-Poly was optimised for the specific interests of Professor John Harding (but applicable to other test cases); these yielded a 3 fold improvement in performance. This came from reworking the I/O portion of the code and recoding for much improved scalability. Generic improvements were passed back to the authors and have been included, or are being considered for inclusion, in the main DL-Poly 3 distribution.

The CASINO application is used by Dr. Dario Alfe of University College London. The code uses large wave function data sets which, for performance purposes, are stored in memory

for recurrent read only access. This is a major limiting factor to the scale of problem that scientists want to run. The CoE worked to reduce the effect of this by allowing a single instance of the read only memory per node and allowing all MPI processes on the node to share it equally. This improvement has pushed forward the scale of simulation able to be performed.

Incompact3D is a DNS application utilised by the Aeronautics group at Imperial College. The CoE assisted this group with regards to their HECToR capability project, providing around 30% improvement in performance through a mixture of compiler tuning and restructuring for cache efficiency.

LUDWIG is another of the HECToR Capability codes where the Cray CoE was involved in assisting users to run the code on the HECToR Cray XT nodes. Current HECToR performance of LUDWIG is shown in Figure A below. A Cray X2 port of LUDWIG is also now available and a key routine has been rewritten to use UPC in place of MPI. The Cray CoE is now involved in optimising the UPC performance of this version of LUDWIG.

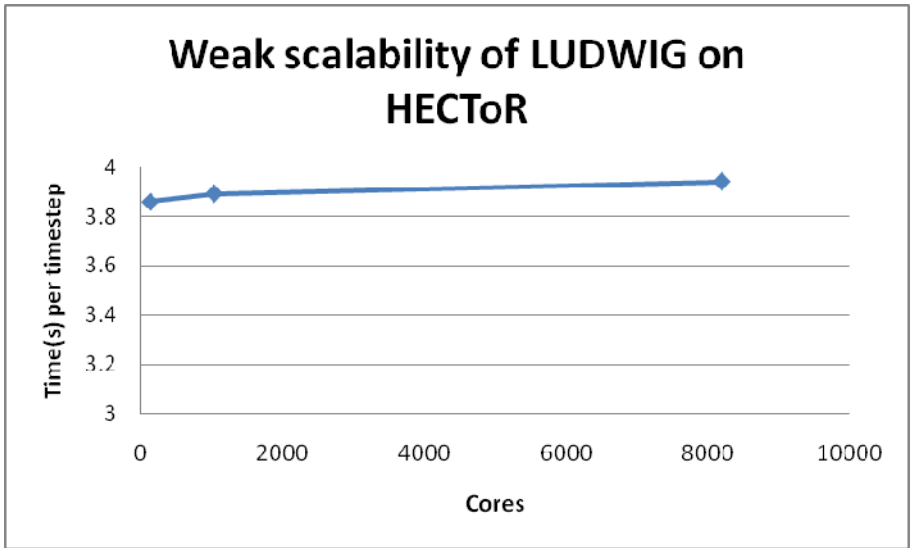


Figure A: LUDWIG performance on HECToR

6.2.2 Cray X2™ Porting and Optimisation

Prior to the service becoming available on HECToR, the Cray CoE exercised a number of the Vector early user codes on a remote Cray X2 to be well prepared for the arrival of the Cray vector supercomputer, both from a porting aspect and performance view. This enabled the user community to begin work immediately after the Cray X2 was installed and to be able to communicate the lessons learnt to all of the HECToR partners. This worked continued after the Cray X2 was available. The main focus of the effort was based around:

- The Unified Model (UM)
- Incompact3D

- LB3D
- SBLI

Training was also made available in the form of two workshops (one provided as part of the NAG training) – more details can be found in Section 6.3.

6.2.3 Other Applications

The Cray CoE has engaged with a number of other projects within the HECToR project, including:

- **SBLI-Shock Boundary Layer Interaction**: SBLI is a code originally developed by Professor Neil Sandham at the University of Southampton for studying turbulent flows. The Cray CoE for HECToR has worked on two aspects of this code; firstly, the single-CPU performance was improved by nearly 20%; and secondly, the scalability of SBLI was improved on HECToR as shown in Figure B.

