

HPCx Service Annual Report for 2008

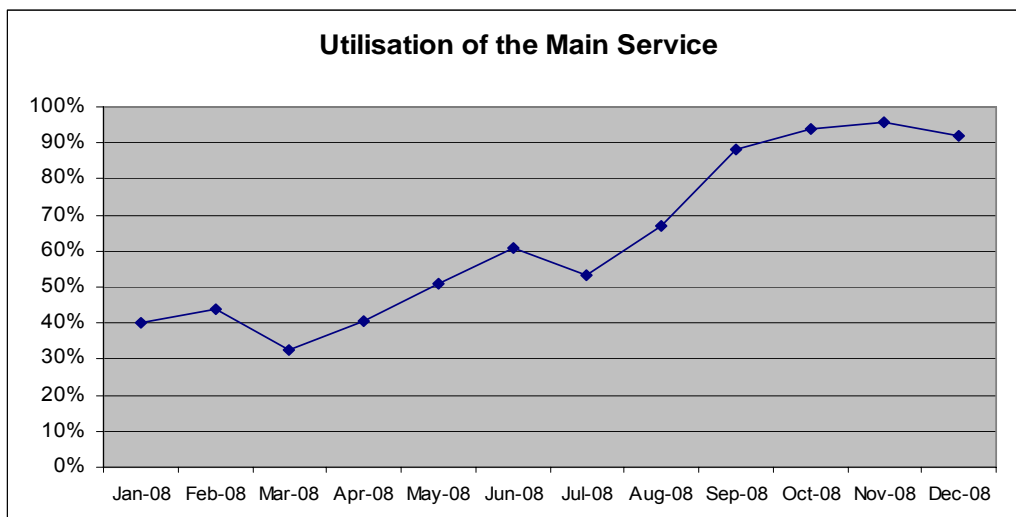
This report has three main sections: the first summarises our progress during the last year, including the key objectives agreed in the Annual Plan; the second reports quantitative information on the utilisation of the service and the agreed performance metrics; and the final section provides more details of the activities and current status of each of the support teams.

1 Summary of Progress

1.1 Overview

Undoubtedly the most notable success of 2008 was the remarkable reliability of HPCx. This is a great credit both to the Systems Team and to IBM.

- The HPCx Open Access initiative ran from August 2008 to January 2009 and encouraged users to run both longer and larger jobs on HPCx. This initiative had a significant effect on the overall utilisation as can be seen in the graph below. Average utilisation for months 1-6 was 45%, compared to 82% over months 7-12. Average utilisation for the year was 65%.



- HPCx entered Phase 4 as of 1 December. Unfortunately, both NERC and BBSRC ceased using HPCx at this stage. The service will now run until January 2010. The HPCx Phase 4 hardware upgrade has been delayed due to hardware availability; exact dates for the hardware upgrade are yet to be confirmed
- The volume of queries received at the Helpdesk was comparable to the previous year. The associated metrics were again excellent.
- HPCx runs alongside the HECToR service, and a major part of its role is to provide a complementary capability computing service. During 2008 a number of changes were implemented in support of this goal. These are detailed in Section 3.1.1 of this report.
- The results from the EPSRC Complementary Capability Computing Challenge were released in late December. Eight new projects have subsequently been set up on HPCx. The total allocation for these projects

amounts to 13.3M AUs. Consortia visits are being planned in direct support of these projects.

- HPCx continues to support and encourage novel computational research. We have continued investigations into the use of Object Oriented techniques for parallel computing. Also, Unified Parallel C (UPC) functionality was integrated into the 'Ludwig' Lattice Boltzmann Code.
- The BBC featured HPCx and the Computer Room in a series on climate change on Sunday the 21st of September at 9pm. The title was "Earth: The Climate Wars", and it was the third and final episode in the series.

1.2 Key Objectives

The Annual Plan for 2008 identified the following specific objectives and targets and this section indicates the progress we have made.

1) *HECToR came into service in 2007, offering a significantly increased computational capability over the HPCx service. However, there are several differences between the systems. HPCx can provide solutions for problems which, for architectural or service related reasons, HECToR is unlikely to address. We will, in 2008, work to exploit these strengths in the HPCx architecture and thus provide a complementary capability service. We will:*

- *Perform an evaluation of parallel visualisation utilities and subsequently enable interactive visualisation on the service, to the benefit of current projects.*

Parallel visualisation was evaluated and enabled on HPCx, allowing for visualisations which are too demanding for workstations. Following the evaluation of several tools, Paraview was determined to be the most appropriate. Several complexities occurring due to the nature of HPCx were resolved to enable interactive and batch mode use, allowing a wide range of visualisations to be performed. This work was documented in the technical report [HPCxTR0803](#) "Parallel Visualisation on HPCx".

This tool has been exploited by HPCx consortia including the "RealityGrid" and "Edinburgh Soft Matter" projects.

- *Exploit the good I/O performance of the system by investigating how the use of parallel I/O will improve the performance of applications.*

Investigations into parallel I/O on HPCx led to an understanding of the most efficient data access patterns, and this was exploited in the adaptation of a real molecular dynamics code resulting in significantly improved performance. This work was reported in [HPCxTR0806](#) "DL_POLY3 Parallel I/O Alternatives at Large Processor Counts". There were some limitations to the approach

taken in terms of memory usage and further investigations have identified a more memory-efficient approach taking advantage of more advanced features of the MPI-IO standard. This further work is planned to be disseminated in an appropriate journal or conference next year.

