Step '09 Experiences at the Lancaster Tier-2

Matt Doidge & Peter Love
Gridpp Storage Workshop July 2009

Step '09 Goals.

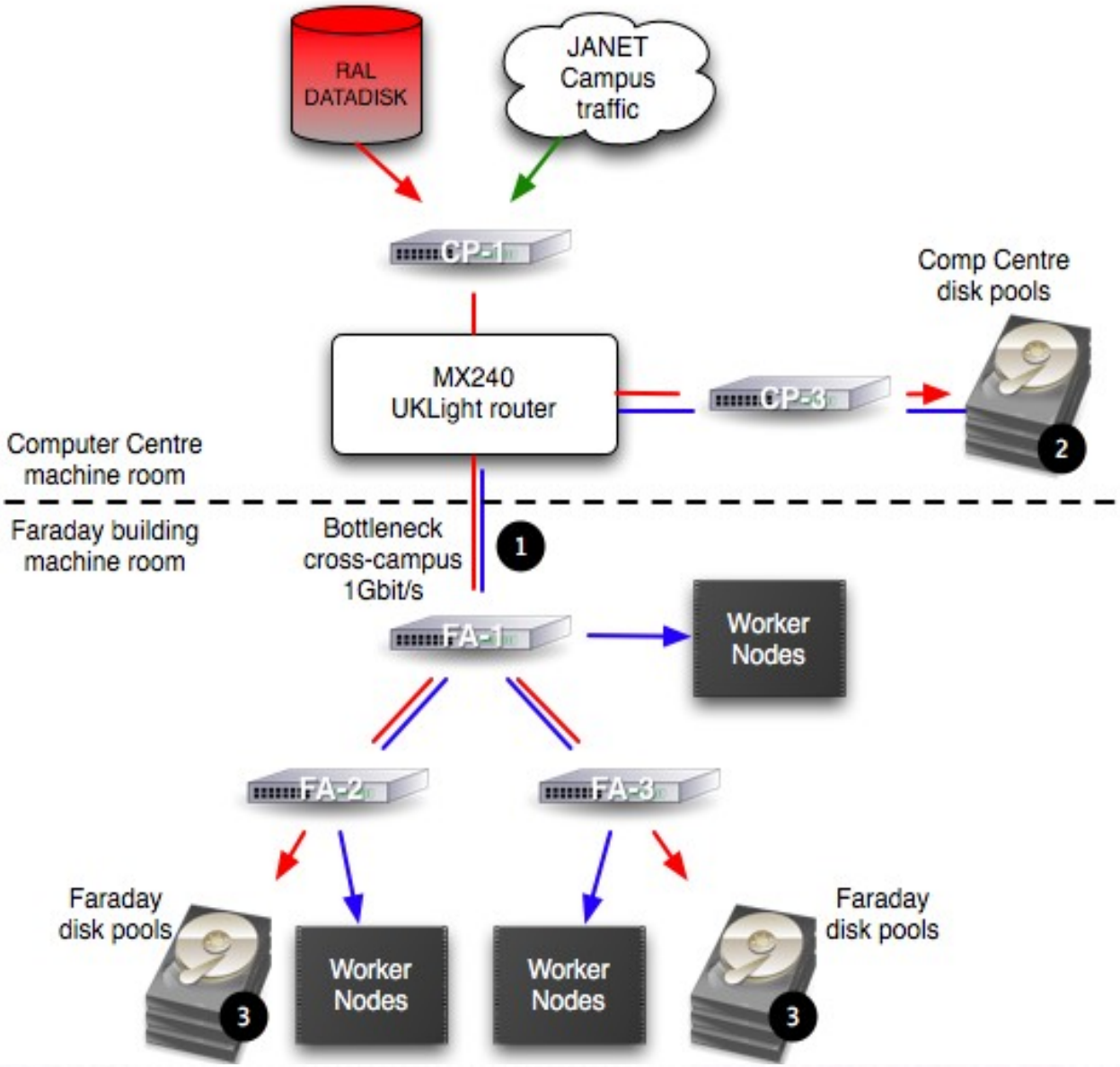
- Lancaster had a number of preset goals to pass during the 2-weeks of Step '09:
  - Run 100 of each type of Analysis Job.
  - Run 200 Production jobs.
  - Accept 30% (34TB) of the total UK AOD and DPD data.
  - Continue to provide resources to other VOs during this time.
  - Do all of the above **concurrently**.

So what hit the fan during Step?

- During Step we had a number of Storage and Network related problems:
  - Network Congestion.
    - Led to a backlog of RAL Transfers.
    - Lowered Job Efficiency.
  - Pool load Imbalance.
  - Some Ill-timed hardware failures.
  - Getting more then our “fair share” of data due to differences in dataset size.
- Plus some other Maui/Software release woes.

Network Congestion.

- A problem with our site it is characterised by having large chunks of storage dangling off a 1Gb Network link.
- Our older and busier nodes are dangling off one of these bits, and contest for bandwidth with traffic from RAL.
- DPM's lack of pool load-balancing led to a greater number of hot files being on our older nodes, further adding to the problem.


