

The HECToR Service

An update

12 October 2010

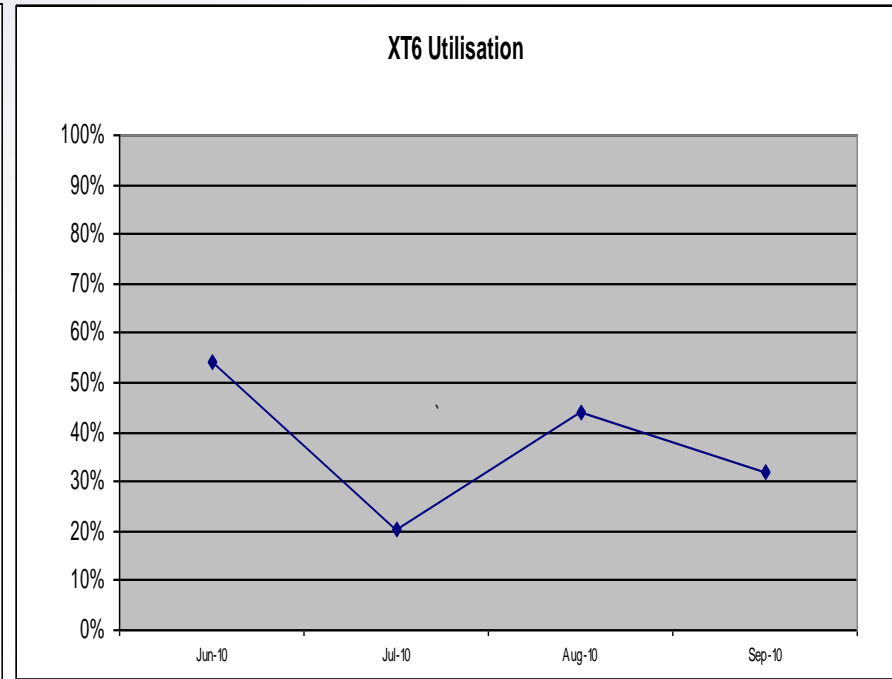
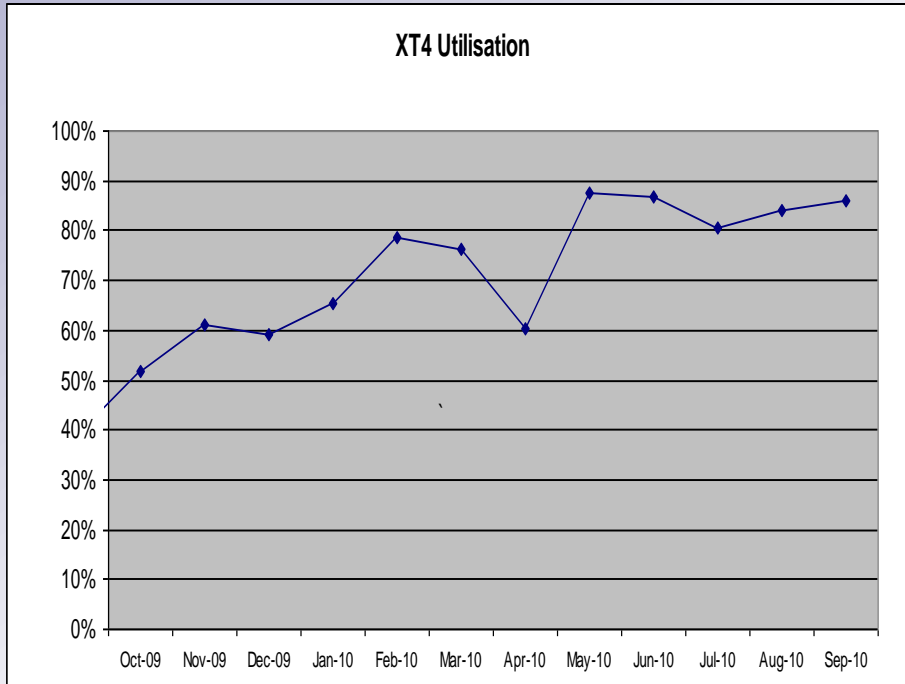
Liz Sim



Overview

- Utilisation
 - XT Utilisation
 - Capability Usage
 - Low Priority Access
 - X2 Utilisation
- Metrics
 - Reliability
 - Helpdesk
 - Power Efficiency
- Service Improvements
 - HECToR Roadmap
 - esFS Lustre
 - Phase 3
- Summary