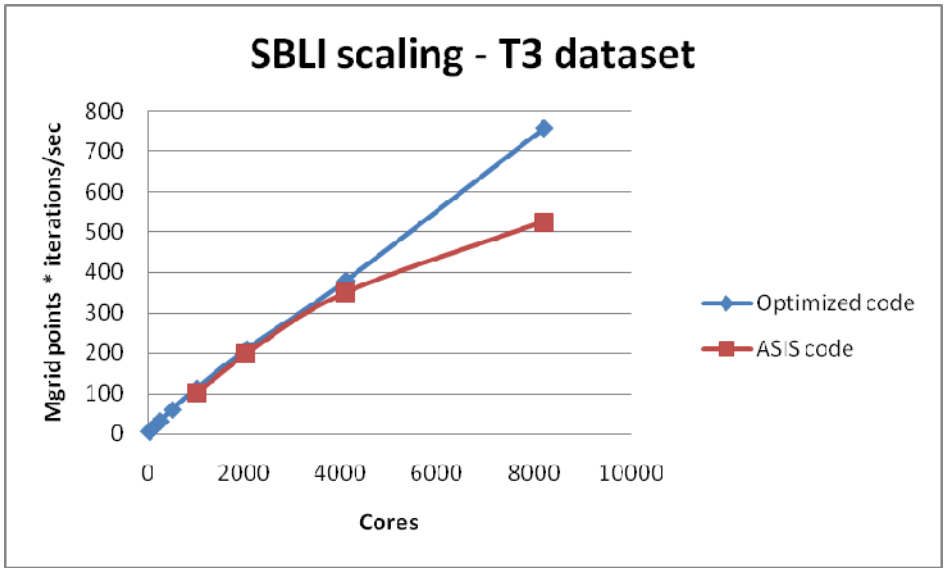


Figure B: Affect of optimisation work in SBLI on scalability and performance.

- **Incompact3D**: Following on from work done for the Capability Challenge project, the CoE identified that a major reworking of the code was needed for scaling to large core counts and worked with the group at Imperial to deliver this work via the dCSE support track.
- **VASP**: The shared memory approach used in CASINO was adapted for use in VASP, this allowed faster alltoall communications and is applicable to nearly all datasets. The improvements vary and depend on the dataset.

- **LB3D & HemeLB**: Work has centred on I/O and single processor optimisation. The single processor optimisation has improved performance by nearly 50% and the I/O work removed a bottleneck in which large runs took exponentially longer to start.

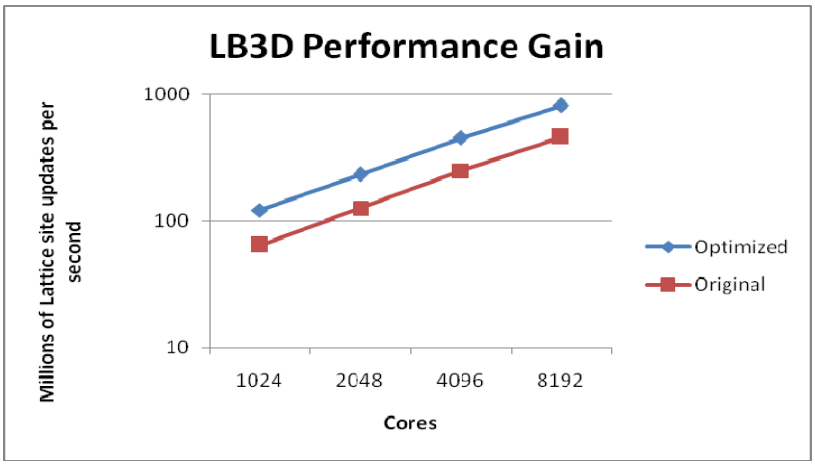


Figure C: Gain achieved with single processor improvements

- **Weather Research and Forecasting Model (WRF):** A group based largely in Leeds University has begun investigating WRF as a scalable application for climate and weather research. The CoE continues to advise and work with the group to ensure good scaling and performance and to bring together the expertise in this application that is available from the US.

- **Unified Model (UM):** The UM is a complex code and the CoE has provided support on a number of levels. Porting and debugging, understanding machine characteristics and its influence on performance (particularly I/O).

- **OpenFOAM**: OpenFOAM is an open source multi-physics package for continuum mechanics which is used both in academia and industry for CFD and solid mechanics simulations. The Cray CoE for HECToR, in conjunction with EPCC, has done a HECToR port of OpenFOAM to enable an industrial customer to run a large number of CFD simulations on HECToR. The CoE continues to support this work.

6.2.4 Phase 2 Development

As with the introduction of the Cray X2™, the CoE has begun work on analysing application performance on quad-core processors to analyse the performance characteristics when increasing the number of cores; these applications use production codes and datasets. This work supports the recent HECToR Phase 2a report provided to the Research Councils in December 2008.

6.3 Workshops and Training Events

The Cray CoE conducted a number of training events in 2008:

- On the 7-10th of April the Cray CoE for HECToR gave an optimisation workshop for the Cray XT4 in Edinburgh. This workshop was given by the Cray CoE staff with assistance from other Cray CoE staff at Oak Ridge National Laboratory (ORNL) and Cray's Application Team.
- On the 24-25th July the Cray CoE, with assistance from Cray technical team, gave a Cray X2 training course to NAG in Oxford.
- On the 4-5th July the Cray CoE, with assistance from Cray UK service staff, gave the first Cray X2 training course to users at the NAG offices in Oxford.