RAL
DATADISK
JANET
Campus
traffic
GP-1
Comp Centre
disk pools
MX240
UKLight router
GP-3
2
Computer Centre
machine room
Faraday building
machine room
Bottleneck
cross-campus
1Gbit/s
1
FA-1
Worker
Nodes
FA-2
FA-3
Faraday
disk pools
Worker
Nodes
Worker
Nodes
Faraday
disk pools
3
3

Network Decongestion.

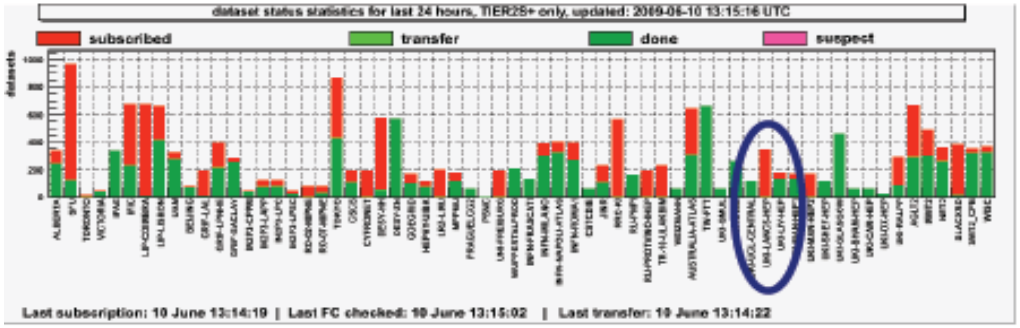
- During step we had to throttle back and kill jobs
  - Not a viable solution during “the real thing”.
- Identifying and shuffling hot files around helped alleviate network balance issues.
  - But this was a clunky procedure.
- Upgrade to a 10G infrastructure, and start stacking “close” switches.
  - Expensive, but Step'09 gave us the ammo we needed to justify the expense. Some of the new equipment arrived this week!

The Data must flow.

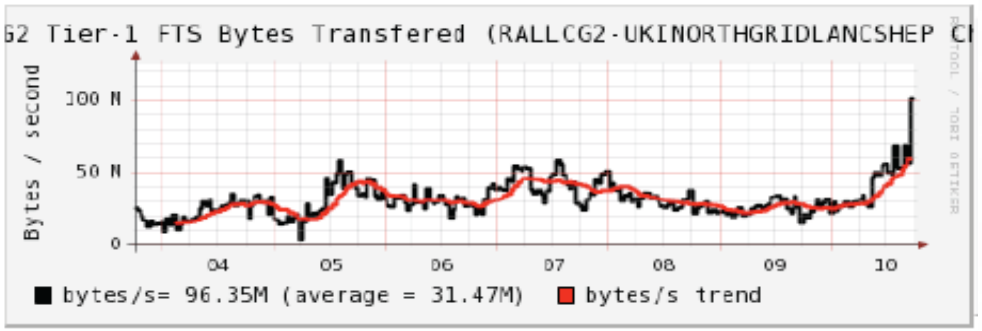
- Due to variations in dataset sizes we actually got more data then the expected 0.3*112TB.
  - This could happen to you too! Always leave margins for error w.r.t. space and bandwidth.
- It doesn't take much of a bandwidth shortfall to create a sizable backlog. Our “nominal”rate was 220 Mbit/s. Our rate in the first week was 200 Mb/s.
- As an experiment we cut down the Muon analysis jobs to just 1 slot. This 1 slot still pulled in 200Mb/s.

Backlog.

- Squint at the opposite plot and you see the lack of green for Lancaster's transfers.