- *Through investigating the performance of shared memory programming techniques within a node, facilitate effective use of the large memory address space available on HPCx.*

The use of “Shared Memory Segment” and “OpenMP” shared memory programming techniques were investigated. A Shared Memory Segment utility library was created to allow users to safely use this. Also, the performance of OpenMP on HPCx was compared to other modern systems.

- *Exploit the job reservation scheme, currently being put in place to support Grid computing, to explore debugging on large numbers of processors.*

Job reservations are now available to all users on HPCx via the normal LoadLeveler reservation mechanism in order to support interactive use on large processor counts, such as pre-arranged debugging or computational steering sessions. We assisted the "Quantum Simulations for Chemical Biology" project to utilise this feature to perform computational steering of a Molecular Dynamics application.

- 2) *We will, while focusing on the usability for real HPC applications, study the performance of at least two emerging architectures. We will continue to evaluate the performance of key representative applications on HPCx compared to current competing and complementary systems (e.g. HECToR, Blue Gene and commodity clusters). This will principally take place through interactions with other projects, e.g. DEISA, PRACE and the TWP ACE exercise.*

HPC programming on Graphics Processing Units (GPUs) was investigated, and a benchmark resembling several types of applications (e.g. those in the area of computational fluid dynamics) was adapted to utilise the GPU architecture; this resulted in a speed-up of up to a factor of 7.5 over the use of a traditional CPU. This work was disseminated in the HPCx technical report [HPCxTR0808](#) “Utilisation of the GPU Architecture for HPC”. Applications were ported to the X2 Vector architecture, and benchmarking was performed resulting in documentation regarding the suitability of porting HPCx applications to this architecture for several application areas. A detailed performance comparison of HPCx verses the HECToR XT4 machine, considering both applications and lower level benchmarks, was performed and documented in the HPCx technical report [HPCxTR0801](#) . Several Fusion applications, currently active on HPCx, were ported to the Blue Gene/P architecture. Previously reported work on the IBM Cell architecture was furthered with a journal submission. The above work was all performed in a complementary fashion to activities undertaken in other projects, in particular DEISA, PRACE and the TWP ACE exercise.

- 3) *Grid access to HPCx continues to grow in its importance. This community is particularly supportive of advance reservations and we are working towards implementing this early in 2008. We will also support at least one significant technology demonstrator using HPCx and the grid.*

Advance reservations are now available via the HARC advance reservation mechanism. HARC is a grid co-scheduler that permits simultaneous reservations to be made across multiple machines to permit multi-site computing. Users make their requests to one of a set of broker daemons that contact resource manager daemons on each of the target sites and attempt to make corresponding local reservations. If all these reservations are successful then the corresponding reservation IDs are returned to the user. If any fail then HARC removes those reservations that did succeed.

In recent years HPCx has taken part in significant technology demonstrations by the GENIUS project as part of the annual Supercomputing conference. This year this project refocused its efforts on developing its production infrastructure rather than attempting another one-off large scale demonstration so they only ran a series of small scale demonstrations at Supercomputing that only required smaller systems than HPCx.

The FireGrid project is utilising the integration between Globus and advanced reservations, which we provided as part of this activity, to support pre-booked demonstrations of their software.

- 4) *We will run two workshops designed to inform the UK community of current developments in HPC software and to better understand the important issues in these areas from a user perspective. The workshops will cover: parallel I/O on large systems, and new HPC languages.*

The *HPCx Parallel I/O Workshop* was held at Daresbury Laboratory in December with a programme of five talks from HPC users with challenging I/O problems, and HPC vendors developing practical I/O solutions. It attracted more than 30 delegates.

The *Novel Parallel Programming Languages for HPC* workshop was held in Edinburgh in June, immediately after the HPCx Annual Seminar. It consisted of talks from HPC vendors, HPC language developers and HPC researchers, with the aim of bringing these three communities together to share their experiences. One of the topics delivered at the workshop is documented in the technical report [HPCxTR0804](#) "*Parallel OO Techniques for HPC Applications*". The workshop was extremely successful, attracting more than 60 delegates and ending with a lively open panel session. The workshop was very timely, with all present being invited to formally comment on proposed HPC extensions to the official Fortran language standard.

Together with the Annual Seminar, our events this year attracted more than 160 attendees.

- 5) *HPC is an increasingly pan-European activity. As part of our collaboration on the continuing eDEISA EU project, we will work towards producing a European HPC benchmark suite. This will involve porting and profiling key scientific codes on HPCx and other DEISA platforms, with a focus on very large numbers of processors. We will also work with the PRACE project investigating hardware and software technologies for future HPC systems.*

The first version of the benchmark suite was delivered as planned under the eDEISA project, and maintenance continued under the follow-on DEISA2 project. There have been strong links with PRACE which has produced EU-wide surveys of HPC usage, and built on the DEISA benchmark to start to develop a benchmark suitable for Petascale systems.

