GridPP

UK Computing for Particle Physics

Storage and Networking Ruminations

Brian Davies

Hepsysman

RAL, June 2017

Science & Technology Facilities Council

GridPP
UK Computing for Particle Physics

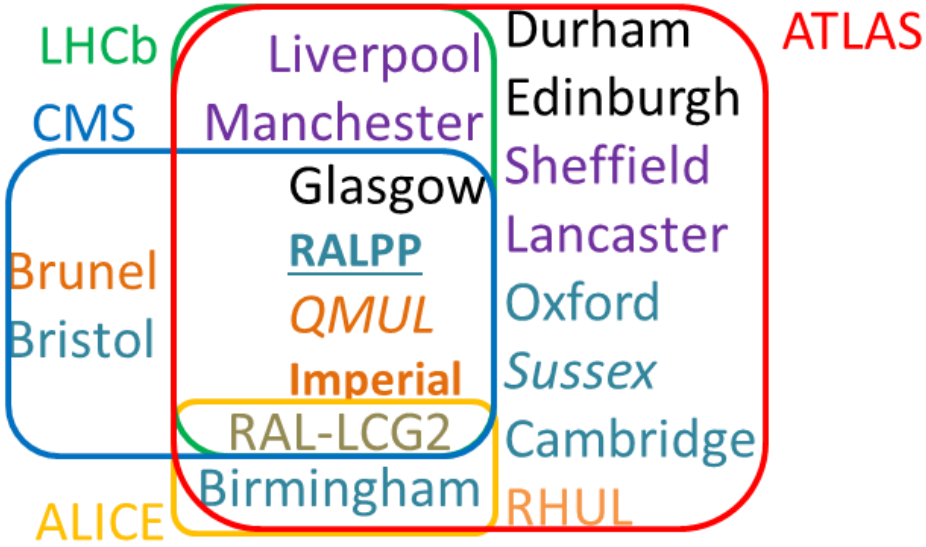
Contents

- Storage
  - Setups
    - Venn Diagram
      - IPV6 readiness
  - Transfer Rates
    - Rates between sites
    - Transfers
    - Closeness
      - Plot
      - Table and Location
  - Deletion Rates
  - Echo results
  - Gfal-* Commands
- Networking
  - RTT between sites
    - Effect on WN traffic speed
  - IPV6 in perfSONAR
    - Roadmap for 4.1
    - Comparison between ipv4 and ipv6 results between sites
    - Mddash results

Science & Technology Facilities Council

Storage: Setups

- Storage
  - Setups
    - IPV6 readiness?
      - » Storage Vs network ready
      - » Icmp vs tcp traceroutes(MD to cover)
      - » https://www.gridpp.ac.uk/wiki/IPv6_site_status
      - » But lets not waste effort. (Do "I" need to update my SE?)



Networking: ATLAS Closeness

- Using current max data rates to determine if site is good.
- https://twiki.cern.ch/twiki/bin/view/AtlasComputing/DDMNetworkMetrics
  - How much data has been transferred in any particular one hour window in the last month
  - Small is good
    - Needs finer grade values and to be verified