XT Utilisation



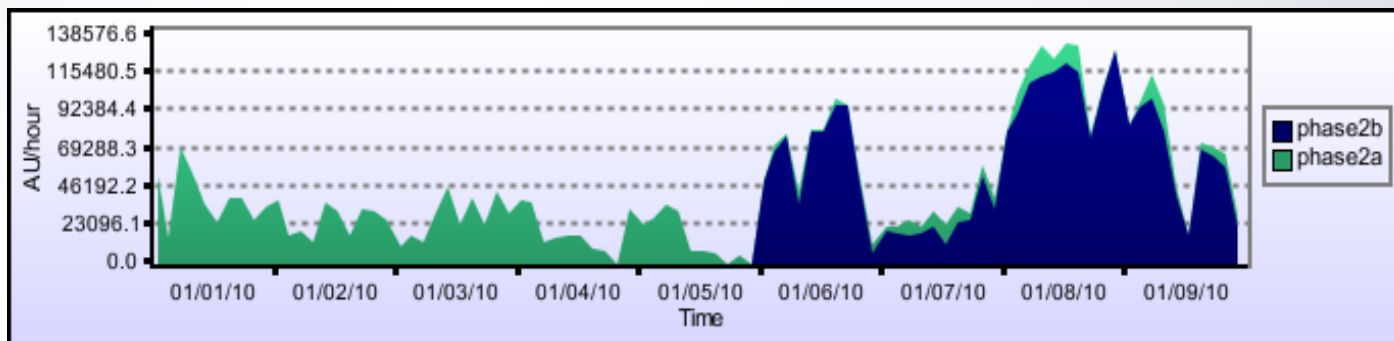
- XT4 utilisation has been very good
- But XT6 utilisation is not yet as high as everyone would like
 - Modified charging rate to encourage usage

Capability Usage

- Capability Incentive levels were revised during 3Q10
 - A single set of incentives now applies on both Phase 2a and Phase 2b

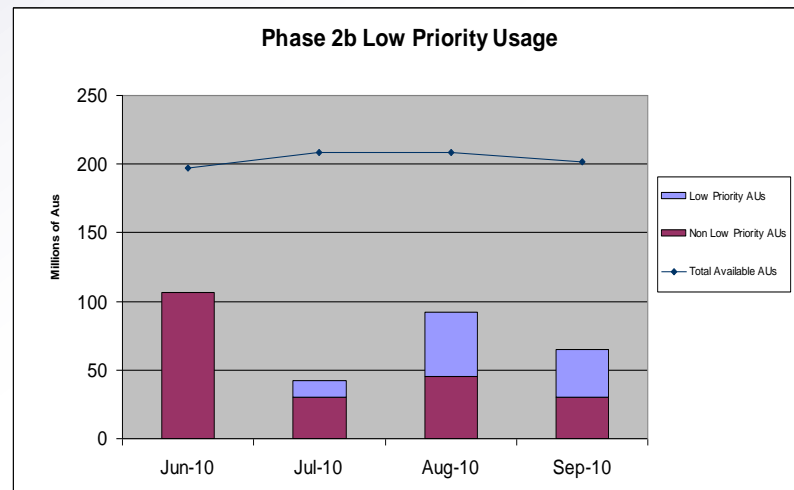
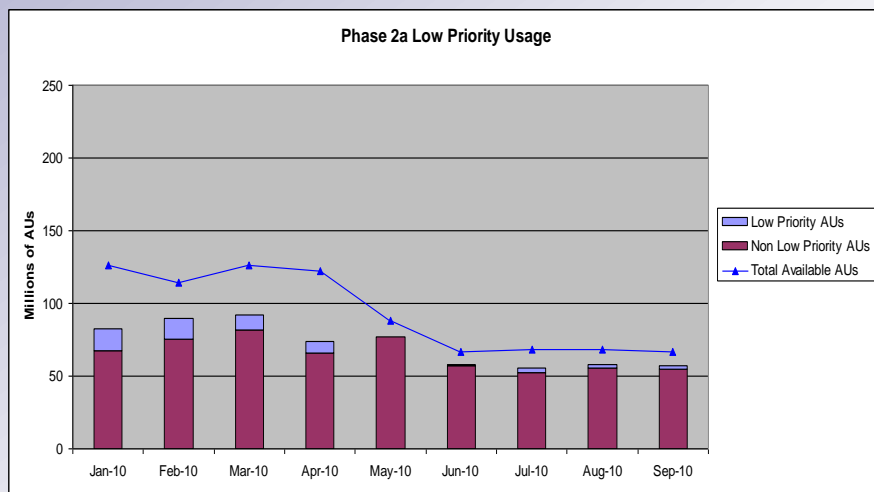
<i>Level</i>	<i>Minimum Number of Cores</i>	<i>Discount</i>
Bronze	2048	5%
Silver	4096	15%
Gold	8192	30%