- 6) *The established programme of technical reports and course delivery will have the following targets for this year:*
 - a) *Eight technical reports with two produced per quarter;*
 - b) *A total of twenty days of training with runs at four different locations.*
 - c) *Production of material covering the new OpenMP 3.0 standard with an associated course run if feasible.*

During this year, we also plan to investigate the impact of these activities during the HPCx service.

We produced technical reports covering all 8 topics listed in the 2008 Annual Plan, at a rate of two per quarter. These are all available online at <http://www.hpcx.ac.uk/research/hpc/>. In order to assess the impact of the technical reports produced during the HPCx service, a survey of downloads was conducted based on data from the HPCx web server covering the period April 2006 to June 2008 (see [HPCx2Q08.pdf](#) section 4.1.2). Care was taken to identify only accesses from real users, removing hits from automated robots and web crawlers. Over this period, there were a total of 36,335 downloads of PDF copies of the 65 reports at a steady rate of around 1,400 per month. More than two thirds of downloads were from four countries: the UK, China, the US and India. Access across technical reports was relatively flat with two thirds (comprising those reports with counts greater than 500) coming from 40% of the titles. The most popular report (more than 3000 accesses) is quite old – a report on eigensolver performance from 2003. However, even the two reports from Q1 of 2008 already had more than 100 accesses each, and a report on single node performance on HPCx from 2007 had in excess of 800 downloads in just over a year. This data shows that these reports are of interest to the local and global HPC community, and are valuable to the dissemination of the activities of the HPCx support teams.

We ran 20 days of training at 5 different locations this year, slightly exceeding the target. We also tried to match the training programme to user requirements. For example, November's *Parallel Programming* course in London was not a standard one, but had content tailored specifically to the requirements of the e05 consortium.

We continued to lead the development of the new OpenMP language standard through EPCC's membership of the OpenMP Architecture Review Board where we are represented by Mark Bull. Stable OpenMP 3.0 compilers were not available in time to run a course. However, an OpenMP resource page is now available at:

<http://www.hpcx.ac.uk/support/documentation/openmp3/>

which includes the presentation we developed and links to other resources.

- 7) *There is a wide range of new and novel languages and programming models being developed for large scale HPC resources. The experience gained in 2007 on these will now be applied to real application kernels.*

Following on from the 2007 evaluation of several emerging novel paradigms, Unified Parallel C, selected due to its relative maturity and portability, was integrated into the Ludwig application (which is currently active on HPCx), and the performance and productivity of the adapted code was compared to the equivalent code from the original message passing version. This work was fed back to the Ludwig developers, and disseminated in a report which will be submitted to an appropriate conference or journal.

Investigations into the use of novel HPC languages for real applications were documented in the technical report [HPCxTR0807](#) "*Is MPI-2 suitable for Quantum Chemistry?*"

- 8) *To ensure that the science support effort is dominated by experienced staff, we will deliver at least 80% of the FSL effort levels from the named staff. All workshops and events will have at least two such staff involved, usually including one team leader; this will ensure that the staff experience directly benefits the user community.*

Experienced staff delivered around 75% of the FSL effort levels. We ensured that HPCx was well represented by experienced staff at a variety of national and international conferences. Most of the core staff attended the 6th HPCx Annual Seminar. The Parallel I/O workshop was supported by 4 of the experienced staff. The novel programming languages workshop was supported by 3 team leaders and 3 additional experienced staff.

2 Quantitative Metrics

2.1 Reliability

The service has continued in 2008 with an exceptional level of stability. The following table shows the number of incidents in 2008 at the various severity levels.

Severity	Number
1	2
2-4	59
Total	61

Incidents at severity level 1, which corresponds to contractual failure, were again at an all time low. Note the 2 contractual failures in 2008 were both as a result of maintenance session overruns.

The following table shows the number of failures, the mean time between failures and the serviceability. These are then attributed to the site, to IBM and to external factors.

Attribution	N	MTBF	Serviceability
IBM	1.5	5856	99.96%
Site	0.5	17568	99.98%
External	0	∞	100.00%
Total	2	4392	99.96%

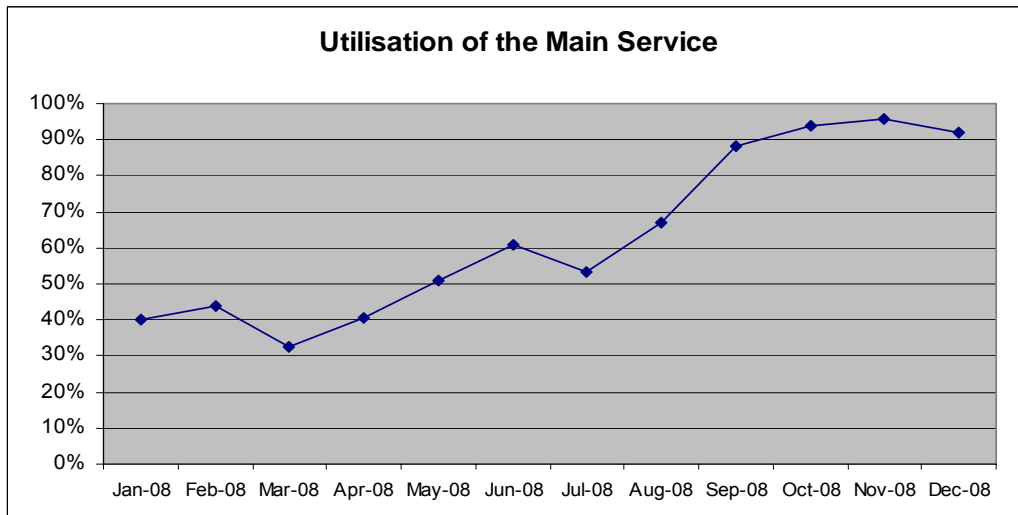
The MTBF and serviceability figures matched that of 2007.