		To	To	To	To	To	To	To	To	To	To	To	To	To	To	To	To
		RAL Castor	**RAL-echo**	B'ham	**rhul**	**qmul**	**glasgow**	**lancaster**	**edinburgh**	**Mancs**	liverpool	sheffield	ralpp	SUSX	Cambridge	Oxford	durham
From	**RAL Castor**		3.00	3.00	4.00	3.00	3.00	3.00	4.00	3.00	4.00	4.00	4.00	5.00	4.00	4.00	4.00
From	**RAL-echo**	4.00		5.00	5.00	4.00	4.00	4.00	5.00	4.00	5.00	6.00	6.00	6.00	4.00	5.00	6.00
From	B'ham	5.00	6.00		5.00	5.00	5.00	5.00	5.00	5.00	5.00	6.00	6.00	5.00	4.00	6.00	11.00
From	**rhul**	3.00	5.00	2.00		4.00	5.00	4.00	3.00	4.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
From	**qmul**	3.00	3.00	4.00	2.00		4.00	4.00	3.00	4.00	4.00	5.00	4.00	4.00	5.00	4.00	5.00
From	**glasgow**	3.00	4.00	3.00	2.00	4.00		4.00	4.00	4.00	4.00	4.00	4.00	5.00	5.00	5.00	6.00
From	**lancaster**	3.00	3.00	2.00	4.00	4.00	4.00		5.00	4.00	3.00	5.00	4.00	5.00	5.00	4.00	5.00
From	**edinburgh**	4.00	4.00	5.00	5.00	4.00	4.00	5.00		5.00	5.00	6.00	5.00	5.00	5.00	6.00	11.00
From	**Mancs**	3.00	4.00	4.00	2.00	4.00	4.00	4.00	4.00		4.00	4.00	4.00	5.00	5.00	4.00	6.00
From	liverpool	4.00	4.00	5.00	5.00	4.00	5.00	5.00	5.00	5.00		5.00	5.00	5.00	5.00	6.00	6.00
From	sheffield	4.00	5.00	3.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00		5.00	5.00	5.00	6.00	6.00
From	ralpp	4.00	5.00	5.00	5.00	4.00	5.00	5.00	5.00	4.00	5.00	5.00		5.00	5.00	5.00	6.00
From	SUSX	5.00	6.00	6.00	6.00	5.00	5.00	5.00	6.00	5.00	5.00	6.00	5.00		5.00	6.00	11.00
From	Cambridge	4.00	5.00	5.00	3.00	4.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00		6.00	11.00
From	Oxford	5.00	5.00	5.00	5.00	4.00	5.00	5.00	4.00	4.00	5.00	5.00	6.00	11.00	5.00		6.00
From	durham	5.00	6.00	11.00	6.00	5.00	6.00	6.00	11.00	8.00	6.00	6.00	11.00	11.00	11.00	11.00	

Networking: Closeness Cont.

- What about other Communities
- Are values believable.?
  - "2"=10GB/s "3" = 1GB/s
  - Catch22 due to volume issues. Needs to be mitigated against.

From T1 to T2s	JINR	pic	Taiwan	RRC	NIKEF	NDGF	Triumf	AL-ECH	RAL	SARA	infn	n2p3	fzk	cern	bnl
RAL	5	4	3	3	3	3	2	4	0	3	3	3	3	2	2
RAL-echo	5	4	3	4	3	3	3	0	3	3	3	3	3	3	3
Mancs	5	4	4	3	4	3	3	4	3	3	3	3	3	3	3
lancaster	5	4	4	4	4	3	3	4	3	3	3	3	3	2	3
qmul	5	4	4	4	4	3	3	4	3	3	3	3	3	3	2
glasgow	5	4	4	4	4	3	2	4	3	3	3	3	3	2	3
rhul	5	4	4	4	4	2	4	5	4	3	4	3	4	3	3
edinburgh	5	5	4	4	4	4	4	5	4	2	4	4	3	3	3
ralpp	5	4	4	4	4	3	4	6	4	4	4	4	4	3	3
liverpool	5	4	5	4	4	4	4	5	4	4	4	4	4	3	3
Oxford	6	4	4	3	4	4	4	5	4	4	4	3	4	3	3
sheffield	6	5	5	6	5	4	4	6	4	5	4	4	4	4	4
B'ham	5	5	5	5	5	3	5	5	3	5	5	4	4	4	4
Cambridge	4	4	4	5	5	4	5	4	4	5	5	4	4	4	4
SUSX	5	5	5	5	5	5	5	6	5	5	5	5	5	4	5
durham	11	5	5	4	5	5	4	6	4	5	5	5	5	5	5

