

Cluster Usage Report

Iain Rae <iainr@inf.ed.ac.uk

Available data.

Unfortunately this information was requested just as we started to upgrade the clusters to SL5, so we have no data on Lutzow. Much of the temporary infrastructure used to produce stats for the “clustering” meeting last year has rotted and died so we're limited to gridengine's basic accounting tools. We've also recently had a large number of problems with job databases being corrupted so we don't have a lot of archival data. I've included stats on townhill because, although it's funded by HCRC it's about to be tied together with hermes (and a lot of the names are the same).

The first set of tables show cluster usage over the last year for Lion,Hermes and Townhill. For brevity each table lists the total time used by the top 20 users, the full listing of these tables were used to identify users for the questionnaire.

Lion (top 20 users)

Total usage in units of hour

Last year user stats for queue all.q

```
wallclock : 206064.04
user time : 158369.71 [76.85%]
system time: 984.20 [0.48%]
cpu time : 166336.69 [80.72%]
```

By User usage sorted by wall

Username	%	wall	user	system	cpu	memory
pkoehn	20	41725	21606	482	23430	26082723
jjamagis	16	32904	28192	190	29090	3911600
s0129866	15	31249	23807	57	24678	892033
tjones1	8	16035	14932	8	15004	534558
s0451222	7	13960	12629	3	13117	12967941
s0234842	7	13605	12962	29	12992	2601881
s0235256	5	10816	8238	65	9315	2127376
s0235704	5	10195	9753	22	9775	311608
s0570628	5	9785	8488	23	8769	300572
s0565349	4	8084	7420	1	7618	5388880
s0565860	1	2804	1095	15	2225	1021845
v1brauch	1	2292	2238	3	2245	150456
v1cshi	1	2217	2173	1	2175	392087
s0344725	1	2016	1307	0	1977	107553
hleather	1	1939	135	8	147	66969
s0454958	1	1380	640	0	714	70446
s0672485	1	1265	478	23	533	43972
s0680896	1	1206	1003	9	1047	168680
s0451815	0	811	318	9	459	851057
s0681566	0	383	79	1	97	69483

Hermes (top 20 users)

Total usage in units of hour

Last year user stats for queue all.q

wallclock : 175643.93
user time : 141678.34 [80.66%]
system time: 1264.01 [0.72%]
cpu time : 153725.49 [87.52%]

By User usage sorted by wall

Username	%	wall	user	system	cpu	memory
s0129866	13	22222	17748	63	17947	1044389
jyamagis	11	19241	16942	76	18502	1402929
s0565741	10	17931	13354	262	15190	8468600
pkoehn	9	16050	9050	148	9263	6058544
s0235704	7	11977	10630	30	11096	1232563
s0235256	6	10099	8178	147	9777	671841
s0451222	5	9046	8356	40	8897	7072006
s0197686	5	8752	7595	174	7784	190644
s0460218	5	8626	7014	6	8479	195249
s9734229	5	8205	5424	68	6994	328597
joe	4	7230	6988	2	7176	201415
s0787953	2	3999	3487	28	3592	2129575
s0344725	2	3624	2808	1	3115	24023
s0681566	2	3571	2428	20	2525	3432139
tjones1	2	3212	2911	2	2956	688294
adielman	2	2952	2733	26	2792	890325
s0677383	1	2495	2040	19	2079	3787450
ggarau	1	2341	1927	50	2654	3409854
s0680896	1	2157	1817	4	2124	527035
v1jyamag	1	2039	1777	5	1843	434026
s0451815	1	1724	1678	5	1692	1617492

Townhill (top 20 users)

Total usage in units of hour

Last year user stats for queue all.q

```
wallclock : 456753.55
user time : 369083.26 [80.81%]
system time: 3344.07 [0.73%]
cpu time : 417127.64 [91.32%]
```