2.2 Utilisation

2.2.1 Overall

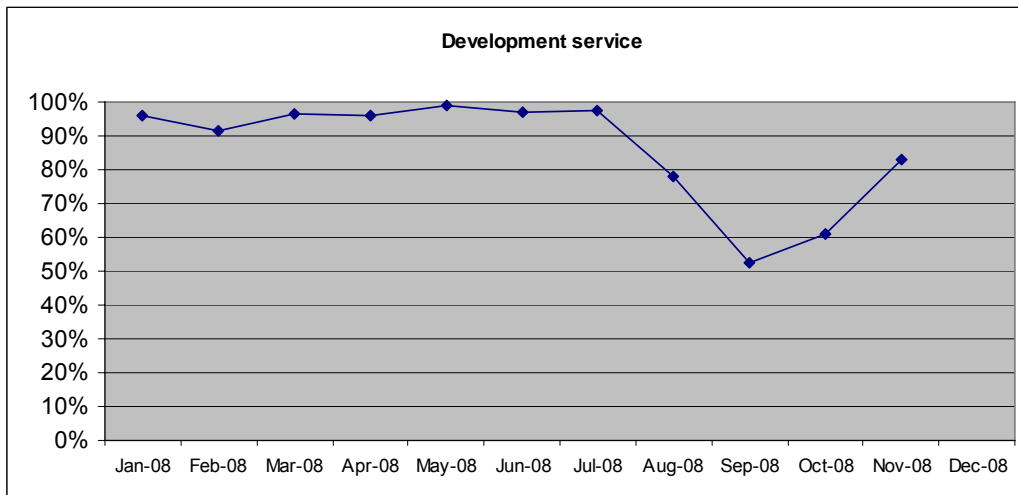
Main Service

The graph below shows the utilisation of the main service, as a percentage of the available time. Average utilisation for the year was 65% compared to 70% in 2007.



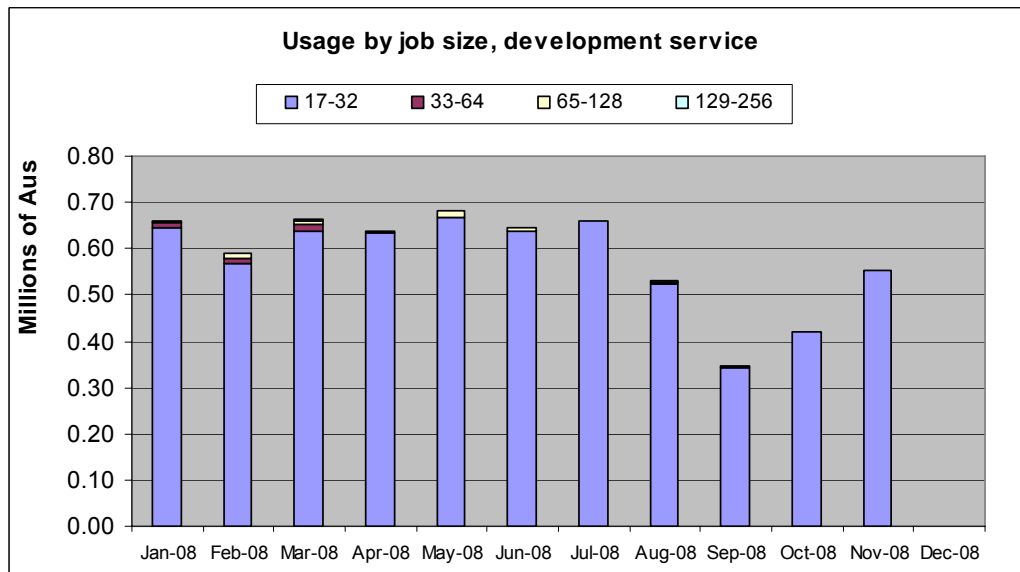
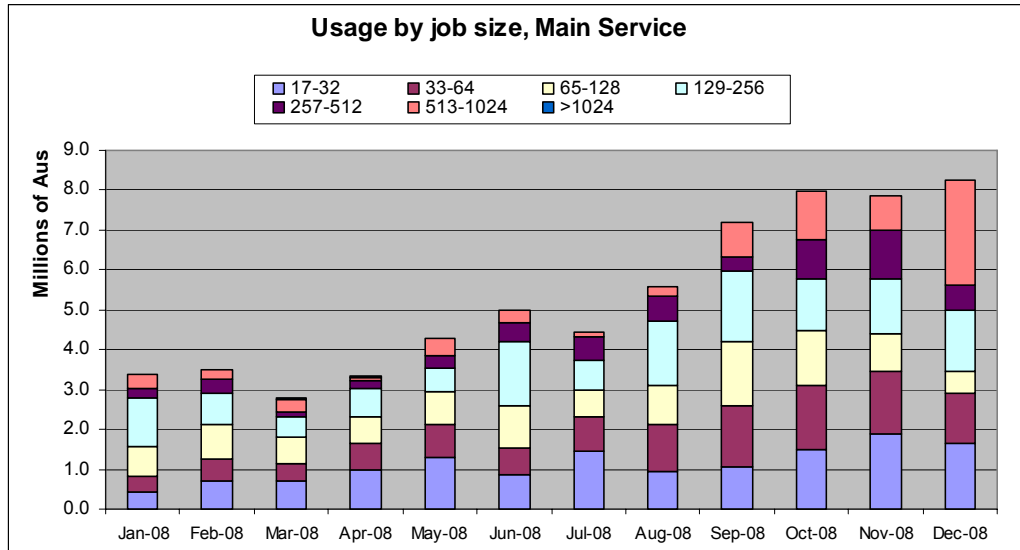
Development Service