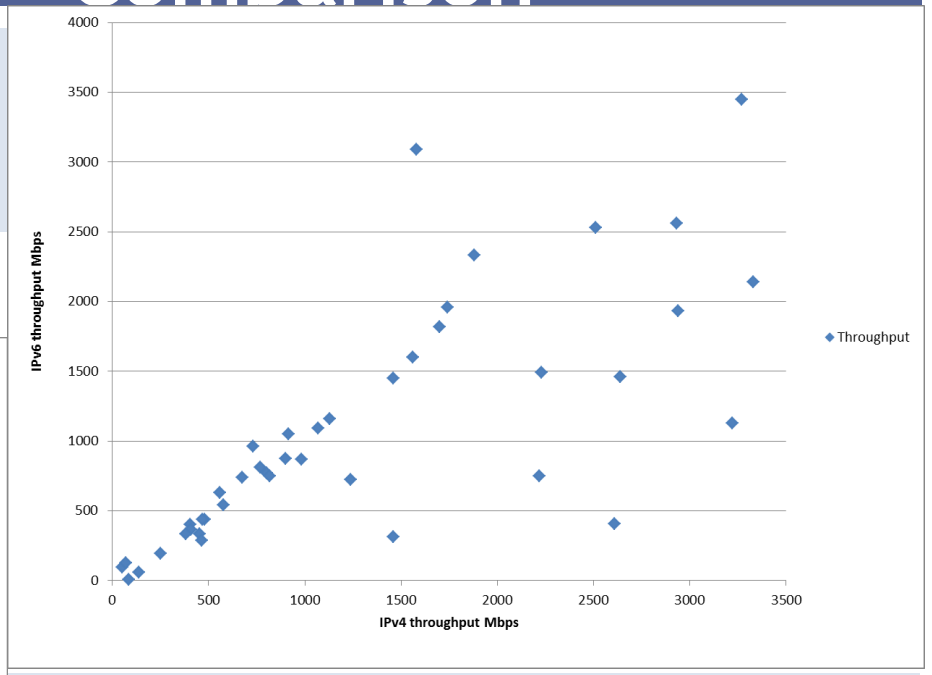
Networking: RTT

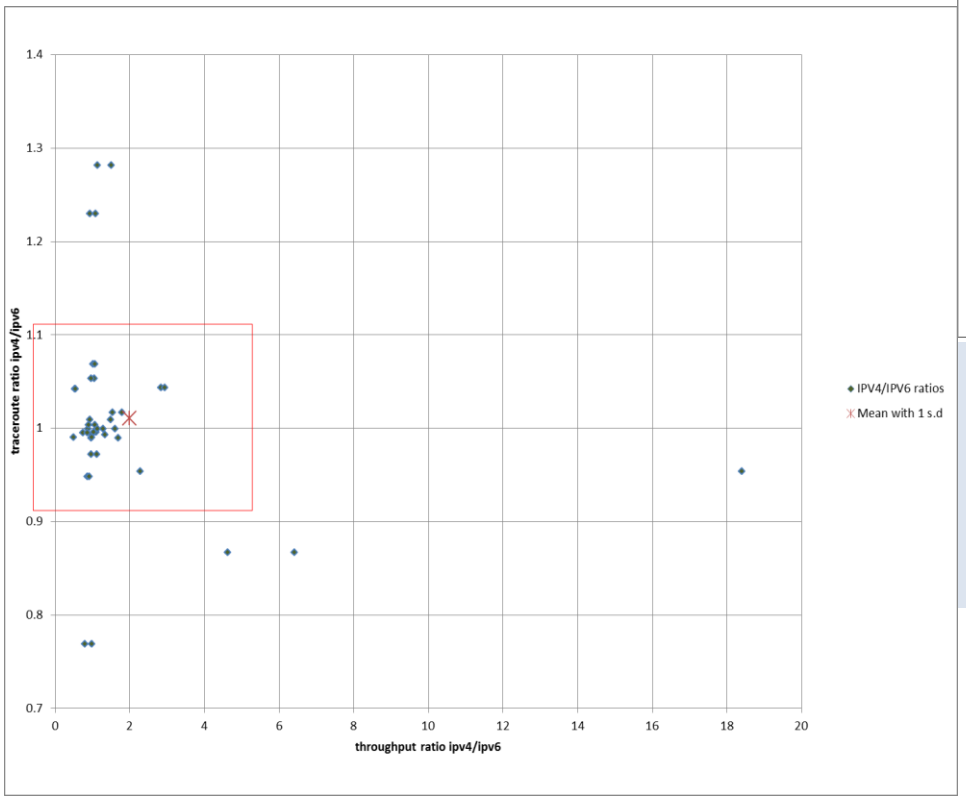
- Networking
  - RTT between sites
    - Effect on WN traffic speed
    - Throughput important for copying, rtt important for streaming

