LHC Data Taking on the Grid

An infrastructure perspective

Jeremy Coles
ISGC2010 - Taipei, Taiwan
9th March 2010
Overview

- Did we see any impacts (learn anything) from first collision data!?
- Situation by experiment
- Tier-2 issues (e.g. data access)
- Tier-1 status and issues
- GridPP project directions
- GridPP infrastructure directions
- The wider grid work
Did we see ANY impact from first LHC data!

- We are small contributors to ALICE so the focus here is LHCb, ATLAS and CMS.

Traffic over the OPN to and from CERN in the November/December 09 period. No issues.

Cumulative normalised Elapsed time – Sept 09 to Feb 10. No sharp rise in usage.
• Due to the low crossing rate got a total of 220GB data.

• **Good latency for jobs** – typically of order 10 minutes. RAL performed well.

• Expect chaotic running in 2010 – changing LHC conditions and commissioning of trigger.

The main LHCb issue has been Tier-1 job efficiencies. Originally we thought this may be related to the number of running jobs (and i/o wait) but further investigation indicates that it is **inefficient user jobs**. The plot here shows the clustering for 2 weeks in January where this was analysed.

Ref: [http://www.gridpp.rl.ac.uk/blog/2010/03/02/low-efficiency-lhcb-jobs/](http://www.gridpp.rl.ac.uk/blog/2010/03/02/low-efficiency-lhcb-jobs/)
• CPU: We did see a peak in activity (jobs running) over the November/December period but nowhere near full utilisation.

Reprocessing: Issue with T1 - final merge step of reprocessing started to run out of wallclock on the 3000M queue. Limit changed and things ran fine.

UK cloud had 8 sites taking subscriptions. Largest was Glasgow – hosts min-bias analysis group. User analysis at this level works.

ATLAS UK had recent problems with both poorly justified data requests and requests to place data in inappropriate space tokens.

Jamboree - noted site downtimes still an issue
Transfers worked. Little impact from CMS activities.
Data management – integrity checking …

… started just before data taking

(1) Accuracy of the information systems

- Differences in what SRM and experiment tools see at a site can waste resources
- In GridPP we are currently cross-checking figures for each site and correcting where necessary.

(2) We have started a programme of checksum validations

- There are several reasons a file may not be written correctly.

(3) Storage optimisation – site issue

(4) Data access efficiencies – site & experiment issue

Ref: S Jézéquel ADC jamboree

Aside: We are not alone!
- 140 TB - Large Synoptic Survey Telescope every 5 days (2016)
- 2.5 PB - Estimate of Wal-Mart's database size
- 1 PB - Estimate of what Google processes every hour
- 1.7 PB - Estimate of amount of information in existence in 2010

Source: The Economist
Data access – improving performance

- Use xfs if possible. Tests showed an almost two-fold performance increase. Also less system load.
- FTS: number of files and number of streams requires careful evaluation. Easy to overload a server and under perform on network usage.
- 1Gb/s external connectivity is a minimum
- Use several servers to spread load across SE – typically get 300Mb/s on each (so 3 min)
- For DPM – optimisation of the mySQL database can lead to large performance gains. The backend is re-written for every transaction.
- Make use of mySQL binary logging. Important issues for backups. Give mySQL plenty of buffer (DBs created as innoDB databases within mySQL – better to let it cache its operations).
- Reduce headnode load by putting DPM and mySQL services on different machines.
- DPM write algorithm is very basic – distributes equally between file systems. Add more fs to balance new/old servers and i/o capabilities.

http://www.gridpp.ac.uk/wiki/Performance_and_Tuning
also for dCache: http://www.gridpp.ac.uk/wiki/Optimising_dCache_Performance

Jeremy Coles – Taipei – 09/03/2010
ATLAS pseudo-random access vs reordered

Local: ext3

Remote: rfio

ISGC2010 – LHC data – An infrastructure perspective

Jeremy Coles – Taipei – 09/03/2010
GPFS vs rfio

Remote: GPFS

Remote: rfio

GPUFS vs rfio

Remote: GPFS

Remote: rfio

Reordered

Reordered
1) As shown reordered files are much more efficient to access.