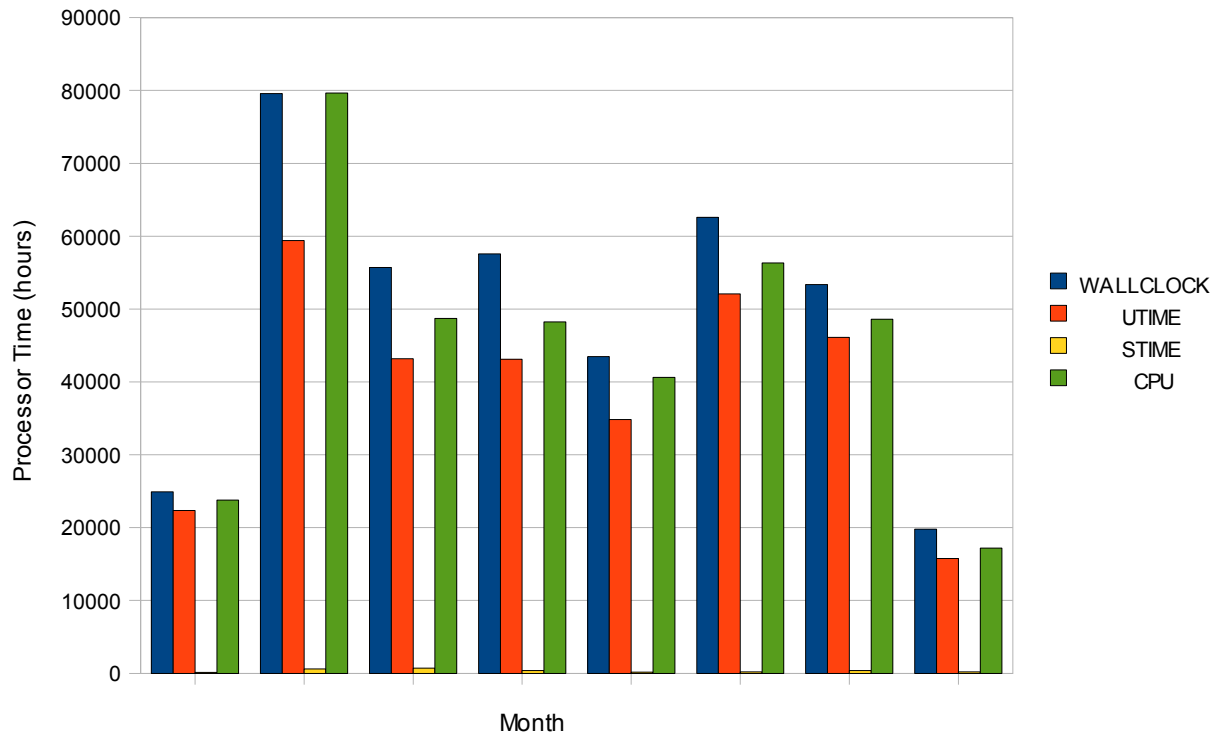
By User usage sorted by wall

Username	%	wall	user	system	cpu	memory	
jyamagis		16	71844	63357	174	65879	17114311
pkoehn		14	64081	45911	691	47466	49055221
v1jtejed		12	52963	42834	54	55915	4835936
s0565741		8	36858	31416	777	34326	35440994
s0129866		8	35866	32073	91	32396	1906061
s0570628		5	24591	22162	78	28975	2618387
dwang2		4	18006	15048	219	15677	18556493
s0197686		4	16108	10823	41	15996	452249
s0681566		3	15679	11312	60	12296	16721100
s0680896		3	13466	12661	26	12804	619146
joe		2	11269	9634	7	11088	418267
s0565860		2	10755	8987	20	9822	3039854
bhadow		2	10682	8824	136	9360	3932107
tcohn		2	9234	7647	7	8965	883925
s0787953		2	7794	6924	221	7308	4846608
s0677383		2	7163	2964	25	3694	5000553
anuthman		2	7046	758	1	6876	101218
s0235256		1	5727	5302	23	5561	487219
s0343799		1	4489	3745	10	4319	492925
jschroe1		1	4387	3598	19	3662	4075577