- The introduction of the Phase 2b system has clearly increased the capability for running large jobs



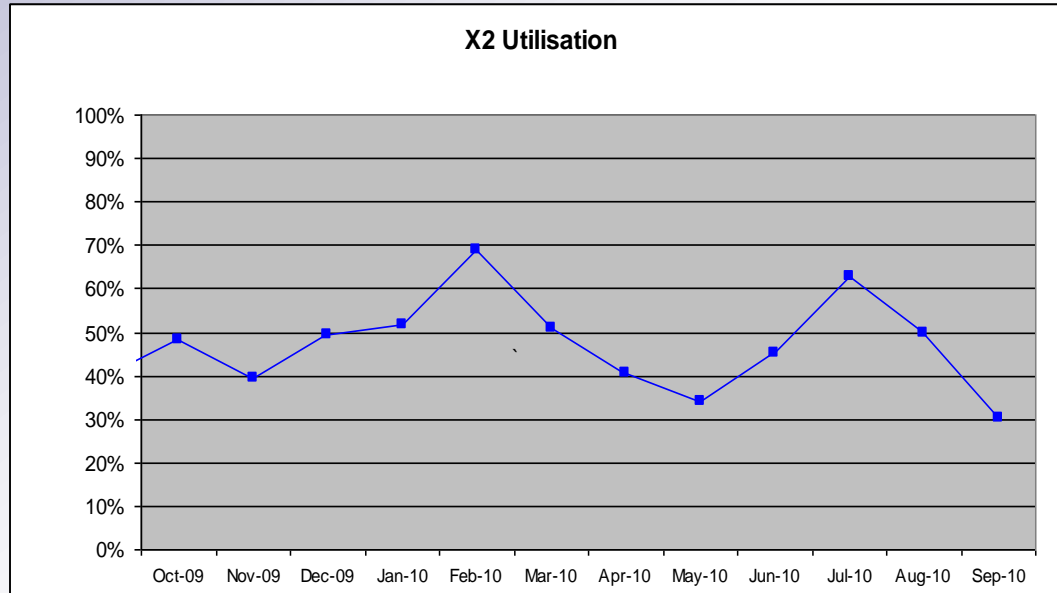
HECToR Jobs using greater than 2048 CPUs

Low Priority Access



- On Phase 2a, only a few low priority jobs have run since May
- On Phase 2b, during August and September there have been more AUs used by Low Priority Jobs than charged jobs

X2 Utilisation

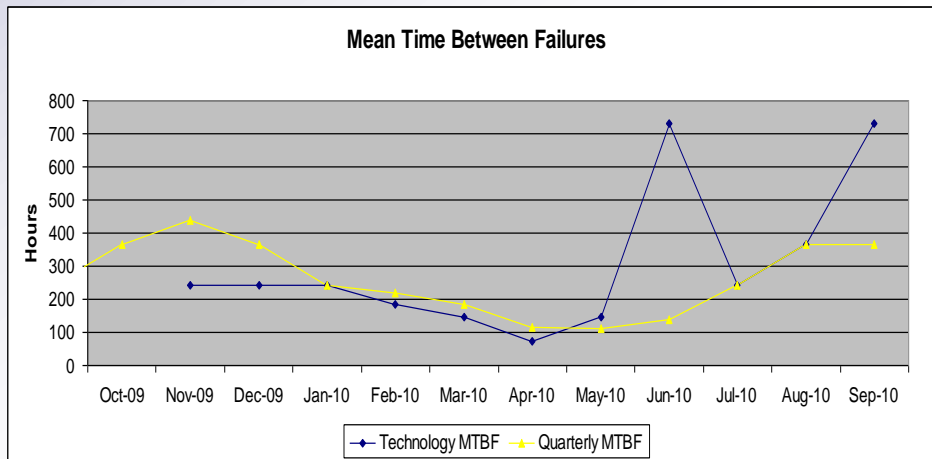


- Charging remains suspended on the X2
- Usage has averaged around 50%
- A very small subset of users use the X2
 - A single user accounts for 75% of the usage to date in 2010
- The requirement to maintain the X2 is being reviewed

