

Apollo HPC Cluster

Jeremy Maris
Research Computing
IT Services
University of Sussex

Apollo Cluster – people

- IT Services
 - Jeremy Maris
 - Alhamdu Bello
 - Bernie Broughton
- Maths and Physical Sciences
 - EPP Vacant (was Matt Raso-Barnett)
 - Albert Asawaroengchai



Apollo Cluster - Aims

- Shared infrastructure and support from IT Services
- Fairshare use of central resources
- Extension of facility by departments
 - Storage (adding Lustre OST, SAN storage)
 - CPU (power paid bt
 - Software Licenses
- Departments guaranteed 90% exclusive of their nodes.
10% sharing with others, plus back fill of idle time.
- Enhancement by IT Services as budgets allow

Apollo Cluster - Data Centre

- 24 x 45 U Water cooled racks
- 96 A per rack, 18Kw cooling
- Current capacity ~ 350Kw
- Upgrade with another 5 racks + PDU
- UPS 5 minutes
- 1MW generator
- 2 x 10Gb JANET

Apollo Cluster - Hardware

- Current total of ~3250 cores.
 - Physics 1216 16 x 64, 8 x 12, 4 x 16 + 2 GPU nodes
 - GridPP 304 4 x 64, 3 x 16
 - Engineering 400 16 x 16, 2 x 64 4x K40 GPU
 - Informatics 256 4 x 64
 - BSMS 128 2 x 64
 - Chem 176 16 x 8 + 3 x 16
 - Lifesci 128 1 x 64 core + 4 x 16 core
 - Economics 16 1 x 16
 - ITS 456 mainly Intel 12 core nodes 48GB RAM/node
- 40 TB Home NFS file systems 500 TB Lustre file system (scratch), QDR IB, IPoIB
- Bright Cluster manager, Univa Grid Engine

Apollo Cluster - Lustre

- Patched 2.5.3 on Centos 6
 - 8 OSS, 21 OST
 - R510, R730 + MD1200 and MD1400
 - Mix of 2, 3 and 6TB disks
- Subscription to Lustre Community edition
 - \$2000 per OSS per annum on 2 x OSS test system
 - Privileged access to repos
 - EDU support area – can see all EDU tickets
 - Can only raise support tickets on test cluster

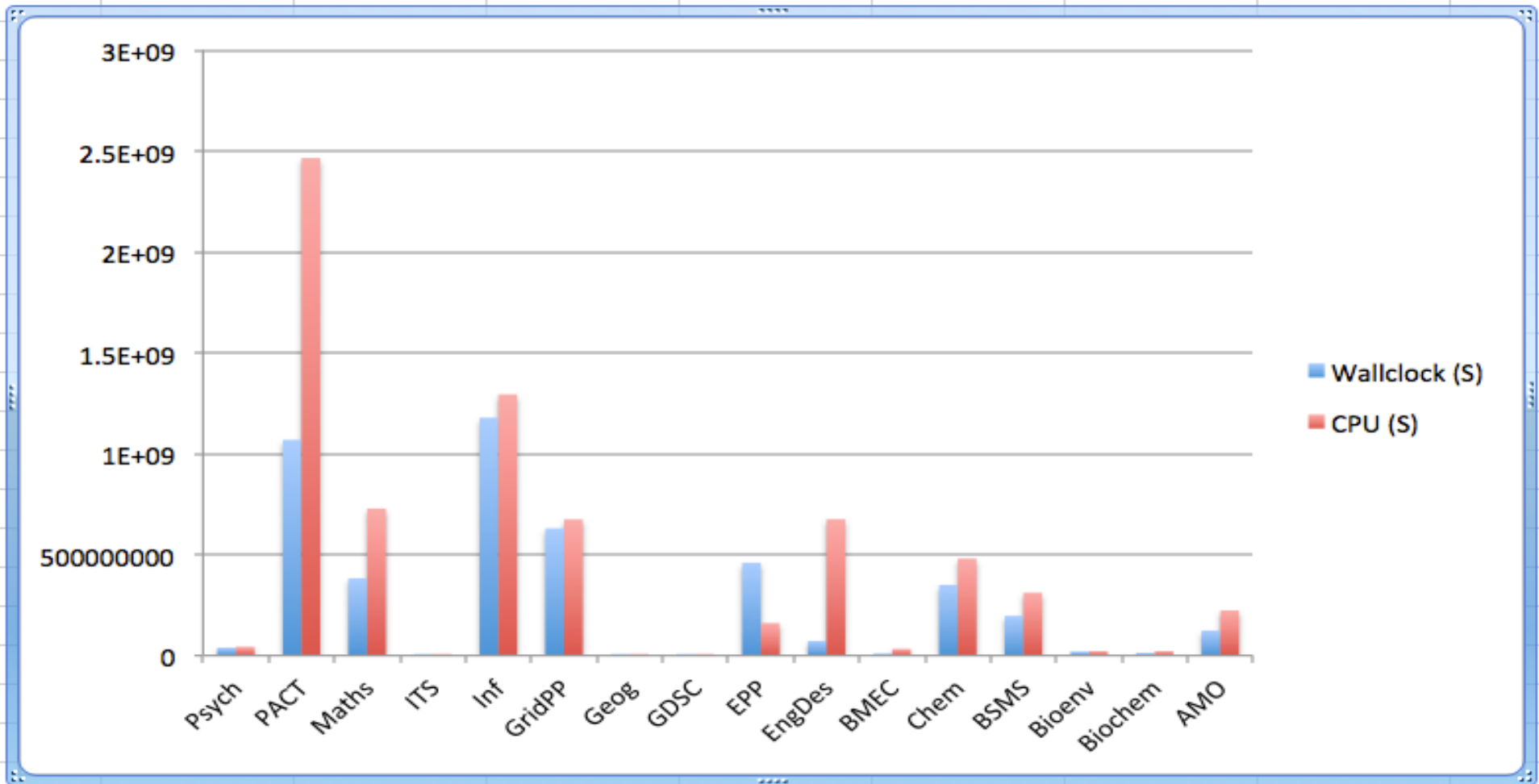
Apollo Cluster - Storage

- Dothill SAN + FalconStor virtualisation
- Nexenta ZFS research storage (140TB)
- NFS home filesystems - R510 and R730
- Astronomy N body simulation from PRACE
 - R730XD, MD1200, MD3640
 - 100TB - > 400TB NFS storage
 - 12 disk RAID6 volumes + LVM
- Backup:
 - Legato Networker + LTO5

Apollo Cluster - provisioning

- Cluster
 - Bright Cluster Manager 6.1 base
 - Image based system,
 - Puppet tailoring for Grid
- Lustre, NFS, Grid service nodes
 - Puppet
 - VMs for Grid service nodes
 - SRM has 1Gb, not Infiniband
 - Need real hardware...

Accounting – 200 active users



Apollo Cluster – Summer Upgrade

- HPE procurement – 55 x 2640 v3 nodes, 880 cores
- Omni-Path half bandwidth tree
- Lustre router between Truescale and Omni-Path
- Centos 7.2 for most nodes, SL6 for Grid
- Bright Cluster Manager 7.1
 - 4 node Hadoop instance
- Univa Grid Engine 8.4.0

Apollo Cluster – Challenges

- ~ 65% utilisation of cluster – fragmented by dedicated queues
- Heterogeneous
 - QDR IB
 - Omnipath
 - AMD nodes
 - Intel nodes
- Use job classes to select appropriate nodes
- Cgroups to limit/manage resources
- More use of fairshare
- Back fill with ATLAS production jobs (~1000 cores..)