The development service continued to be well utilised until such time that NERC left HPCx in December 2008



2.2.2 Usage by Job Size

The following graphs illustrate the way that the size of jobs has varied on the two services this year.



Generally, the average job size has increased throughout the year. In January, the modal job size was 129-256; in December 512-1024.

2.2.3 Usage by Consortium

A total of 98,363,352 AUs were available during this period.

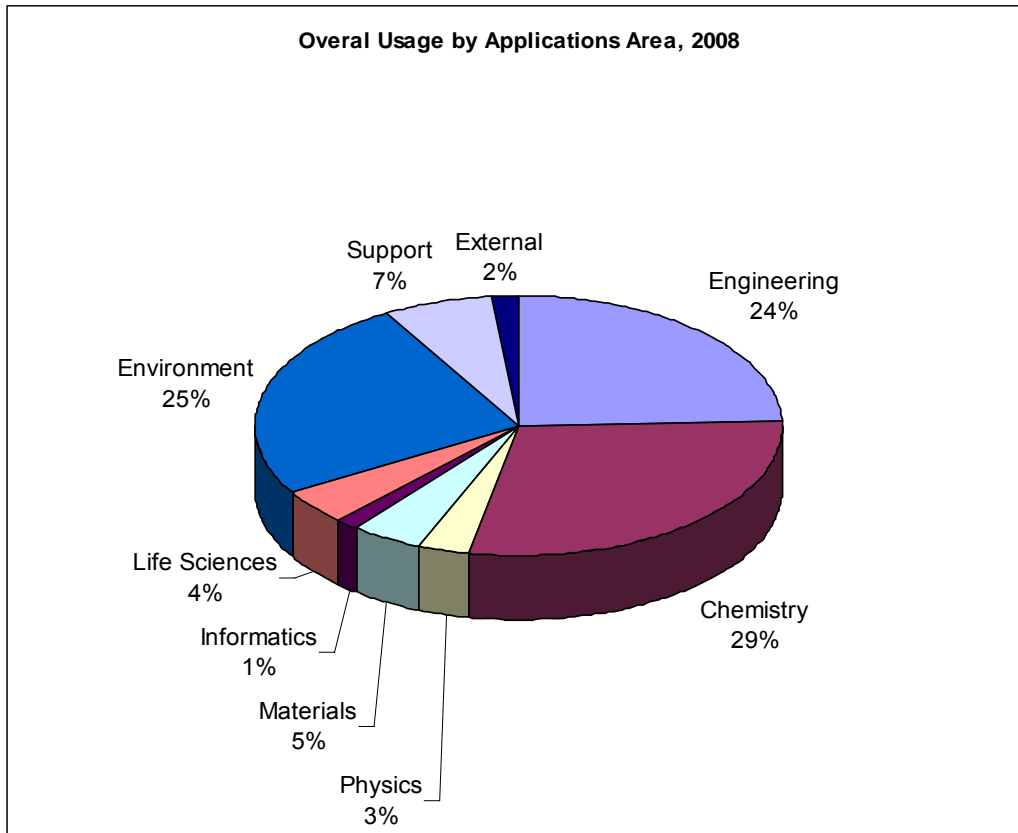
Project	Area	AUs	Number of Jobs	%age of Use	Utilisation
e01	Engineering	3,656,574	15103	11.76%	10.40%
e03	Physics	8,231	2582	0.03%	0.01%
e05	Chemistry	4,272,334	15753	13.74%	13.46%
e06	Materials	121,589	10	0.39%	0.12%
e08	Chemistry	863,527	2582	2.78%	0.88%
e10	Chemistry	70,150	237	0.23%	0.31%
e17	Life Sciences	1,235	3	0.00%	0.00%
e18	Engineering	28,292	275	0.09%	0.03%
e19	Materials	110,673	600	0.36%	0.77%
e20	Engineering	2,575	20	0.01%	0.00%
e24	Engineering	317,524	12847	1.02%	0.32%
e26	Life Sciences	162	18	0.00%	0.00%
e27	Chemistry	11,233	6	0.04%	0.01%
e30	Physics	0	557	0.00%	0.17%
e33	Engineering	1,091,612	1652	3.51%	2.03%
e35	Materials	537,362	4442	1.73%	1.12%
e36	Engineering	5,169	1044	0.02%	0.10%
e37	Engineering	0	14	0.00%	0.61%
e38	Environment	681,805	951	2.19%	0.69%
e39	Chemistry	473,289	1260	1.52%	0.66%
e40	Physics	0	77	0.00%	0.80%
e41	Engineering	835,631	187	2.69%	2.56%
e42	Engineering	7	100	0.00%	0.03%
e44	Engineering	91,767	246	0.30%	0.24%
e45	Materials	215,850	40	0.69%	0.22%
e46	Materials	1,814	25	0.01%	0.06%
e47	Engineering	260,101	199	0.84%	0.26%
e50	Life Sciences	216,061	831	0.70%	0.51%
e55	Engineering	3,112	10	0.01%	0.00%
e57	Life Sciences	256,711	640	0.83%	0.34%
e58	Chemistry	2,892,338	13991	9.30%	7.82%
e59	Engineering	14,343	213	0.05%	0.48%
e60	Physics	633,999	1142	2.04%	1.40%
e62	Physics	44,663	873	0.14%	0.25%
e63	Engineering	1,284,325	2634	4.13%	3.43%
e64	Engineering	0	351	0.00%	1.35%
e65	Engineering	6,207	2415	0.02%	0.04%
e67	Engineering	951	15	0.00%	0.00%
e68	Chemistry	172,403	1898	0.55%	0.89%
e93	Informatics	413,528	361	1.33%	0.42%
e94	Chemistry	112,290	48	0.36%	0.11%
e95	Engineering	0	15	0.00%	0.00%
EPSRC Total		19,709,435	86267	63.40%	52.92%