6.4 Seminars

The Cray CoE has given a number of external seminars of interest to the UK HPC community:

- CUG 2008: The Cray CoE gave a presentation on "HECToR, the CoE and Large-Scale Application Performance on CLE" at the 2008 Cray User's Group (CUG) meeting. The presentation highlighted the new HECToR service and some of our initial scalability results from the HECToR early access codes.
- Novel Languages Workshop: The HECToR CoE gave a presentation "Applications of UPC" at the Novel Parallel Programming Languages Day in Edinburgh on 18th June. The presentation discussed the basics of Unified Parallel C (UPC) and use of UPC in HPC applications.
- Cray Technical Workshop Europe: The Cray CoE gave a presentation "CFD Applications at Scale" at the 5th European Cray Technical Workshop 2008 in Edinburgh.
- HPCx Parallel I/O Workshop: The Cray CoE gave a presentation "How To write(101) data on a large system like a Cray XT called HECToR" at the HPCx parallel I/O Workshop. The presentation discussed the I/O capabilities of HECToR, techniques for optimising I/O on the Cray supercomputer and some results from I/O optimisation of several HECToR codes.

7 The HECToR Computational Science and Engineering (CSE) Service

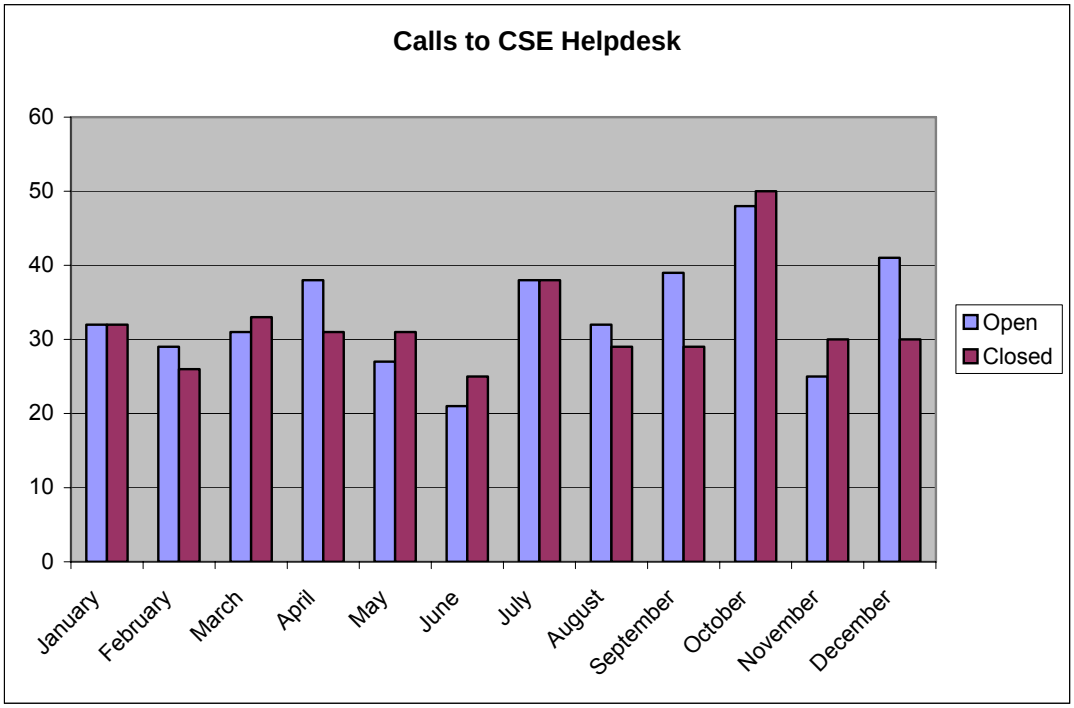
7.1 Overview of the CSE Service

The Computational Science and Engineering (CSE) service exists to help the user community to make the best use of the HECToR hardware by providing training, web-based resources, and assistance with porting, optimisation and tuning of software. The service is provided by the Numerical Algorithms Group Ltd (NAG), a not-for-profit Company with offices in Oxford and Manchester, and over 30 years experience developing mathematical and statistical software. The *Core Team*, made up entirely of NAG staff, responds to in-depth software problems reported by users via the HECToR helpdesk, processes Technical Assessments related to applications for access to HECToR, runs a range of training courses and maintains a range of good practice guides and reference material as part of the service website. The *Distributed Team*, made up of a mixture of NAG staff and staff employed via a contract with a third party, provides dedicated support for particular projects or groups of users for periods of between six months and two years.