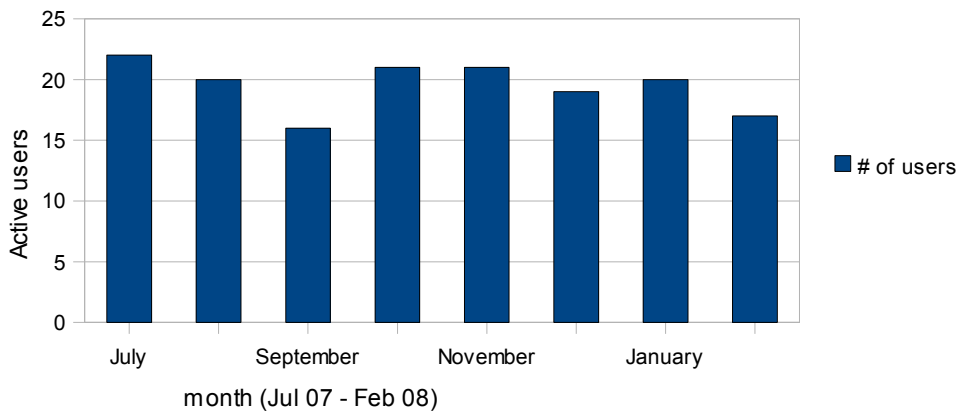
The next graphs show wall, user, system and CPU time totals for all job run per month for the last 12 months on Townhill and lion. These graphs should give an idea of how busy the clusters were. If people were switching to Eddie you'd expect to see a drop in throughput.

Processor time totals per month (townhill)