2) Optimising local access is probably still important (protocols, hot files (e.g. using pCache)). Last year GridPP did various studies on rfio buffer size impacts on efficiency. (see plot for Glasgow server performance issues)

Tuning the rfio buffer size. Factors include number of concurrent jobs, the file access type (random or sequential), the size of file and amount of data being processed.

Buffering should maximise the amount of data being accessed from worker node RAM, minimise the number of requests to the storage servers over the network and minimise the number of IO operations the storage has to process.

Reading directly from the network with no buffering incurs a heavy penalty in overhead (TCP/IP headers, network latency and IO operations) compared to reading from RAM. But, reading in more of the file into buffer than is needed also incurs a overhead in wasted network bandwidth and wasted cpu time while waiting for the read to complete.

3) What next? How to manage upgrades in GridPP once we are taking data

4) Performance of, and future for, other file systems

5) Monitoring to ensure (and improve) performance. Changes as experiment approaches evolve.
Reliability is still a core need but....

<table>
<thead>
<tr>
<th></th>
<th>CE</th>
<th>SRMv2</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours ago: 9 8 7 6 5 4 3 2 1 0</td>
<td>9 8 7 6 5 4 3 2 1 0</td>
<td>24 Hrs Week Month 6 Mon</td>
<td></td>
</tr>
<tr>
<td>EFDA-JET</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>96% 71% 92% 97%</td>
</tr>
<tr>
<td>RAL-LCG2_Tier-1</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>100% 100% 98% 90%</td>
</tr>
<tr>
<td>UKI-LT2-Brunel</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>100% 78% 94% 96%</td>
</tr>
<tr>
<td>UKI-LT2-IC-HEP</td>
<td>P E E E E E E E E E E E E E</td>
<td>P E E E E E E E E E E E E E</td>
<td>83% 75% 84% 86%</td>
</tr>
<tr>
<td>UKI-LT2-QMUL</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>100% 78% 89% 88%</td>
</tr>
<tr>
<td>UKI-LT2-RHUL</td>
<td>E E E E E E E E E E</td>
<td>P P P P P P P P P P</td>
<td>0% 0% 18% 76%</td>
</tr>
<tr>
<td>UKI-LT2-UCL-CENTRAL</td>
<td>M M M M M M M M M M</td>
<td>P P P P P P P P P P</td>
<td>100% 99% 37% 46%</td>
</tr>
<tr>
<td>UKI-LT2-UCL-HEP</td>
<td>M M M M M M M M M M</td>
<td>M M M M M M M M M M</td>
<td>0% 84% 91% 93%</td>
</tr>
<tr>
<td>UKI-NORTHGRID-LANCS-HEP</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>91% 75% 94% 87%</td>
</tr>
<tr>
<td>UKI-NORTHGRID-LIV-HEP</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>100% 78% 91% 98%</td>
</tr>
<tr>
<td>UKI-NORTHGRID-MAN-HEP</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>100% 96% 94% 91%</td>
</tr>
<tr>
<td>UKI-NORTHGRID-SHEF-HEP</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>78% 75% 88% 96%</td>
</tr>
<tr>
<td>UKI-SCOTGRID-DURHAM</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>100% 99% 96% 94%</td>
</tr>
<tr>
<td>UKI-SCOTGRID-ECDF</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>100% 76% 92% 93%</td>
</tr>
<tr>
<td>UKI-SCOTGRID-GLASGOW</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>100% 96% 99% 97%</td>
</tr>
<tr>
<td>UKI-SOUTHGRID-BHAM-HEP</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>100% 64% 86% 85%</td>
</tr>
<tr>
<td>UKI-SOUTHGRID-BRIS-HEP</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>100% 73% 75% 95%</td>
</tr>
<tr>
<td>UKI-SOUTHGRID-CAM-HEP</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>100% 64% 76% 92%</td>
</tr>
<tr>
<td>UKI-SOUTHGRID-OX-HEP</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>100% 75% 93% 88%</td>
</tr>
<tr>
<td>UKI-SOUTHGRID-RALPP</td>
<td>P P P P P P P P P P</td>
<td>P P P P P P P P P P</td>
<td>100% 75% 77% 83%</td>
</tr>
<tr>
<td>csTCDi</td>
<td>P P F F F F F F F F</td>
<td>P P F F F F F F F F</td>
<td>83% 82% 93% 94%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>92% 81% 85% 89%</td>
</tr>
</tbody>
</table>
Sites ran very smoothly during December – and the Christmas vacation!
Red line shows the Tier-1 availability – several outages (explained next)
Green line shows GridPP/UK average.
Didn’t we address resilience?