7.2 The CSE Helpdesk

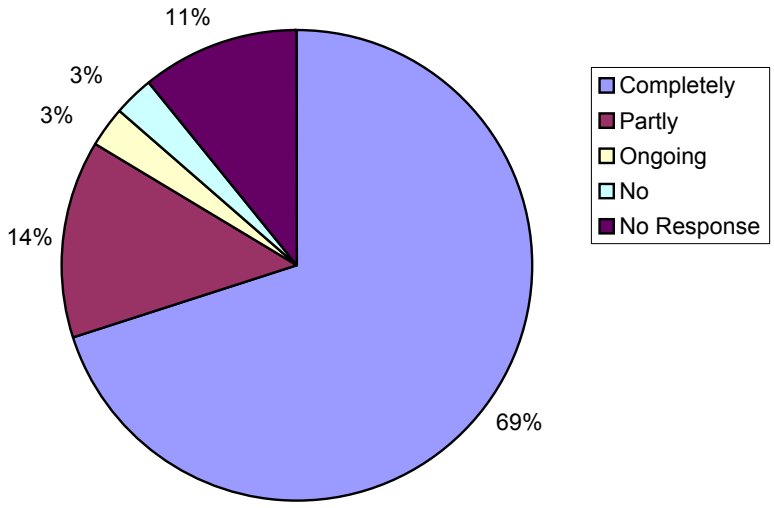
The Core CSE team handles queries from users forwarded by the service helpdesk, carries out technical assessments of applications for HECToR time, undertakes various outreach activities and runs the training courses. The latter two activities are handled separately in this report.

The queries received by the CSE team vary from straightforward requests for advice to requests for assistance in porting, tuning etc. Some queries are resolved straight away while others develop into small projects lasting weeks or even months. The team resolves most queries but if, after investigation, they are found to be connected to system issues, then they will be re-assigned to the Service Provider or to Cray. The following table shows how many queries were opened and closed each month. The peak in October was related to requests for Technical Assessments of proposals submitted to the *Software Development for HPC* call.

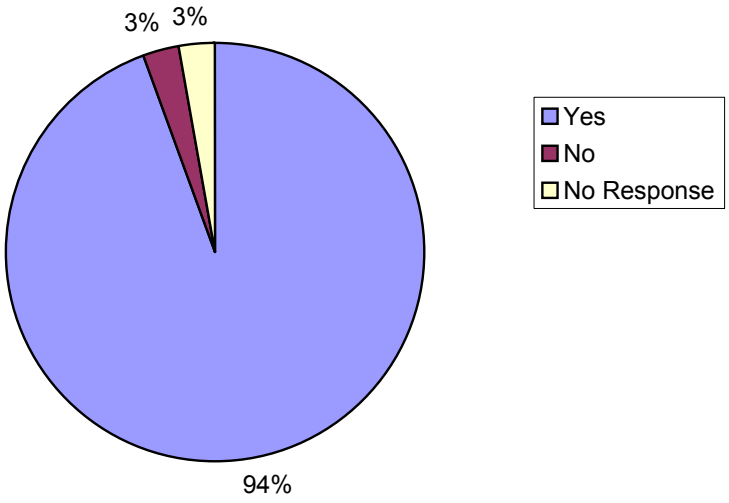


In cases where a technical query (as opposed to a request for a Technical Assessment) is resolved by the team (strictly speaking where the query is closed within the SAFE system by the CSE team), the user is invited to fill in a questionnaire giving feedback about his or her experience and satisfaction with the outcome. Since the start of the service 179 questionnaires have been sent out and 73 returned, a response rate of 41% which is quite high for a survey of this kind. The responses to the questions are summarised as follows:

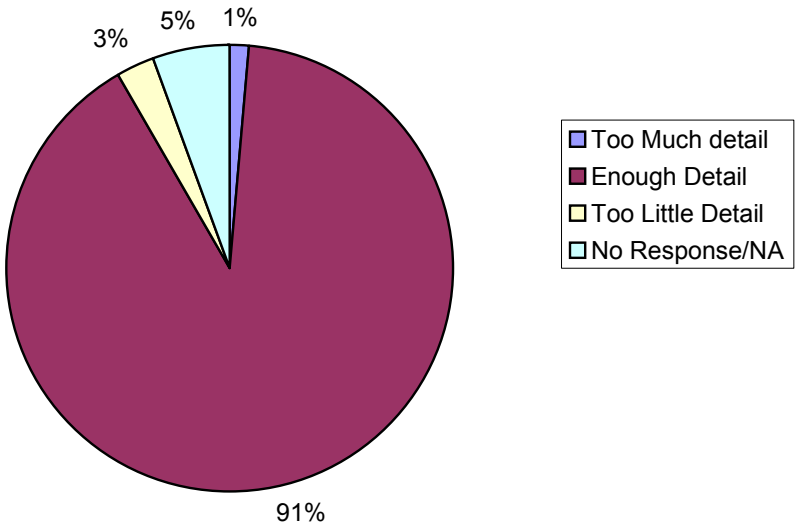
Has the problem raised in your query been resolved by the information provided by the helpdesk?