			lancaster	glasgow	dinburgh	durham	sheffiled	liverpool	Mancs	Cam'br	B'ham	qmul	IC-HEP	rhul	ristol	RAL	Oxford	SUSX	runel	UCL
fal-pygrid-30.lancs.ac.uk	4	lancaster		6.52	13.8	13.6	13.7	9.59	10.2	14.6	12.4	13.3	16.1	17	18	14.9	18.2	20.3		
svr018.gla.scotgrid.ac.uk	1	glasgow				8.38		7.94	7.92	10	9.24	11.4		13		13.1	14.5	14.6		
srm.glite.ecdf.ed.ac.uk	2	edinburgh				7.1	5.5	9.04	7.05	8.39	7.82	9.24	12.8	14	16	14.9	14.8	17.11		
se01.dur.scotgrid.ac.uk	3	durham					4.68	7.56	6.39	7.5	6.99	8.32	10.6	11	14	12.2	13.6			
lcgse0.shef.ac.uk	5	sheffield						5.98	4.39	5.85	5.33	6.76		11		11	11.8			
hepgrid11.ph.liv.ac.uk	6	liverpool							5	8.57	4.72	9.51	8.42	9.5	12	11.8	10.6	12.6		
bohr3226.tier2.hep.manchester.ac.uk	7	Mancs								6.98	4.78	7.57	8.59	9.6	12	10	10.6	12.4		
serv02.hep.phy.cam.ac.uk	11	Cambridge									4.78	3.61	8.5	7.2	12	10.1	10.7	12.4		
epgse1.ph.bham.ac.uk	8	B'ham										5.65	5.36	5.9	8.2	5.4	6.82	9.08		
se03.esc.qmul.ac.uk	15	qmul											1.18	2.9	4.4	2.84	3.38	5.52	2.3	
gfe02.grid.hep.ph.ic.ac.uk	17	IC-HEP												2			2.76			
se2.ppgrid1.rhul.ac.uk	14	rhul													5.1		4.2	6.13		1.09
lcgnetmon02.phy.bris.ac.uk	9	Bristol														5.4	5.98			
srm-atlas.gridpp.rl.ac.uk	13	RAL															4.3			
t2se01.physics.ox.ac.uk	10	Oxford																7.71		
grid-storm-02.hpc.susx.ac.uk	19	SUSX																		
dc2-grid-64.brunel.ac.uk	16	brunel																		
	18	UCL																		