“The best defense is a good offense”. Being proactive here can save time, money and effort. There are several areas we need to consider, here some major ones:

• Duplicate services or machines
• Increase the hardware’s capacity (to handle faults)
• Use (good) fault detection
• Implement automatic restarts
• Provide fast intervention
• Fully investigate failures
• Report bugs -> ask for better middleware

The image on the left illustrates the trade-offs between implementation costs, technology availability, and risk. The image on the right depicts a scenario where a chief information officer

Chief Information Officer
"The buck stops here"


A full redesign?
No chance!
I'm tired of throwing money at Oracle problems!
Just move it to a faster server!
Tier-I – what can go wrong!?

- Mid-July: Move to new building… thereafter plenty of “tests” of our disaster management process!
- 12th August: Water leak onto the tape robot. Turned out to be a leak from air-conditioning condensate.
- 10th-12th August: Air conditioning failure (twice) during the night. Castor and batch services down. One failure was caused by the chillers stopping following a reboot of the building management system. The other failure was due to an over pressure alarm.
- 4th October. Four disk arrays failed in a short timescale. Two of these arrays hosted the Oracle databases behind Castor, the other two the Oracle databases behind the LFC, FTS and 3D services.
- November: Data loss! The problem occurred with the database that included the Castor ‘nameserver’. Database used “incorrect” backup.
- January: disk array issues traced to noise on the electrical current supplied by UPS – impedance mismatch.
- February: While carrying out the planned migration of the Oracle Castor databases encountered stability problems - a reboot of one of the nodes in an Oracle RAC would cause other nodes in that RAC to crash. SAN issue.
- Equipment acceptance issues

“Most disasters are the result of a collection of relatively small events happening at the same time in response to a common trigger”.

IBM
Are we reliable enough?