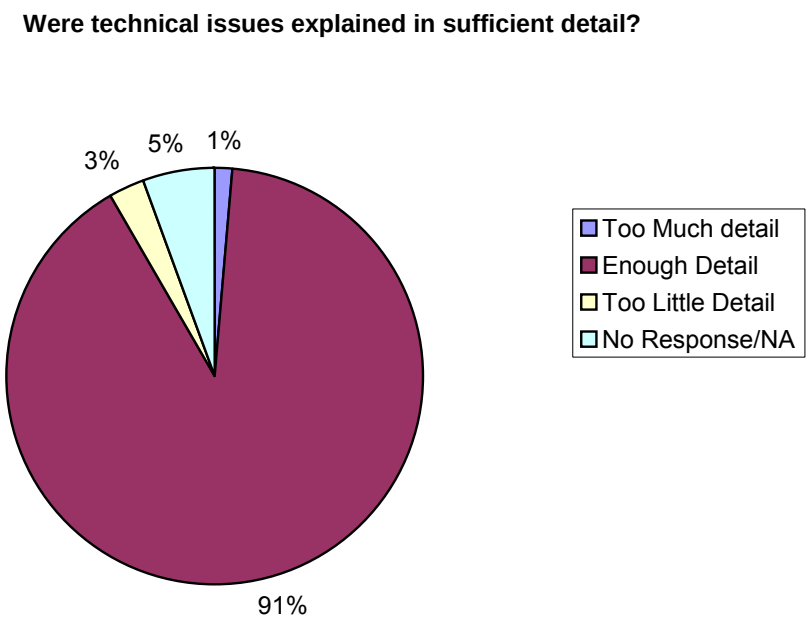


Were all communications replied to promptly and the text clear and understandable?



Were technical issues explained in sufficient detail?





On the relatively rare occasions that a negative response is received in feedback the CSE team will attempt to understand the reasons behind the response and, if necessary, change or improve their procedures.

7.3 Training

During this period NAG has offered a range of training course, varying from HECToR-specific courses (e.g. *Programming Tricks for HECToR),* through more general HPC courses (e.g. *Parallel Programming with MPI*), to courses teaching good practice in software engineering (e.g. *Programming for Portability and Maintainability*). Most courses have been run by NAG staff, although we have run two courses in collaboration with the University of Warwick (*Introduction to HPC*).

Take-up for courses has been mixed, with the HECToR-specific and basic HPC programming courses being the most popular. The majority of courses were initially delivered in Oxford, but during this year courses were also delivered in Warwick, Manchester, Bath and Birmingham. The latter two courses covered MPI programming with an introduction to HECToR, and were run in collaboration with local HPC services supporting mid-range clusters. These courses were very popular (the Birmingham one was heavily over-subscribed) and we plan to continue to run courses which can also be of benefit to users of local services (who we hope will go on to use the national services) in the future.

The complete list of courses offered in 2008 is as follows. The courses range in duration from one day to a week, and complementary courses are grouped together for convenience. Many of these courses were run several times in different locations.

- *Introduction to HECToR*
- *Tools and Techniques for Optimising Parallel Codes*
- *Testing and Benchmarking*
- *Scientific Visualisation*
- *Pitfalls of Numerical Software Engineering*
- *Techniques for Achieving Scalability*

- *Introduction to High Performance Computing*
- *Programming Tricks for HECToR*
- *Programming for Portability and Maintainability*
- *Programming the X2 Vector System*
- *Code Surgery*
- *Data Management*
- *Introduction to MPI*
- *Advanced MPI*
- *One-sided MPI Communication and MPI-IO*
- *OpenMP and Mixed-Mode Programming*

7.4 Outreach

The CSE team has attended a number of national and international meetings to promote HECToR and to share experiences with other HPC users and support services.

In the second half of 2008 a concerted effort was made to contact the Principal Investigator of every consortium on HECToR and arrange a visit by a member of the CSE service. The aim was to talk to the consortia to ensure that they were making the best use of the services on offer. To date most of the planned visits have taken place and, based on the feedback received, a new training course on FORTRAN has been introduced and some of the support material on the HECToR website has been revised. A number of Distributed CSE applications have also been submitted as a result of these visits.

Towards the end of 2008 the CSE service began a programme of visits to other HPC services worldwide that have a similar remit to HECToR, to engage in knowledge transfer activities and to ensure that the CSE service is promoting best practice. So far we have visited San Diego Supercomputing Center, Argonne National laboratory, Pittsburgh Supercomputer Center, Penn State University, and Ohio Supercomputer Center.

In September the CSE service organised the second HECToR User Group meeting in Edinburgh, attended by 46 users and representatives of the service providers and funding agencies. There were presentations from capability challenge users, Distributed CSE projects and X2 early users. The day ended with a panel session where the audience was able to question representatives from EPCC, Cray, EPSRC and NAG. The discussion was lively at times but on the whole constructive and positive, and EPCC and NAG took actions to address some of the issues that were raised. Notes of the panel session, along with copies of the presentations given on the day, are available from the HECToR website.

7.5 Other Core CSE Activities

The CSE team launched an electronic newsletter for HECToR users in November, designed to complement the regular mailings from the Service Provider. This will be sent out approximately every month and will contain information about software issues, workarounds, coding tips etc. It will also carry news of forthcoming events (e.g. workshops, training, proposal deadlines) and completed Distributed CSE projects.

The CSE team worked with the Cray Centre of Excellence to help the early adopters of the X2 vector system port their codes to the new hardware. Some impressive performance results were obtained but unfortunately the limited size of the system makes it unsuitable for many real applications.