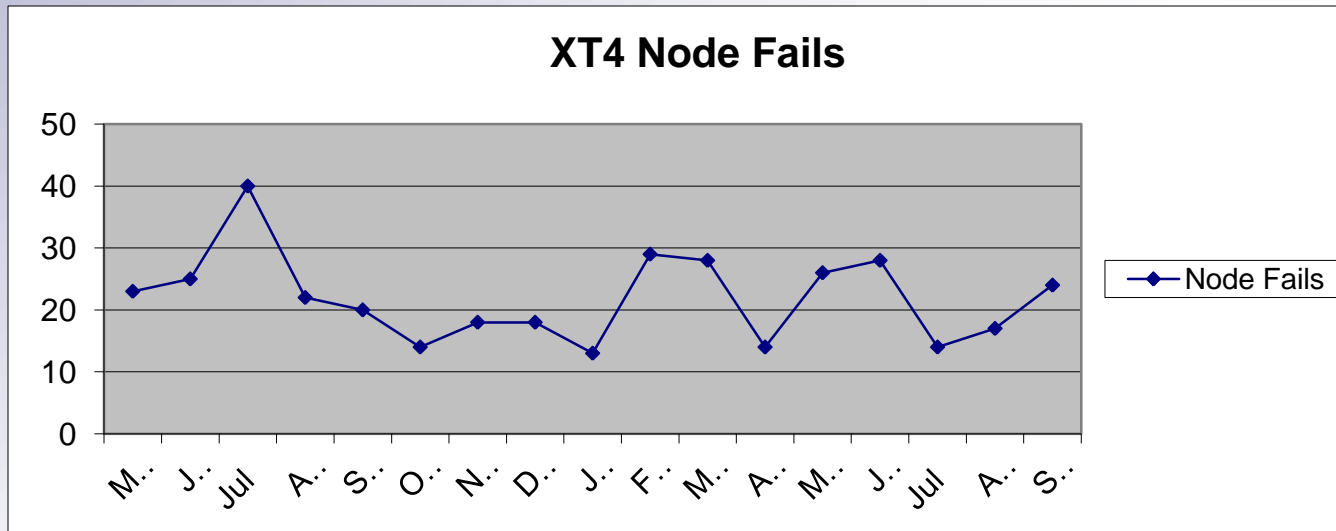
Overall Reliability (Phase 2a)

	2Q10			3Q10		
	Apr-10	May-10	Jun-10	Jul-10	Aug-10	Sep-10
Cray	10	5	1	3	2	1
Site	0	0	0	0	0	0
External	0	0	0	0	0	1

- Since the serious problems in April/May, reliability on the XT4 has been generally good



Single Node Failures



- Node failure rates as a result of memory and processor failures remain very low
- Most node failures on the XT4 relate to comms errors
- On the XT6 failure rates also low (but utilisation also low)
 - Majority of errors relate to 'Out of Memory' condition
 - Issue fixed in CLE3.1, which was installed 29th Sep

Helpdesk

Metric (1 Mar 10 - 1 Oct10)	Pass	Total	Fraction	Target
All queries finished in 1 day	1745	1766	98.8%	97.0%
Admin queries finished in 1 day	1575	1589	99.1%	97.0%
Queries assigned in 30 min	2055	2057	99.9%	97.0%
Technical assessments in 10 days	61	63	96.8%	97.0%

- Above numbers do not include Phase 2b specific queries
- Helpdesk has been busy, but metrics are excellent
- High number of Technical Assessments
 - Due to RAP and software development call

Performance Metrics: Summary

Metric	TSL(%)	FSL(%)	Apr-10	May-10	Jun-10	Jul-10	Aug-10	Sep-10
Technology reliability (%)	85.0%	98.5%	80.9%	98.6%	99.3%	99.3%	98.6%	99.9%
Technology MTBF (hours)	100	126.4	73.2	146.4	732.0	244.0	366.0	366.0
Technology Throughput, hours/year	7000	8367	6984	8393	8636	8601	8348	8307
Capability jobs completion rate	70%	90%	97.4%	94.7%	100.0%	97.7%	98.2%	96.7%
Non in-depth queries resolved within 1 day (%)	85%	97%	98.7%	98.5%	98.7%	100.0%	99.0%	97.7%
Number of SP FTEs	7.3	8.0	8.6	10.3	10.7	9.0	9.0	TBA
SP serviceability (%)	80.0%	99.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Colour Coding :

Exceeds FSL	
Between TSL and FSL	
Below TSL	

Definitions:

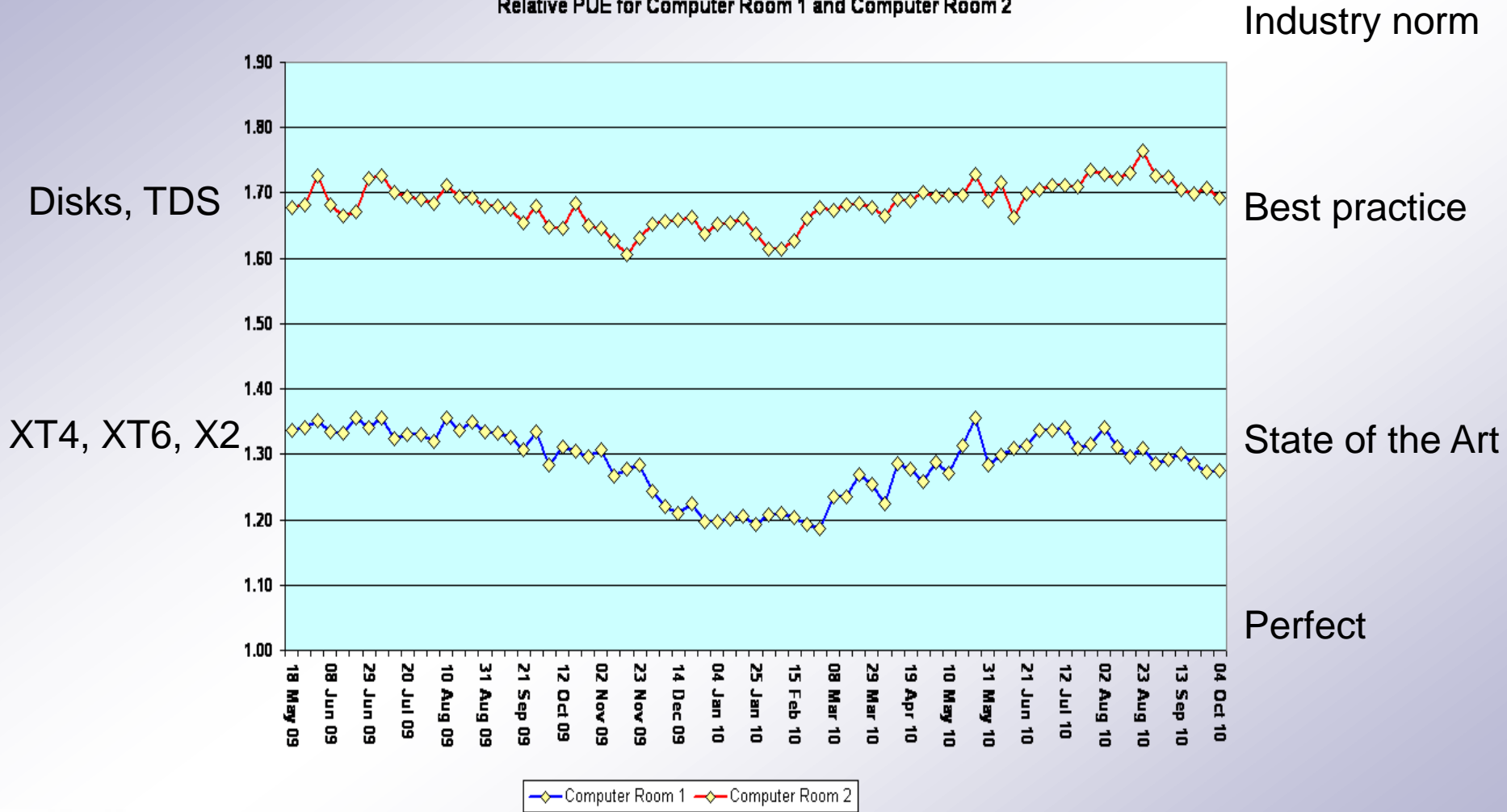
TSL	Threshold Service Level
FSL	Full Service Level
SDT	Scheduled Down Time
UDT	Unscheduled Down Time
WCT	Wall Clock Time
MTBF	Mean Time Between Failures = 732/Number of Failures
SP	Service Provision
SP Serviceability%	$100 * (WCT - SDT - UDT(SP)) / (WCT - SDT)$
Technology Reliability %	$100 * (1 - (UDT(Technology)) / (WCT - SDT))$

Process Improvements

- Reviewing options for opting in/out of email communications
- Option to 'opt-out' in SAFE from different levels of detail
- Default - Service Updates Only
 - e.g. Upgrade Plans, CSE Newsletters, Research Council Announcements
- Additional option - System Status Updates
 - e.g. System Available/Unavailable, Scheduled maintenance reminders
- Existing option in SAFE to opt-out of all emails will remain