July 2007- February 2008

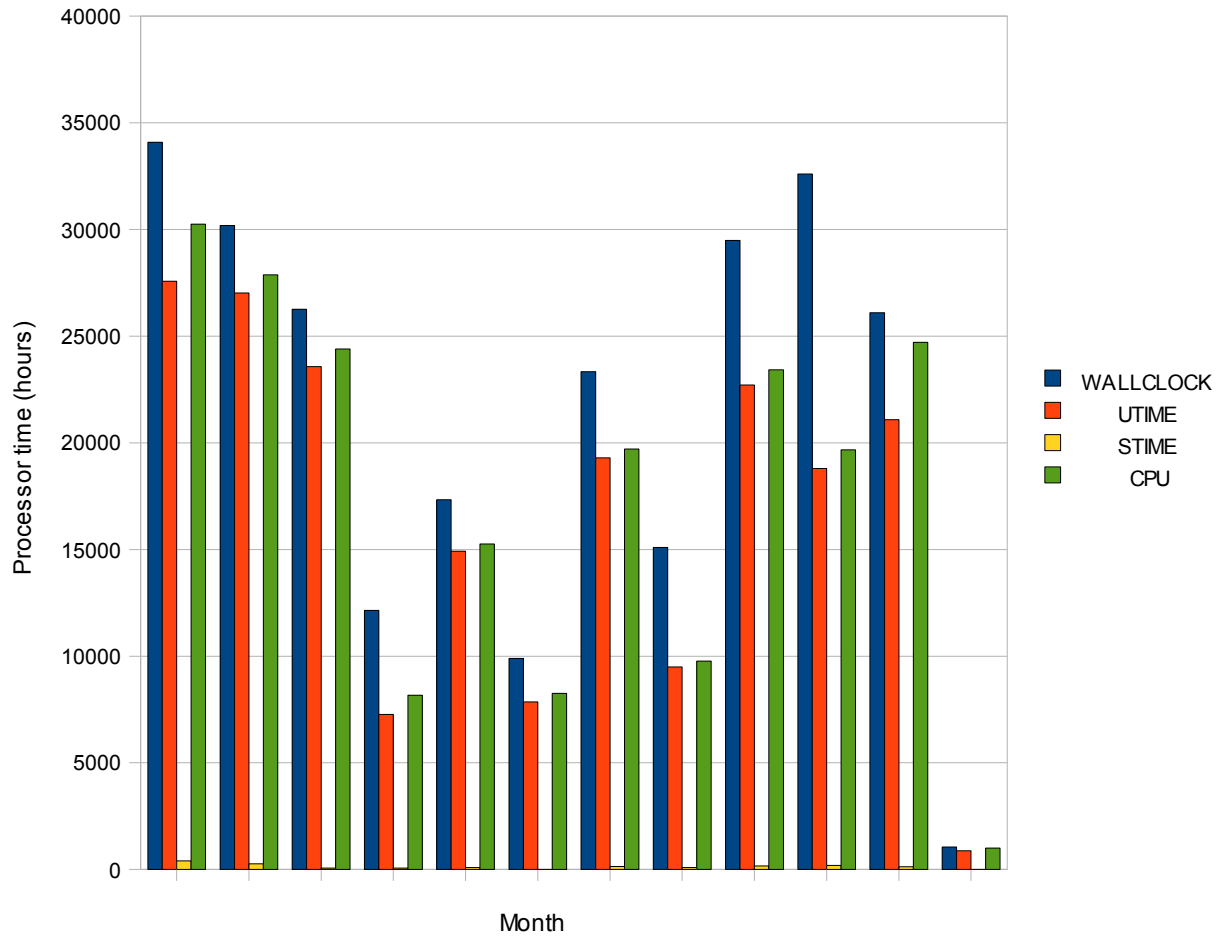


Cluster usage (townhill) (Active users per month)



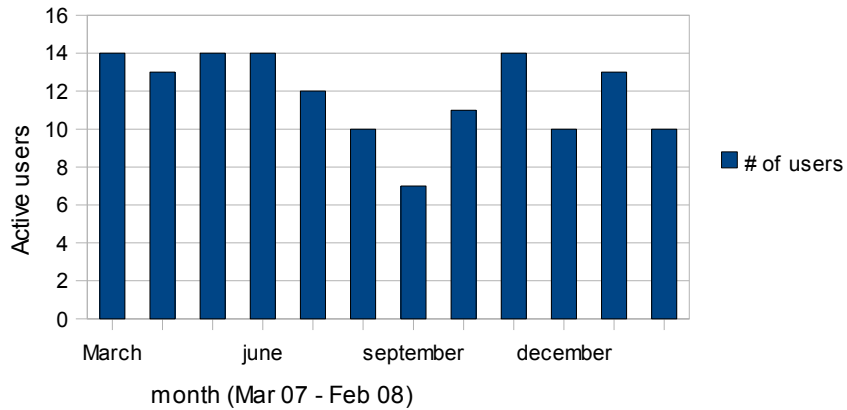
Processor Time totals per month (Lion)

February 2007-February 2008



Cluster usage (Lion)

(active users per month)



Usage survey

All users submitting jobs to the clusters over the last month were surveyed with the following results (from 46 users, 17 of whom replied)

Question	Results
Which clusters have you used in the past	Hermes: 12 Townhill: 12 Lion:12 Lutzow:8 Most people replying used all clusters available to them
Which clusters are you currently using	Hermes:9 Townhill:10 Lion:7 Lutzow: 4 of those that answered 7 use Eddie in addition to our clusters.
If you've stopped using the clusters why have you stopped?	On maternity leave: 1 finished research: 1 clusters too busy: 2 unfair queuing policy: 2 Of the latter two categories I get the impression that they'd still be using our clusters if they could schedules jobs quickly.
Do you use Eddie?	Yes: 8 ("intending to" counted as yes.)
Baseline or paid	Baseline: 2 Paid: 4 one person would use paid service in future if necessary
If not why not?	Jobs run longer than Eddie job limit: 1 not allowed access: 1 insufficient disk space/resource limits: 3 using condor: 1 never heard of it: 2
User comments	provide a cluster for few but long running (2 weeks) jobs. "Not strictly related to clusters, but I would like to say that the Condor system is a fantastic idea

and would like to see more computers added to the pool.”

“I used eddie for my phd experiments. I was using the informatics clusters, but found that: The scheduling on eddie seems better - newly submitted jobs run immediately. It must restrict the jobs of individual users. With the informatics clusters I had to wait over a week sometimes for new jobs to run, which is a disaster for debugging. The eddie cluster is much, much faster. I used the baseline service but performance was many times faster than the informatics cluster, I very soon switched completely.”

provide for more people, coz they are always full!!