7.6 Distributed CSE

The aim of the Distributed CSE (DCSE) programme is to provide researchers with resources to enable them to:

- port their codes onto HECToR, in particular to work with new codes or to enable previously unsupported features in existing codes;
- improve the performance of their codes on HECToR;
- re-factor their codes to improve long-term maintainability;
- take advantage of algorithmic improvements in the field of high-performance computing.

DCSE is designed to complement the usual instruments for funding research which often cannot be used to support software development and maintenance of this kind. The necessary staff can be provided by NAG, recruited specially, or seconded from an existing project at the host institution. Where necessary the CSE service will provide special training to equip the DCSE team member with the skills necessary to perform his or her project.

Applications for DCSE support are considered by an independent review panel, chaired by Professor Chris Bischof of the University of Aachen. The panel recommends which applications should be funded wholly or partially, and NAG then works to agree a contract with the applicant which covers the mechanism for resourcing the project, timescales and a programme of work. The panel met four times between the start of the service and the end of 2008, and considered a total of twenty applications. Of those, nine have been recommended for full funding and six for partial funding, two have been referred to a future panel, and three have been rejected. Twelve contracts have so far been agreed, which together will deliver more than 11 years of effort to the community.

The first DCSE contract to finish was *Enhanced parallel hierarchy for the materials modelling code CASTEP for the UKCP* in July 2008. The Principal Investigator was Keith Refson of Rutherford Appleton Laboratory, with the work being carried out by Phil Hasnip of the University of York. The aim of this project was to improve the scaling of CASTEP on 1024 (or more) cores, where it was operating at only 12% efficiency. Improving the parallelisation strategies used and parallelising some serial parts of the code increased that efficiency to 42%.

The other contracts which have been agreed and which are either underway or about to start are:

- *OCEANS 2025 (NEMO)* Dr Andrew Coward (University of Southampton)
- *Parallel Algorithms for Efficient Massively-parallel tools for the Study of Catalytic Chemistry* Dr Paul Sherwood (Daresbury Laboratories) & Professor Richard Catlow (UCL)
- *Improving the parallelisation and adding functionality to the quantum Monte Carlo code CASINO* Professor Dario Alfe (UCL)
- *Parallel Algorithms for the Materials Modelling code CRYSTAL* Professor Nic Harrison (Daresbury Laboratories)
- *Future-proof parallelism for the electron-atom scattering codes PRMAT* Dr Martin Plummer (Daresbury laboratories)
- *Cloud and Aerosol Research on Massively-Parallel Architectures (CARMA)* Dr Paul Connolly (University of Manchester)
- *Porting and Optimisation of Code_Saturne on HECToR and Black Widow* Professor David Emerson (Daresbury Laboratories)
- *Improve scalability of Domain Decomposition within CP2K* Dr Ben Slater (UCL)
- *WRF code optimisation for Meso-scale Process Studies (WOMPS)* Dr Alan Gadian (University of Leeds)

- *Support for multigrid improvements to Citcom* Dr Jeroen van Hunen (University of Durham)
- *Hybrid time-dependent density functional theory in the Castep code* Dr Keith Refson (Rutherford Appleton Laboratory)
- *Performance enhancements for the GLOMAP aerosol model* Dr Graham Mann (University of Leeds)

7.7 Plans for 2009

7.7.1 Training

A number of changes have been made to the existing courses, based on feedback from attendees. In 2009 a new three-day course on *Parallel Programming with MPI* will be offered, combining elements of the existing introductory and advanced courses. A new course on *Fortran-95*, aimed particularly at new postgraduates and post-doctoral researchers whose previous programming experience has been in high-level languages and environments such as Java or MATLAB, will be introduced. Such a course has been requested by several HECToR users.

Courses will continue to be run at NAG's Oxford and Manchester offices, but an increasing number will be run at other locations. Plans already finalised for the first quarter of 2009 include running courses at Imperial College London and the Universities of Bristol and Southampton. The CSE team is always open to requests for specific training targeted at particular projects or communities.

7.7.2 Outreach

The CSE service will continue its programme of visits to HECToR users. To date we have focussed on Principal Investigators but some of these lead very large consortia and we would like to start visiting users in those consortia as well. The programme of establishing contacts with leading HPC services worldwide will continue, with particular emphasis on sites with Cray hardware or a significant programme providing in-depth support to users. We will also be organising another User Meeting.

7.7.3 Distributed CSE

Now that we have a critical mass of DCSE projects underway we will be organising workshops involving all the DCSE staff. We will use our visits to HECToR users to explore ways in which DCSE projects can help them to improve their code and hence their research output.