n01	Environment	51,050	59	0.16%	0.05%
n02	Environment	5,045,028	12205	16.23%	5.13%
n03	Materials	440,415	552	1.42%	0.45%
n04	Environment	2,111,523	1696	6.79%	2.52%
NERC Total		7,648,015	14512	24.60%	8.15%
b08	Life Sciences	903,445	167	2.91%	0.92%
b09	Life Sciences	1,893	13	0.01%	0.00%
BBSRC Total		905,338	180	2.91%	0.92%
p01	Physics	253,460	1946	0.82%	0.27%
PPARC Total		253,460	1946	0.82%	0.27%
x05	Engineering	572	89	0.00%	0.00%
z001	Support	637,235	7207	2.05%	0.65%
z002	Support	2	5	0.00%	0.00%
z004	Support	39,404	4550	0.13%	0.04%
z05	Support	5,679	118	0.02%	0.01%
z06	Support	854	98	0.00%	0.00%
z10	Support	49	26834	0.00%	0.00%
HPCx Total		683,795	38903	2.20%	0.70%
x01	External	535,896	1515	1.72%	0.54%
External Total		535,896	1515	1.72%	0.54%
c01	Support	1,350,095	7170	4.34%	1.39%
Support Total		1,350,095	7170	4.34%	1.39%
Overall Total		31,086,034	150493	100.00%	64.89%

As in 2007, the e01 and e05 consortia accounted for a significant amount of the utilisation of the service from an EPSRC perspective. In contrast to the previous year, e58 was also a major contributor to the overall utilisation.

From a NERC perspective, n02 was the main user group.

2.2.4 Usage by Applications Area



The overall utilisation in 2008 was slightly lower than in 2007. The proportion of overall utilisation by applications area in 2008 remained comparable to the previous year.

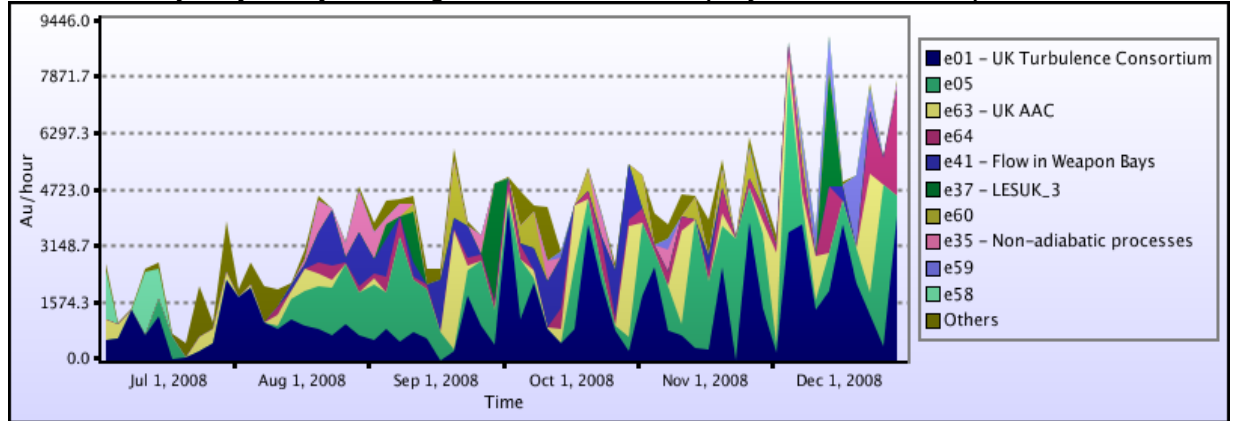
2.3 Open Access Initiative

In August 2008, the EPSRC Open Access Initiative was launched with the aim of encouraging users to explore complementarity. The section below illustrates the impact of this initiative on utilisation of HPCx. This initiative ran from August 2008 to January 2009.

