# Cluster / Grid Status Update
# University of Sussex
Emyr James

# Contents

- History of the Cluster
- Hardware / Architecture
- Software Inventory
- Usage
- Problems So Far...
- Future Plans

# History

- Money available from grant for Atlas work to pay for hardware & support
- Design – Jeremy Maris, Alces, Dell
- Installation : Above + Clustervision
- Hardware and basic software (OS, SGE, lustre, Atlas) ready to go when I joined

# Architecture : Feynman

- Cluster split into 2 parts – Feynman / Apollo
- Lustre filesystem shared by both
- Infiniband interconnect
- Head Nodes : PowerEdge R610, dual 6 core Xeon 2.67GHz, 48GB Ram (4GB/Core)
- Feynman Compute Nodes : 8x Dell PowerEdge 410 as above
- Feynman : 120 Cores, ~420GB RAM

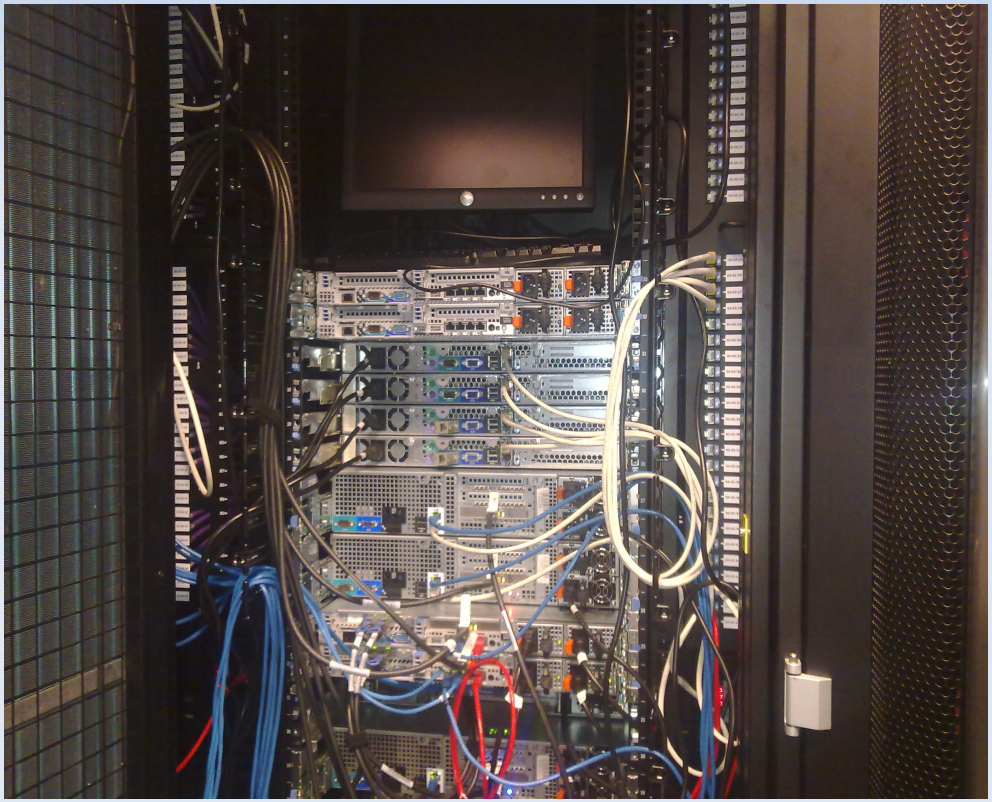
# Architecture : Apollo

- Apollo : 16 Nodes
- 1-10, 13-16 as feynman
- 11,12 :  PowerEdge R815, 4xAMD Opteron 12 cores @ 2.2GHz, 256GB RAM
- Total of 288 cores, ~1.2 TB Ram

# Architecture : Lustre

- 2 MDS's – Automatic Failover
- 3 OSS's each with 3x10TB RAID6 OST's
- Total of 81T available after formatting
- Currently  48T used but needs cleaning up
- Great performance, has been in use with no problems at all for 9 months
- Also NFS server for home directories

# Architecture : Data Centre



# Architecture : Data Centre


# Software

- Sun Grid Engine Scheduler
- Ganglia for Monitoring
- Bright Cluster Manager
- Python scripts for viewing queue / accounting

# Usage

- Mainly Atlas at the moment
- Typical usage : download data to lustre using dq2-get on head node, run Root analysis jobs on compute nodes
- Analysis done contributed to 2 Atlas papers so far – SUSY related
- Snoplus group starting to use it
- CryoEDM developing MC code

# Lustre Performance

| Servers | All-OSS | | OSS1 | OSS2 | OSS3 |
|---|---|---|---|---|---|
| **Clients** | **18** | | **6** | **6** | **6** |
| **metric** | **KB/s** | **IOPS** | **KB/s** | **KB/s** | **KB/s** |
| write | 3,854,980 | 60,235 | 1,273,054 | 1,272,119 | 1,273,239 |
| rewrite | 3,005,168 | 46,956 | 1,039,489 | 1,117,662 | 1,020,844 |
| read | 2,982,143 | 46,596 | 1,143,591 | 1,164,641 | 1,165,076 |
| reread | 18,411,516 | 287,680 | 9,127,979 | 8,918,498 | 7,724,083 |
| random read | 151,069 | 2,361 | | | |
| random write | 261,675 | 4,089 | | | |
| reverse read | 251,587 | 3,932 | | | |
| stride read | 303,852 | 4,748 | | | |
| mixed | 280,632 | 4,385 | | | |

# Problems

- Using NFS server for job IO – user education and boilerplate scripts
- LDAP connection issues – use nscd
- DNS registration issue
- Lustre RAID Server battery cache
- Thunder!
- No problems due to lustre filesystem itself

# Future

- 'Gridification' ongoing to become part of SouthGrid GridPP
- Investigating procurement options for 60TB addition to lustre FS
- Adding more nodes – snoplus grant application, pooling money from ITS & other research groups
- Reorganise into 1 cluster