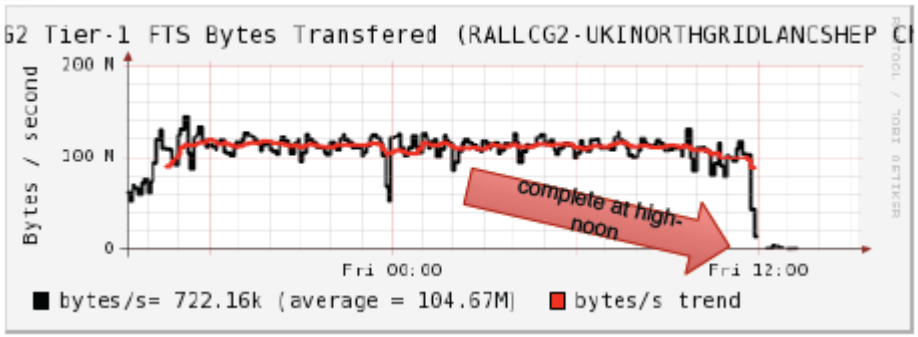


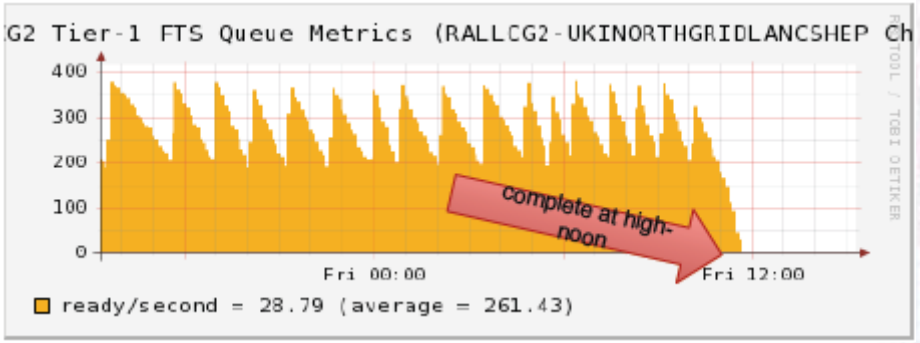
- But drastic measures managed to quadruple our rates towards the end of the challenge.



In the nick of time.

- We managed to obtain all the needed data by:
  - Uping concurrent FTS transfers from 8 to 12.
  - Pruning the number of running jobs (at one point to zero...).
  - Repairing physical problems.



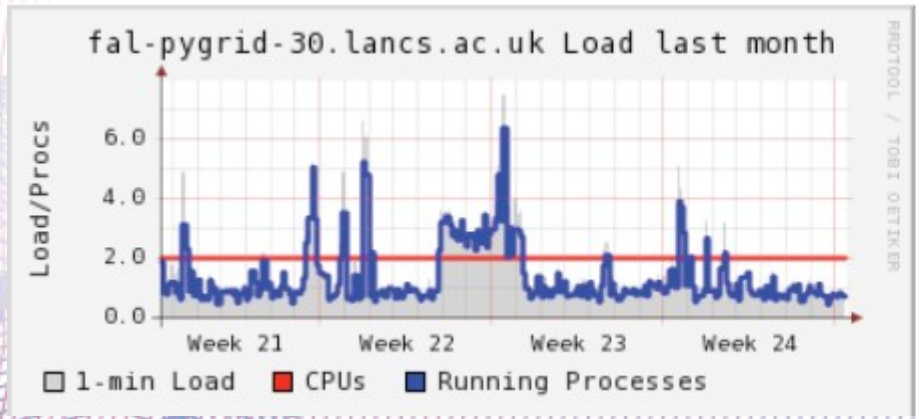


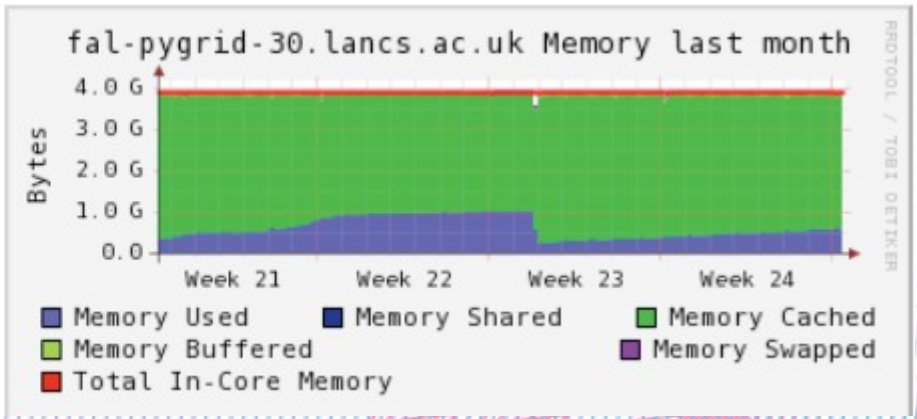
Those Physical Frailties.

- We had a few breakdowns that chose to strike during Step'09.
  - A disk server popped it's system disk
    - Thanks to cfengine and a few years of doing this stuff we were up and running really quickly.
  - Two disk servers had their link degrade to 100Mb/s.
    - It appeared to be due to damaged cat5 cables.
  - The percentage of packets being dropped rose.
    - Need to investigate further, appears to be mainly at the NIC of our older disk servers.

DPM Performance.

- Despite it's crustiness and previous experiences our DPM headnode wasn't strained at all.
  - We had to restart DPM services once due to a memory leak issue that will hopefully begone with the 1.7 upgrade.
- Increased IOWait on the pool nodes.
  - No chance during Step to tune RFIO.
- DPM “file shuffling” was a royal pain to impliment but really helped out.





DPM Woes.

- As I've mentioned a few times, hot files on old nodes proved a problem.
- The DPM round-robin file writing falls down when you have pools of greatly differing ages
  - Hopefully after the 1.7 upgrade we'll be able to spread the files with IF the Token-Trashing has been fixed.
- DPM proving too feature-poor? Would jazzing it up make it lose its charm?

Summary.

- Step '09 was a really useful experience
  - Helped us identify the need, justify the expense and obtain funding for 10G infrastructure.
  - Slammed home the need to be prepared and be flexible.
- But like any exercise, one two-week session is not enough.
  - Step helped us identify core problems, but didn't allow us time to tune.