Trying to better integrate a shared university cluster

<table>
<thead>
<tr>
<th></th>
<th>2Q07</th>
<th>3Q07</th>
<th>4Q07</th>
<th>1Q08</th>
<th>2Q08</th>
<th>3Q08</th>
<th>4Q08</th>
<th>1Q09</th>
<th>2Q09</th>
<th>3Q09</th>
<th>4Q09</th>
<th>1Q10</th>
<th>Recent</th>
</tr>
</thead>
<tbody>
<tr>
<td>EFDA-JET</td>
<td>98%</td>
<td>78%</td>
<td>83%</td>
<td>83%</td>
<td>82%</td>
<td>81%</td>
<td>83%</td>
<td>91%</td>
<td>92%</td>
<td>77%</td>
<td>84%</td>
<td>93%</td>
<td>98%</td>
</tr>
<tr>
<td>RAL-LCG2_Tier-1</td>
<td>83%</td>
<td>84%</td>
<td>86%</td>
<td>85%</td>
<td>94%</td>
<td>90%</td>
<td>94%</td>
<td>85%</td>
<td>88%</td>
<td>90%</td>
<td>85%</td>
<td>95%</td>
<td>94%</td>
</tr>
<tr>
<td>UKI-LT2-Brunel</td>
<td>95%</td>
<td>92%</td>
<td>95%</td>
<td>95%</td>
<td>96%</td>
<td>95%</td>
<td>96%</td>
<td>92%</td>
<td>93%</td>
<td>97%</td>
<td>93%</td>
<td>94%</td>
<td>96%</td>
</tr>
<tr>
<td>UKI-LT2-IC-HEP</td>
<td>93%</td>
<td>93%</td>
<td>97%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>92%</td>
<td>94%</td>
<td>88%</td>
<td>94%</td>
<td>98%</td>
<td>88%</td>
</tr>
<tr>
<td>UKI-LT2-QMUL</td>
<td>98%</td>
<td>94%</td>
<td>89%</td>
<td>83%</td>
<td>91%</td>
<td>89%</td>
<td>85%</td>
<td>92%</td>
<td>91%</td>
<td>84%</td>
<td>89%</td>
<td>86%</td>
<td>85%</td>
</tr>
<tr>
<td>UKI-LT2-RHUL</td>
<td>99%</td>
<td>85%</td>
<td>86%</td>
<td>86%</td>
<td>86%</td>
<td>85%</td>
<td>83%</td>
<td>83%</td>
<td>83%</td>
<td>83%</td>
<td>83%</td>
<td>83%</td>
<td>83%</td>
</tr>
<tr>
<td>UKI-LT2-UCL-CENTRAL</td>
<td>96%</td>
<td>85%</td>
<td>83%</td>
<td>86%</td>
<td>89%</td>
<td>88%</td>
<td>80%</td>
<td>73%</td>
<td>81%</td>
<td>80%</td>
<td>73%</td>
<td>81%</td>
<td>81%</td>
</tr>
<tr>
<td>UKI-LT2-UCL-HEP</td>
<td>89%</td>
<td>82%</td>
<td>81%</td>
<td>85%</td>
<td>81%</td>
<td>76%</td>
<td>88%</td>
<td>81%</td>
<td>86%</td>
<td>78%</td>
<td>82%</td>
<td>85%</td>
<td>85%</td>
</tr>
<tr>
<td>UKI-NORTHGRID-LANCS-HEP</td>
<td>96%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
</tr>
<tr>
<td>UKI-NORTHGRID-LIV-HEP</td>
<td>98%</td>
<td>93%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
</tr>
<tr>
<td>UKI-NORTHGRID-MAN-HEP</td>
<td>97%</td>
<td>98%</td>
<td>98%</td>
<td>98%</td>
<td>98%</td>
<td>98%</td>
<td>98%</td>
<td>98%</td>
<td>98%</td>
<td>98%</td>
<td>98%</td>
<td>98%</td>
<td>98%</td>
</tr>
<tr>
<td>UKI-NORTHGRID-SHEF-HEP</td>
<td>97%</td>
<td>96%</td>
<td>93%</td>
<td>93%</td>
<td>95%</td>
<td>94%</td>
<td>94%</td>
<td>94%</td>
<td>94%</td>
<td>94%</td>
<td>94%</td>
<td>94%</td>
<td>94%</td>
</tr>
<tr>
<td>UKI-SCOTGRID-DURHAM</td>
<td>94%</td>
<td>91%</td>
<td>88%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
</tr>
<tr>
<td>UKI-SCOTGRID-ECDF</td>
<td>85%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
</tr>
<tr>
<td>UKI-SCOTGRID-GLASGOW</td>
<td>89%</td>
<td>89%</td>
<td>89%</td>
<td>89%</td>
<td>89%</td>
<td>89%</td>
<td>89%</td>
<td>89%</td>
<td>89%</td>
<td>89%</td>
<td>89%</td>
<td>89%</td>
<td>89%</td>
</tr>
<tr>
<td>UKI-SOUTHGRID-BHAM-HEP</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
<td>88%</td>
</tr>
<tr>
<td>UKI-SOUTHGRID-BRIS-HEP</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
</tr>
<tr>
<td>UKI-SOUTHGRID-CAM-HEP</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
</tr>
<tr>
<td>UKI-SOUTHGRID-OX-HEP</td>
<td>79%</td>
<td>80%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
</tr>
<tr>
<td>UKI-SOUTHGRID-RALPP</td>
<td>94%</td>
<td>91%</td>
<td>90%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
</tr>
<tr>
<td>csTCDie</td>
<td>82%</td>
<td>82%</td>
<td>81%</td>
<td>81%</td>
<td>81%</td>
<td>81%</td>
<td>81%</td>
<td>81%</td>
<td>81%</td>
<td>81%</td>
<td>81%</td>
<td>81%</td>
<td>81%</td>
</tr>
<tr>
<td>LondonGrid</td>
<td>78%</td>
<td>78%</td>
<td>79%</td>
<td>81%</td>
<td>77%</td>
<td>77%</td>
<td>71%</td>
<td>69%</td>
<td>69%</td>
<td>67%</td>
<td>66%</td>
<td>66%</td>
<td>66%</td>
</tr>
<tr>
<td>NorthGrid</td>
<td>73%</td>
<td>73%</td>
<td>79%</td>
<td>89%</td>
<td>90%</td>
<td>90%</td>
<td>90%</td>
<td>90%</td>
<td>90%</td>
<td>90%</td>
<td>90%</td>
<td>90%</td>
<td>90%</td>
</tr>
<tr>
<td>ScotGrid</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
</tr>
<tr>
<td>SouthGrid</td>
<td>80%</td>
<td>80%</td>
<td>86%</td>
<td>87%</td>
<td>90%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
</tr>
<tr>
<td>Overall</td>
<td>79%</td>
<td>79%</td>
<td>83%</td>
<td>86%</td>
<td>84%</td>
<td>89%</td>
<td>87%</td>
<td>84%</td>
<td>93%</td>
<td>89%</td>
<td>89%</td>
<td>89%</td>
<td>89%</td>
</tr>
</tbody>
</table>
An example of minor things having large impacts…..