2.3.1 Impact on Utilisation

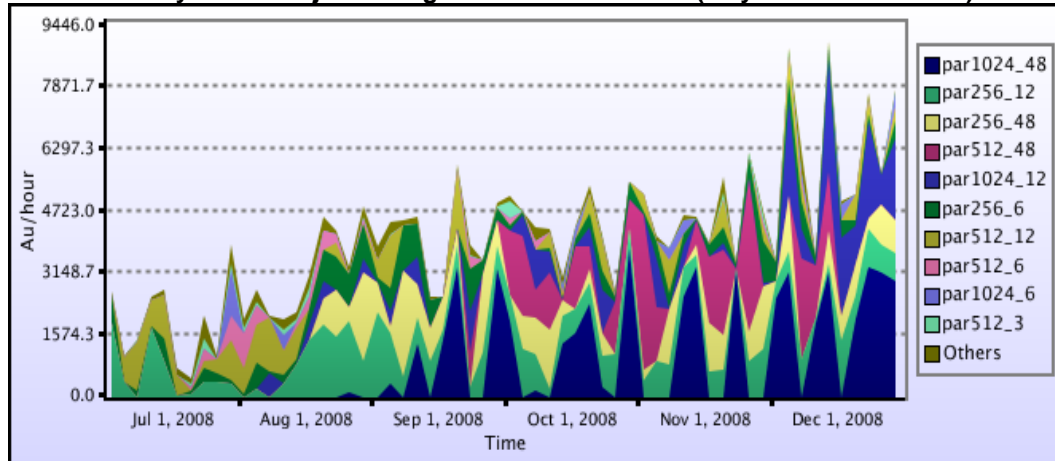
The graph below illustrates the utilisation by project of large jobs (256 processors +).

Raw AU use by Project of jobs using more than 255 CPUs (July – December 2008):



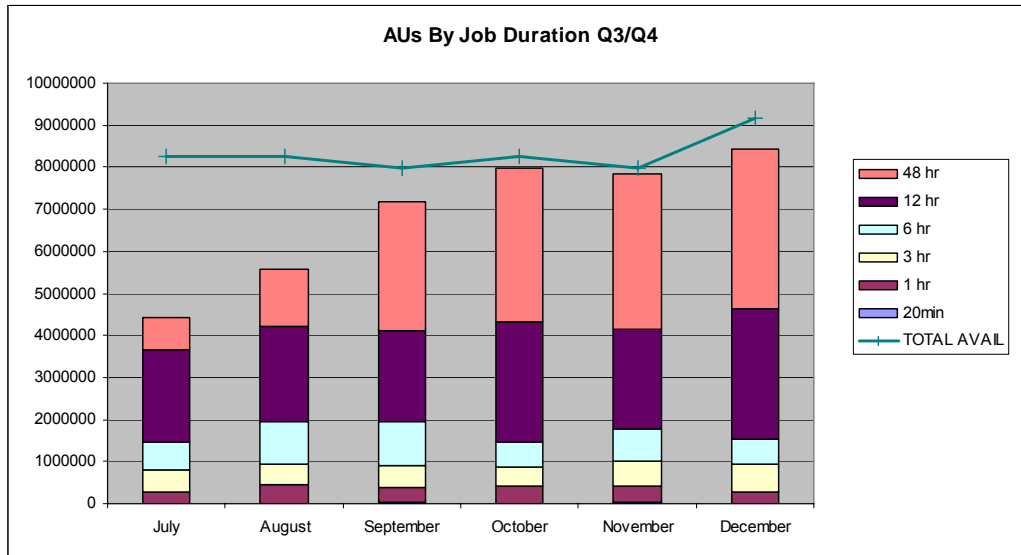
The graph below illustrates the breakdown by job size.

Raw AU use by Queue of jobs using more than 255 CPUs (July – December 2008):