With Eddie being provided, it seems redundant to have school clusters as well. However, loss of these might increase the load on Eddie and make jobs take longer to run. If the school clusters were to be discontinued, it would be good to transfer the funding to Eddie to provide more nodes and more disk space.

It would be nice to be able to specify job priority.

A node or two should be kept available for interactive shell jobs. It is almost impossible to get one most of the time, which means it is difficult to debug a script on an actual node. This results in long wait time while a script is submitted, waits to run on a node, and then needs to be corrected.

The queuing and job management is still quite awkward. It would be nice to have more sophisticated queuing scripts available (for instance, not store input and output in the default directory). These are options in the current system, but require effort and overhead

It would be a better use of our resources to put the money into time and resources with ECDF. Otherwise, we need a fairer queuing system and

	<p>faster disk access.</p> <p>Provide a unified support resources page which lists all the servers/clusters and their different characteristics and documentation, for users to get a unified vision of what they have available.</p> <p>townhill seems not fast enough so i'm moving my jobs to Eddie.</p> <p>make it reliable !</p>
--	---

SSH access to individual nodes on Lion.

As this has been an issue in the past I emailed the beowulf support mailing lists, left messages on the MOTD and looked at the “last” logs since December 2007. Finally I mailed people who had commented in the past. I have been unable to identify anyone still accessing the clusters in this way.

Cost.

In order to give an estimate of the costs involved in running the clusters the following table was compiled. The “published power consumption” figures are those given in the Manufacturers specifications and will Usually apply to machines with all expansion bays filled. The measured power consumption was measured using an inline power meter provided by the MPU. The costs are based on a £0.10/KWh tariff. Note that the costs don't include the the file servers currently serving the clusters as these are not dedicated beowulf machines.

Model	Nominal power (Watts)	Observed Power draw (Kwh)	Period (hrs)	Observed power (watts)	Annual cost based on published power consumption	Annual cost based on measured power consumption
Gx240	160	18.76	329.5	56.93	£140.26	£49.91
P530	460	26.97	337.5	79.91	£403.24	£70.05
Pe1425(hermes)	460	14.94	141.58	105.52	£403.24	£92.50
PE1425(townhill)	460	96.21	333	288.92	£403.24	£253.27
Illustrious (PE1425)	460	2.02	19.25	104.94	£403.24	£91.99
Lion(gx240)	160	18.76	329.5	56.93	£140.26	£49.91
Lutzow(P530)	460	26.97	337.5	79.91	£403.24	£70.05
Townhill(d530)	185	9.41	85.5	110.06	£162.17	£96.48
moselle(gx270)	160	0.17	3	56.67	£140.26	£49.67
seville(gx620)	275	7.04	59	119.32	£241.07	£104.60
arnie(pe650)	230	0.22	3.2	68.75	£201.62	£60.27
beatty(PE4600)	300	6.2	23.5	263.83	£262.98	£231.27
Procurve 2650	100	0.76	19.25	39.48	£87.66	£34.61
Procurve 4108gl	630	0.98	4.5	217.78	£552.26	£190.90
Procurve 8000	157	2.25	22	102.27	£137.63	£89.65
Belkin KVM	15	0.17	23.5	7.23	£13.15	£6.34

This shows costings for the configuration we've been using up until Feb this year.

	# nodes	Type	No of items	Networking kit	No of items	Additional kit	Annual Cost based on published power consumption	Annual cost based on observed power consumption
Lion	81	Gx240	1	Procurve 4108	4	Belkin KVM	£11,965.59	£4,239.87
Lutzow	17	P530	1	Procurve 8000	1	Belkin KVM	£7,005.79	£1,286.84
Hermes	24	PE1425(hermes)	1	Procurve 2650				
	1	Gx620					£10,006.39	£2,359.25
Townhill	34	PE1425(townhill)	1	Procurve 2650				
	1	D530					£13,959.86	£8,742.14
Supporting machines	1	beatty(PE4600)					£262.98	£231.27
	1	Illustrious					£403.24	£91.99
	1	Seville					£241.07	£104.60
						Total Cost	£43,844.90	£17,055.97

This shows the configuration we're switching to.