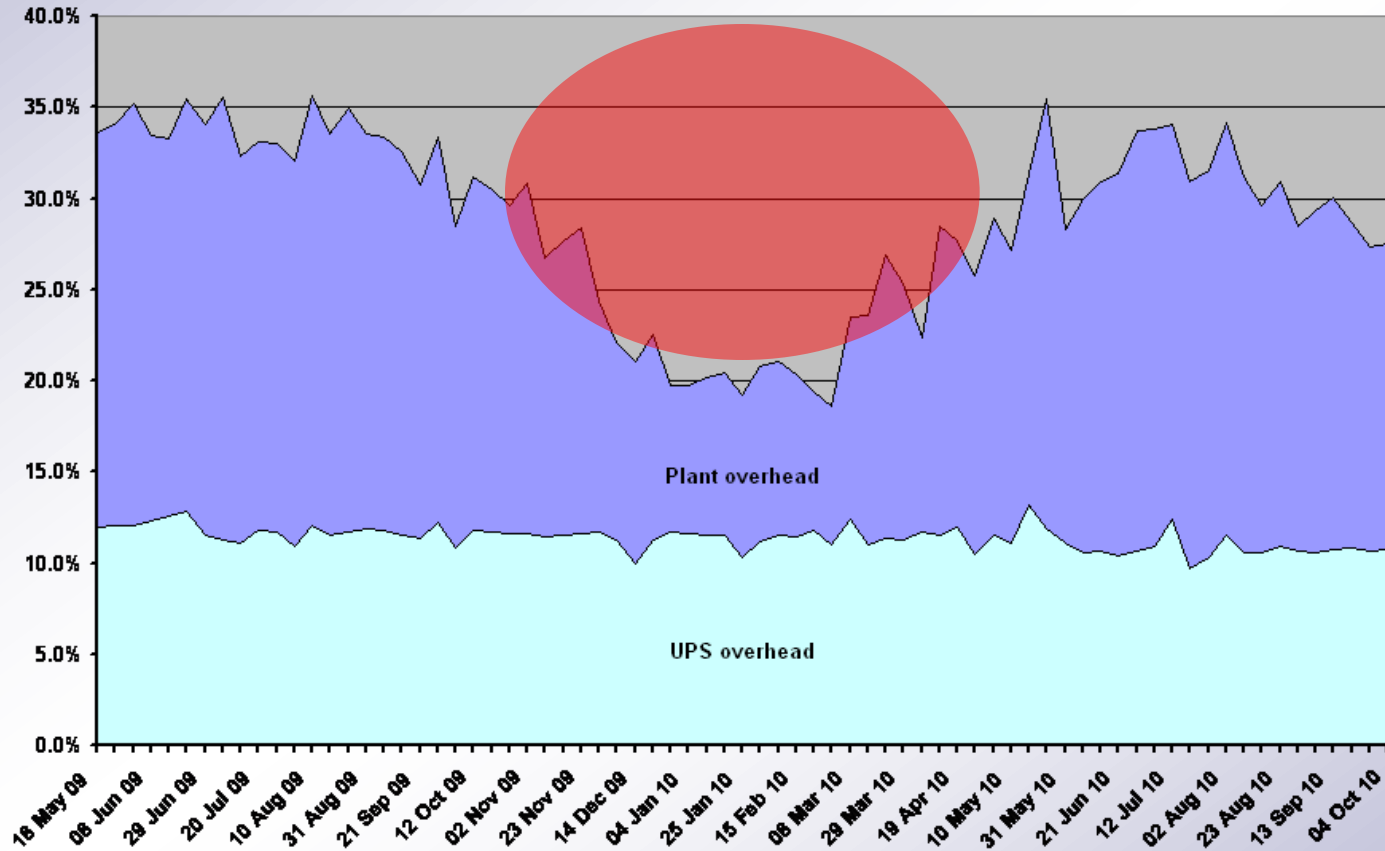
Power Efficiency

Relative PUE for Computer Room 1 and Computer Room 2



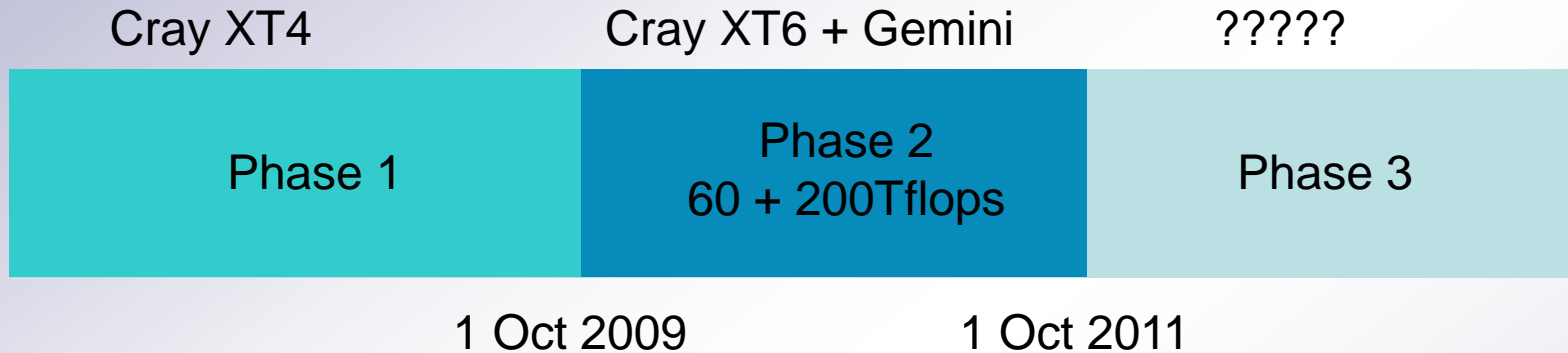
...Where does it go?