Appendix A: Terminology

TSL : Threshold Service Level

FSL : Full Service Level

SDT : Scheduled Down Time

UDT : Unscheduled Down Time

WCT : Wall Clock Time

MTBF : Mean Time Between Failures = 732/Number of Failures

SP : Service Provision

SP Serviceability% = 100*(WCT-SDT-UDT(SP))/(WCT-SDT)

Technology Reliability % = 100*(1-(UDT(Technology)/(WCT-SDT))

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 file systems 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: Projects on HECToR

Code	Title	Funding Body	Class	Area	PI	AUs allocated	AUs used	AUs left
EPSRC Projects								
c01	Support of EPSRC/STFC SLA	EPSRC	Class1	support	Dr Richard Blake	12,803,723	6,753,052	6,050,671
e01	UK Turbulence Consortium	EPSRC	Class1	Engineering	Dr Gary N Coleman	3,107,500	1,672,366	1,435,134
e05	Materials Chemistry HPC Consortium	EPSRC	Class1	Chemistry	Prof C Richard A Catlow	1,129,267,228	6,419,663	1,122,847,565
e10	GENIUS	EPSRC	Class1	Chemistry	Prof Peter Coveney	9,257,856	5,436,742	3,821,114
e100	Large scale MD and quantum embedding for biological systems	EPSRC	Class2	Materials	Prof Zheng X Guo	100,000	10	99,990
e101	Optimization of HPCx LES code	EPSRC	Class2	Engineering	Prof Michael Leschziner	100,000	0	100,000
e102	Numerical investigation of aerofoil noise	EPSRC	Class1	Engineering	Dr Richard D Sandberg	5,000,000	69,097	4,930,903
e103	Micromagnetic simulations on HPC architectures	EPSRC	Class2	Engineering	Dr Hans Fangohr	100,000	0	100,000
e104	Fluid-Mechanical Models applied to Heart Failure	EPSRC	Class1	Physics	Dr Nicolas Smiths	2,400,000	0	2,400,000
e105	Joint Euler/Lagrange Method for Multi-Scale Problems	EPSRC	Class1	Engineering	Dr Andreas M Kempf	1,300,000	0	1,300,000
e106	Numerical Simulation of Multiphase Flow: From Mesocales to	EPSRC	Class1	Engineering	Prof Kai Luo	3,650,000	0	3,650,000
e107	Parallel Brain Surgery Simulation	EPSRC	Class1	Life Sciences	Dr Stephane P. A. Bordas	6,000,000	1	5,999,999
e108	Unsteady Propeller Noise	EPSRC	Class2	Engineering	Dr Sergey Karabasov	100,000	0	100,000
e109	Nonlinear modelling of tokamak plasma eruptions	EPSRC	Class2	Physics	Prof Howard Wilson	100,000	0	100,000
e110	Computational Aeroacoustics Consortium	EPSRC	Class1	Engineering	Prof Paul Tucker	39,000,000	0	39,000,000

e111	The Modelling of New Catalysts for Fuel Cell Application	EPSRC	Class2	Physics	Prof Dario Alfe`	100,000	0	100,000
e24	DEISA	EPSRC	Class1	support	Mrs Alison Kennedy	21,902,294	8,336,942	13,565,352
e34	Hydrogen vacancy distribution in magnesium hydride	EPSRC	Class2	Chemistry	Prof Nora de Leeuw	100,000	14,902	85,098
e35	Non-adiabatic processes	EPSRC	Class1	Materials	Dr Tchavdar Todorov	3,000,000	29,939	2,970,061
e42	Computational Combustion for Engineering Applications	EPSRC	Class1	Engineering	Prof Kai Luo	32,000,000	4,137,540	27,862,460
e59	Turbulence in Breaking Gravity Waves	EPSRC	Class1	Engineering	Prof Ian P Castro	400,000	18,319	381,681
e63	UK Applied Aerodynamics Consortium 2	EPSRC	Class1	Engineering	Dr Nick Hills	13,500,000	3,647,748	9,852,252
e68	Hydrogenation Reactions at Metal Surfaces	EPSRC	Class1	Chemistry	Dr Angelos Michaelides	50,000,000	20,163,250	29,836,750
e69	Simulations of a Subsonic Cylindrical Cavity Flow	EPSRC	Class2	Engineering	Dr Aldo Rona	100,000	81,812	18,188
e70	Computation of Electron Transfer Properties	EPSRC	Class1	Chemistry	Dr Jochen Blumberger	960,000	132,753	827,247
e72	Ultrascalable Modelling of Materials	EPSRC	Class2	Materials	Dr Lee Margetts	8,622,547	8,459,577	162,970
e74	Quantum Monte Carlo Methods	EPSRC	Class1	Materials	Prof Dario Alfe`	30,008,735	32,477,320	-2,468,585
e75	Terascale DNS of Turbulence	EPSRC	Class1	Engineering	Prof Christos Vassilicos	27,760,000	24,890,344	2,869,656
e76	HELIUM Developments	EPSRC	Class1	Physics	Prof Ken Taylor	6,000,000	1,861,801	4,138,199
e78	Q-Espresso CP/PWSCF Codes on HECToR	EPSRC	Class2	Chemistry	Dr Antonio Tilocca	100,000	91,184	8,816
e79	SMEAGOL	EPSRC	Class1	Physics	Prof Colin Lambert	3,000,000	20	2,999,980
e81	e-Collision experiments using HPC	EPSRC	Class2	Physics	Prof NS Scott	200,000	100	199,900
e82	ONETEP: linear-scaling method on High Performance Computers	EPSRC	Class2	Materials	Dr Peter Haynes	100,000	90,727	9,273
e83	Ab initio study of high pressure disordered ice	EPSRC	Class2	Physics	Dr Simon P Bates	100,000	100,589	-589
e84	Vortical Mode Interactions	EPSRC	Class1	Engineering	Dr Tamer Zaki	3,200,000	16	3,199,984