	# nodes	Type	No of items	Networking kit	No of items	Additional kit	Annual Cost based on published power consumption	Annual cost based on observed power consumption
Desktop Cluster	81	GX240	1	Procurve 4801gl	5	Belkin KVM		
	18	PE530	1	Procurve 8000				
	1	GX270						
	1	PE4600						
							£19,777.85	£5,865.03
Rackmounted Cluster	34	PE1425(townhill)	2	Procurve 2650				
	24	PE1425(hermes)						
	1	seville(gx620)						
	1	hermes(gx620)						
	1	arnie(PE650)						
							£24,246.76	£11,169.78
Supporting Machines	1	Illutstrious					£403.24	£91.99
						Total	£44,427.84	£17,126.80

Unfortunately there isn't a good way to measure relative performance of the different clusters without making arbitrary assumptions about the kind of jobs we're running.

Support Issues.

Support is split between the RAT unit and frontline support. As an indicative estimate, the support cost for all the clusters (including townhill) over the past six months was approximately 0.3FTE (ie. a third of one persons time). This includes all operational and development effort from the RAT Unit but does not include effort from frontline support. This is a significant figure in relation to the £17K electricity Bill, in monetary value (at FEC rates) it would be about £28k per year. The cluster reconfiguration should reduce it quite a bit however.

The majority of support time is split between development and recovering disabled nodes.

In the past this has generally been caused by the nodes running low on memory until db4 based services (usually slapd, sometimes updaterpms) start dropping records, corrupting the database or until the OOM killer is invoked and more or less random processes are killed. This has been a particular problem for Lion nodes because we allow SSH access to the individual nodes and clashes between locally started jobs and scheduled jobs happened fairly frequently. Nodes in this condition require significant support intervention to get restarted, rebooting the node will not recover it.

The head nodes are particularly susceptible to database corruption. Database recovery can take several hours and needs a certain amount of knowledge of db4 and gridengine. This hasn't been an issue until recently but we've had a spike in db4 corruptions since the start of the year. Since the software has been pretty much static for the last 5-6 months and the problems are dropping back to a lower level it's assumed this is largely down to a new group of users who were unclear about how to use the clusters.

In the reconfigured clusters we will no longer run a local slapd and the sgeqmaster processes has been moved to a separate host. We've also doubled the number of hosts that jobs can be submitted from. All this should mean that nodes can be rebooted to recover from problems and support costs should go down . Unfortunately it also introduces a new single point of failure in the remote LDAP server

Increasingly over the last 2 years we've seen nodes disabled because amd has failed to mount a remote filesystem or because NFS has IO problems with the mount. Given the load on the fileserver at the time and the network traffic it appears that the NFS server becomes too heavily loaded and client requests timeout. The client node can usually be recovered by restarting amd but this will usually kill any other jobs on the node which are accessing any other NFS filesystem. Given that NFS doesn't scale this will be an increasing problem and effectively prevents us from using the full network bandwidth of the 1425 based machines.

Comments.

In general I think it's fair to say that anyone in the school running jobs on eddie will run them on any school cluster that it's possible for them to run their jobs on. Also I can't see any technical reasons why any jobs which run on our clusters can't run on eddie, ECDF may set policies that make it difficult or impossible for some people to run jobs but I don't think it's in their interest to do this.

I can't see the point in moving the desktop cluster as it currently exists. The hardware is now well out of warranty and we're starting to see hardware failures. Moving it will mean essentially taking it to pieces and reassembling it. If we are to move it I would suggest moving the infrastructure and using whatever kit support usually uses as cold spares with a core of older machines.

Finally this is perhaps a good time to consider the Schools medium to long term requirement for dedicated cluster provision. The rackmount cluster still has several years of usable life and there's a considerable amount of sensible development that can be done (GPFS, integration with the DICE monitoring system, accounting, quotas...) but it may be more appropriate to spend that development time on other services that can't be provided by eddie - condor?