- The plot red line shows the number of jobs transferring during the ATLAS reprocessing exercise in February. To fix this problem took some investigation and to fix required only a minor configuration change to the ATLAS LSF (24hr to 4hrs for the log files)
Current issues

- ATLAS has started to run jobs that require >3GB. At present most RAL T1 WNs are 8 core with 16GB RAM which is 50% overcommitted. This means 8 jobs can run. Changing the requirement to 4GB/slot leads to a possible capacity reduction of 25%.

Ref: http://www.gridpp.rl.ac.uk/blog/2010/03/04/atlas-memory-limit-changes/
http://www.gridpp.rl.ac.uk/status/
Future directions – The project

- GridPP is funded in (approximately) 4 year cycles. GridPP3 finishes in March 2011. The collaboration has recently submitted a proposal for GridPP4 “Computing in the LHC era”.

Some highlights:
- The overall funding request will be of order £30M over 4 years
- Contains 6 work packages (T1, T2, Deployment ops & support, Expt. Support, Management, Impact.)
- Some changes in Tier-2 and operations structures are required to match experiment needs
- We need to address the topical areas of economic impact and knowledge transfer
- Takes account of GridPP contribution to UK NGI
Future directions – the infrastructure

- CREAM CE deployed at several UK sites. Main issue is with Lease and Proxy Renew processes for ATLAS condor submissions. Awaiting CREAM 1.6 and a version for the SGE and Condor batch systems.

- SCAS/glexec now in production at several sites.

- Equipment renewal

- Virtualisation. Several sites now have VM based services. Not storage headnodes.

- Cloud computing. GridPP4 funding request includes a study of running on a commercial cloud vs MC performed at Tier-2s.

Ref: http://scotgrid.blogspot.com/2010/02/cream-sours.html
Wider community support

- ATLAS dominates usage of UK resources but we have good usage from some smaller VOs such as H1 and UK phenomenology studies. Outside of particle physics our resource are used by various projects.

Graph: UKI Cumulative Normalised Elapsed time (HEPSPEC08) by VO and DATE

• Analysed 18M images Nov/Dec 09

Text analytics
- e.g. Improve on Google’s n-gram corpus using a web spider to hit $10^{12}$ words.

Developing next generation search engines

Search image content
- Analysed 18M images Nov/Dec 09

Text analytics
- e.g. Improve on Google’s n-gram corpus using a web spider to hit $10^{12}$ words.
Summary

- No major issues observed with any of the LHC experiments with first data taking, reprocessing and analysis. But then nothing was really stressed.
- Stability of sites (especially storage) remains a concern for the experiments.
- Still plenty of improvement opportunities in the areas of storage and data management.
- GridPP project wise there are some changes coming. GridPP4 proposal has been submitted and is under review. Takes account of the new NGI/EGI world.
- Infrastructure wise things are also still moving – CREAM and SCAS are being tested. Active with virtualisation studies. Will look at computing cloud performance (for MC).
- GridPP continues to support non-LHC and non-physics VOs. We expect more resource contention in the coming months.