Free cooling operates 75% of the year but is most effective in winter

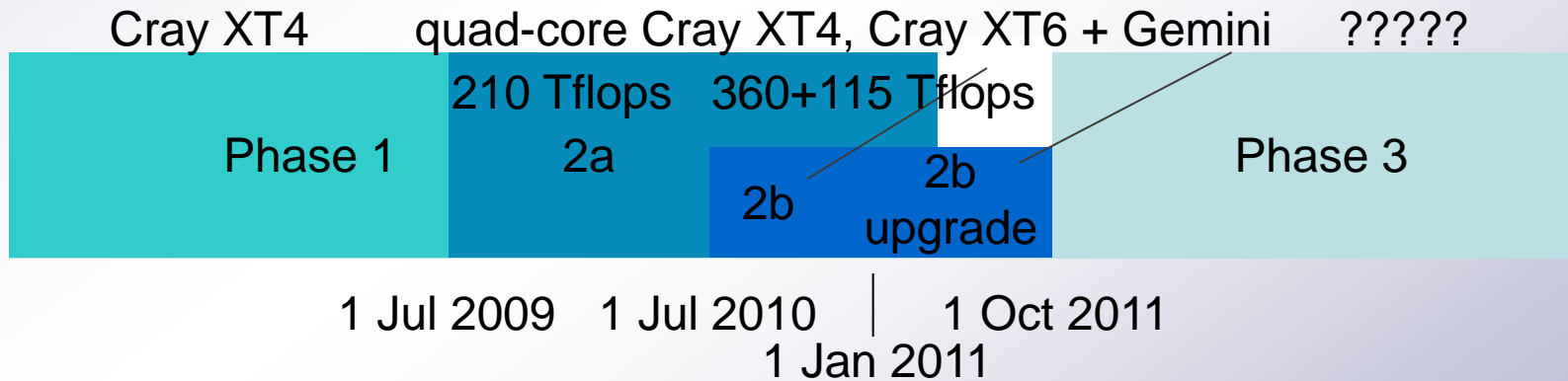


The HECToR Roadmap

- in the beginning:



- but then the new processor was early and Gemini was late:

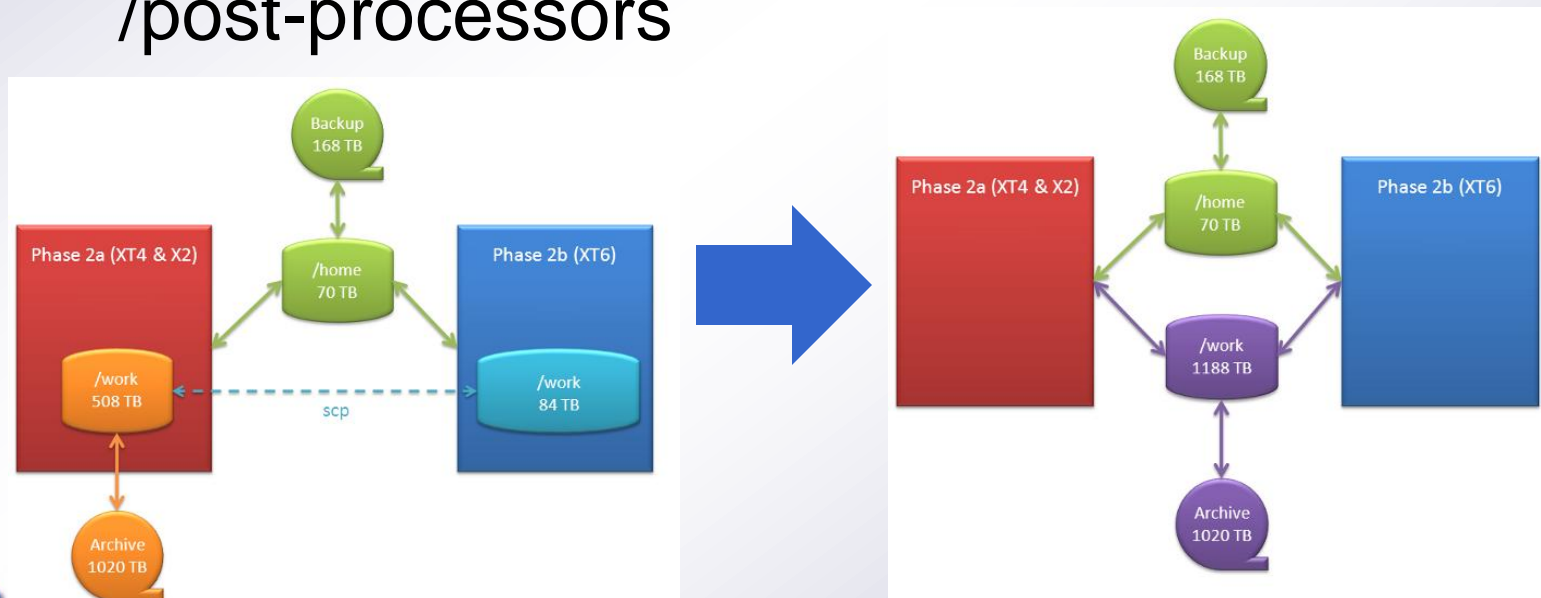


Phase II

- Phase IIa: Q2 2009
 - Dual core -> Quad core
 - 210 Tflops peak
 - memory grew from 6GB/node to 8GB/node
 - X2, interconnect, filesystem etc ... unaltered
- Phase IIb: Q2 2010
 - XT4 re-sized to 33 cabinets (115 Tflops peak)
 - 20 cabinet XT6 installed (360 Tflops peak)
- Phase IIb Upgrade: Q4 2010
 - network to be upgraded with Gemini

esFS Lustre

- in parallel we want to upgrade the filesystem
- ... to solve two problems:
 - seamless data transfers between Phases
 - supporting multiple hosts, eg. additional pre-/post-processors



esFS woes

- the plan had been to install esFS Lustre before Phase 2b
 - a tight timetable but possible, if things worked
 - ... they didn't.
 - problems forced us to reverse the order. Cray installed 170 TB new disk to compensate
 - unfortunately, the problems (thought solved) recurred in July ... finally fixed in August
 - new problems in September

esFS – the current plan

- we believe have a solution for known Lustre problems
- ... but want to check this on the XT4 before risking transfer of users' data
 - hence, 10-day test-period from 15 October
 - most users migrated during this period
 - large data users, customised migration thereafter
- Phase 2b Upgrade (Gemini) installation once esFS fixed

what will we do?

- *12.00 15 October*: shutdown the XT4 and XT6
 - transfer archiver to XT6
 - *15.00 15 October*: reboot XT6
- run standard Lustre & esFS reliability tests
- run “user-inspired” Lustre & esFS tests with advice from n02
- remount esFS on the XT6 and verify functionality
- migrate (most) users’ data to esFS
- *12.00 25 October*: return the XT4 to normal service
- start large data user migration in liaison with consortia

Phase III

- requires a contract extension for Cray, or new contract
- plan to minimize costs by requiring Phase III to fit within infrastructural limits
 - Max power: 1.8 MW
 - Max cooling: 1.8 MW
 - Size: 3000 sq ft

Cray Phase III proposals

- save cost by building on existing Phase 2b, Filesystem and archive
 - 1: Magny-Cours
 - 12-core, 2.1 GHz 1.33 GB/core
 - 72,192 cores - 606Tflops
 - 2: Interlagos
 - 16-core, 2.3 GHz, 1 GB/core
 - 65,280 cores - 601 Tflops
 - 83,712 cores – 770 Tflops
 - 96,000 cores – 883 Tflops
- Magny-Cours - more familiar to users, larger memory/core, higher running costs, no scalability
- Interlagos – some code changes may be required, reduced memory/core but faster

Phase III questions

- should we:
 - change the balance between data and compute?
 - currently the proposals only upgrade compute
 - increase GB/core (expensive)
 - ride the microprocessor curve for increased effectiveness or minimize change?
 - diversify architectures (e.g. add GPGPU-based system)?

Summary

- XT4 has been consistently busy in recent months
- ... and reliability is better
- However, XT6 utilisation remains low
- Service performance metrics are good
- esFS installation has been very problematic
 - 10-day shutdown to address issues
- Gemini is now likely to be post-Xmas
- Phase 3 is being planned/discussed