e85	Study of Interacting Turbulent Flames	EPSRC	Class1	Engineering	Dr N Swaminathan	2,083,000	0	2,083,000
e86	Single molecule vibrational microscopy and spectroscopy	EPSRC	Class2	Materials	Prof Andrew Fisher	100,000	103,105	-3,105
e87	Model Parameters for Unsaturated Elasto-plastic Models	EPSRC	Class2	Engineering	Dr Charles Augarde	100,000	91,363	8,637
e89	Support for UK Car-Parrinello Consortium	EPSRC	Class1	Physics	Dr Matt Probert	140,000,000	17,936,482	122,063,518
e90	Network modelling of wireless cities	EPSRC	Class2	Engineering	Prof Jonathan M Pitts	100,000	286	99,714
e92	Dynamo Action In Compressible Convection	EPSRC	Class2	Physics	Mr Paul Bushby	75,000	50,856	24,144
e94	Porting the Linear Scaling DTF Code Conquest to HECToR	EPSRC	Class2	Physics	Dr David Bowler	100,000	5,570	94,430
e96	Materials Property Relationships	EPSRC	Class2	Materials	Dr Shoufeng Yang	100,000	0	100,000
e97	Discovery of innovative hydrogen storage materials	EPSRC	Class2	Chemistry	Prof Zheng X Guo	100,000	0	100,000
e98	Non-linear magnetohydrodynamic modelling of tokamak plasmas	EPSRC	Class2	Physics	Mr Ian T Chapman	100,000	25,680	74,320
e99	New Developments in Modelling Electron Energy Loss Spectroscopy	EPSRC	Class2	Materials	Mr Andrew J Scott	100,000	138,818	-38,818
u02	Materials simulation using AIMPRO	EPSRC	Early use	Materials	Dr Patrick R Briddon	4,000,000	3,080,443	919,557
u03	DNS of NACA-0012 aerofoil at Mach 0.4	EPSRC	Early use	Engineering	Dr Gary N Coleman	2,500,000	2,301,049	198,951
u10	Turbulent Plasma Transport in Tokamaks	EPSRC	Early use	Physics	Dr Colin M Roach	2,500,000	2,121,848	378,152
y08	Testing	EPSRC	Early use	support	Dr David Jenkins	1,000	0	1,000
NERC projects								
n01	Global Ocean Modelling Consortium	NERC	Class1	Environment	Dr Thomas Anderson	9,830,000	6,143,815	3,686,185
n02	NCAS (National Centre for Atmospheric Science)	NERC	Class1	Environment	Dr Lois Steenman-Clark	52,500,000	26,927,314	25,572,686
n03	Computational Mineral Physics Consortium	NERC	Class1	Environment	Prof John P Brodholt	72,779,000	50,739,382	22,039,618

n04	Shelf Seas Consortium	NERC	Class1	Environment	Dr Roger Proctor	8,250,000	1,296,940	6,953,060
BBSRC projects								
b01	Biomarkers for patient classification	BBSRC	Class2	Life Sciences	Prof. Peter Ghazal	100,000	1	99,999
b08	Int BioSim	BBSRC	Class1	Life Sciences	Mr Mark M Sansom	866,000	0	866,000
b09	Circadian Clock	BBSRC	Class1	Materials	Prof Andrew A Millar	2,000,000	0	2,000,000
External projects								
x01	HPC-Europa	External	Class1	External	Dr Judy Hardy	375,000	483,338	-108,338
T01	NIMES: New Improved Muds from Environmental Sources.	External	Class1	Environment	Dr Chris Greenwell	4,113,669	19,910	4,093,759
Director's Time								
d03	EUFORIA	Directors Time	Service	Physics	Mr Adrian Jackson	1,000,000	24,559	975,441
d04	MSc Projects	Directors Time	Service	External	Dr David Henty	93,000	27,444	65,556
d05	Icon-DT	Directors Time	Service	Engineering	Mr Paul Graham	250,000	68,990	181,010
d07	Thermal ellipsiods and proton transfer	Directors Time	Service	Chemistry	Dr Carole A Morrison	1,116,000	265,788	850,212
d08	Oncology	Directors Time	Service	Chemistry	Mr Florian Scharinger	35,000	1,264	33,736
y09	Director's Time	Directors Time	Service	External	Prof Arthur S Trew	846,708	82,538	764,170