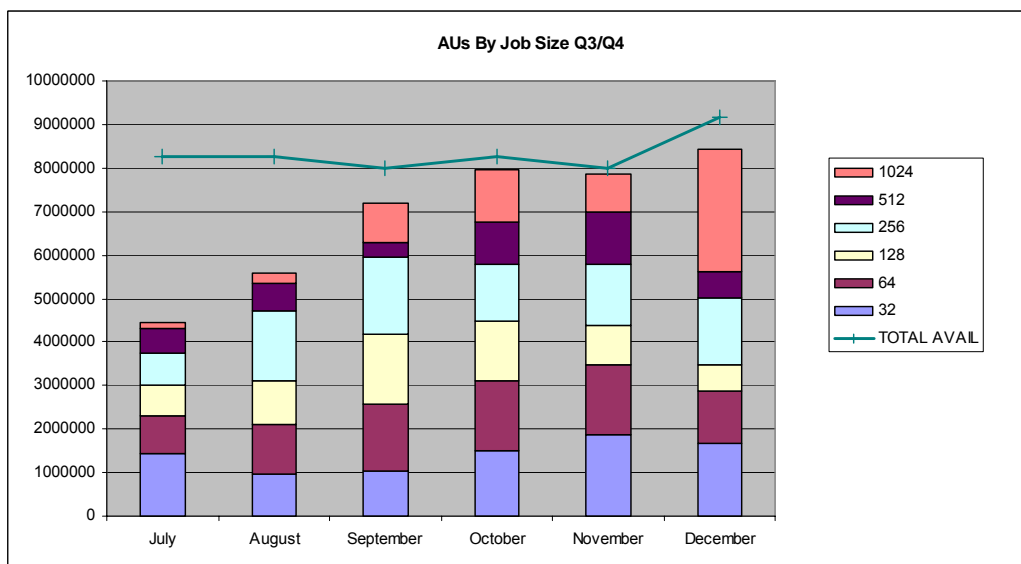
2.3.2 Complementarity

A further aim of the free time initiative was to promote new and innovative uses of HPCx.

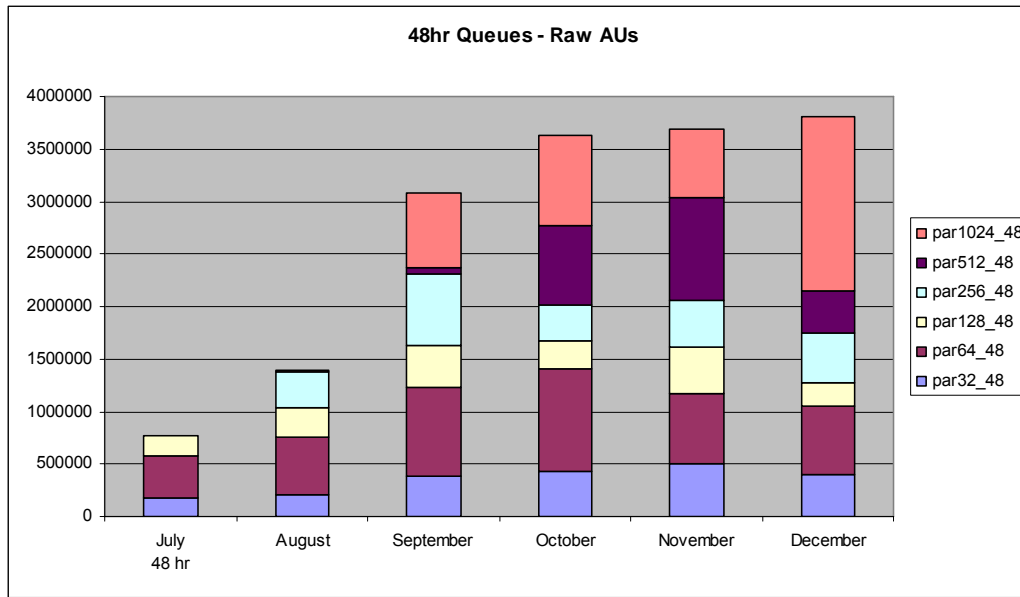


The graph above illustrates the continued change in user behaviour as a result of the free time. Users clearly took the opportunity to run longer jobs.

The graph below illustrates the change in user behaviour in relation to the size of jobs. The increase in capability jobs can clearly be seen.



Users were less conservative and grasped the opportunity to run longer and larger jobs. The 48 hour queues were utilized at all job sizes as can be seen below.



2.4 Performance Metrics 2008

Metric	TSL	FSL	January	February	March	April	May	June	July	August	September	October	November	December	Value for the Year
Technology serviceability (%)	80	99.2	100.0%	99.8%	100.0%	100.0%	100.0%	100.0%	99.8%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Technology MTBF (hours)	200	300	8	1464	8	8	8	8	732	8	8	8	8	8	5856
Number of AV FTEs	7.5	10	10.8	9.7	10.5	12.1	10.5	11.9	11.0	11.3	12.0	12.5	11.6	6.8 (Note 1)	10.3
Number of training days per month	22.5/12	30/12	4/1	4/2	10/3	10/4	10/5	13/6	14/7	14/8	14/9	15/10	20/11	21/12	21/12
Non in-depth queries resolved within 3 days (%)	85	97	100%	100%	98%	98%	100%	98%	99%	99%	98%	100%	98%	98%	98.8%
Number of A&M FTEs	3.75	5.75	5.9	5.3	5.4	5.9	5.8	5.8	5.9	5.5	5.6	5.8	5.8	4.8 (Note 2)	5.2
A&M serviceability (%)	80	99.6	100.0%	99.8%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Colour	Meaning
	Exceeds FSL
	Between TSL and FSL
	Below TSL

The number of training days is reported as a running total since the start of the year and the key metric is the annual target which was met. The months below FSL simply reflect the planned training timetable

Note 1: The Revised AV FTE TSL/FSL for Phase4 is 3.25/5.25

Note 2: The Revised A&M FTE TSL/FSL for Phase4 is 2.75/4.75