PerfSONAR ipV4/IPV6 Comparison

- IPv4 vs IPv6 Throughput measurements

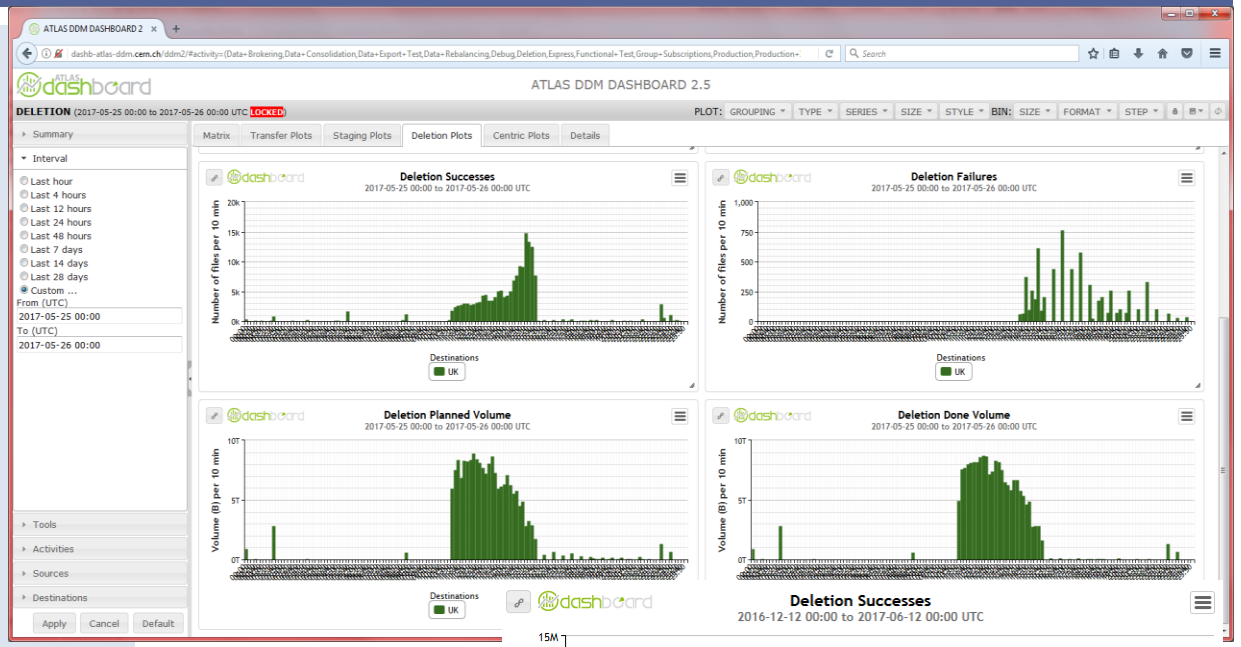


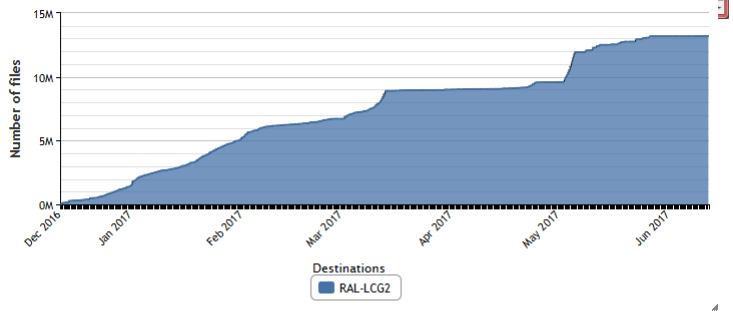


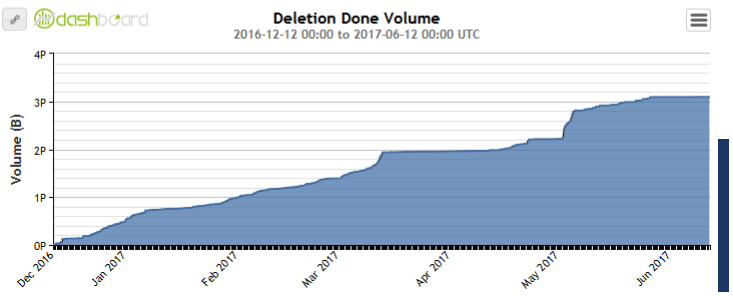
- Traceroute ipv4/ipv6 vs Throughput ipv4/ipv6 ratio
  - Small VOs
    - What/how to do?

Storage: Deletion Rates

- Storage
  - Deletion Rates
  - Lack of non-ATLAS
  - Churn rate per site.
    - 5.8PB datadisk at RAL
    - >3.1PB deleted in 6Mths
    - 1.3M files
    - Echo deletes at 25Hz
      - 10 x Castor rate

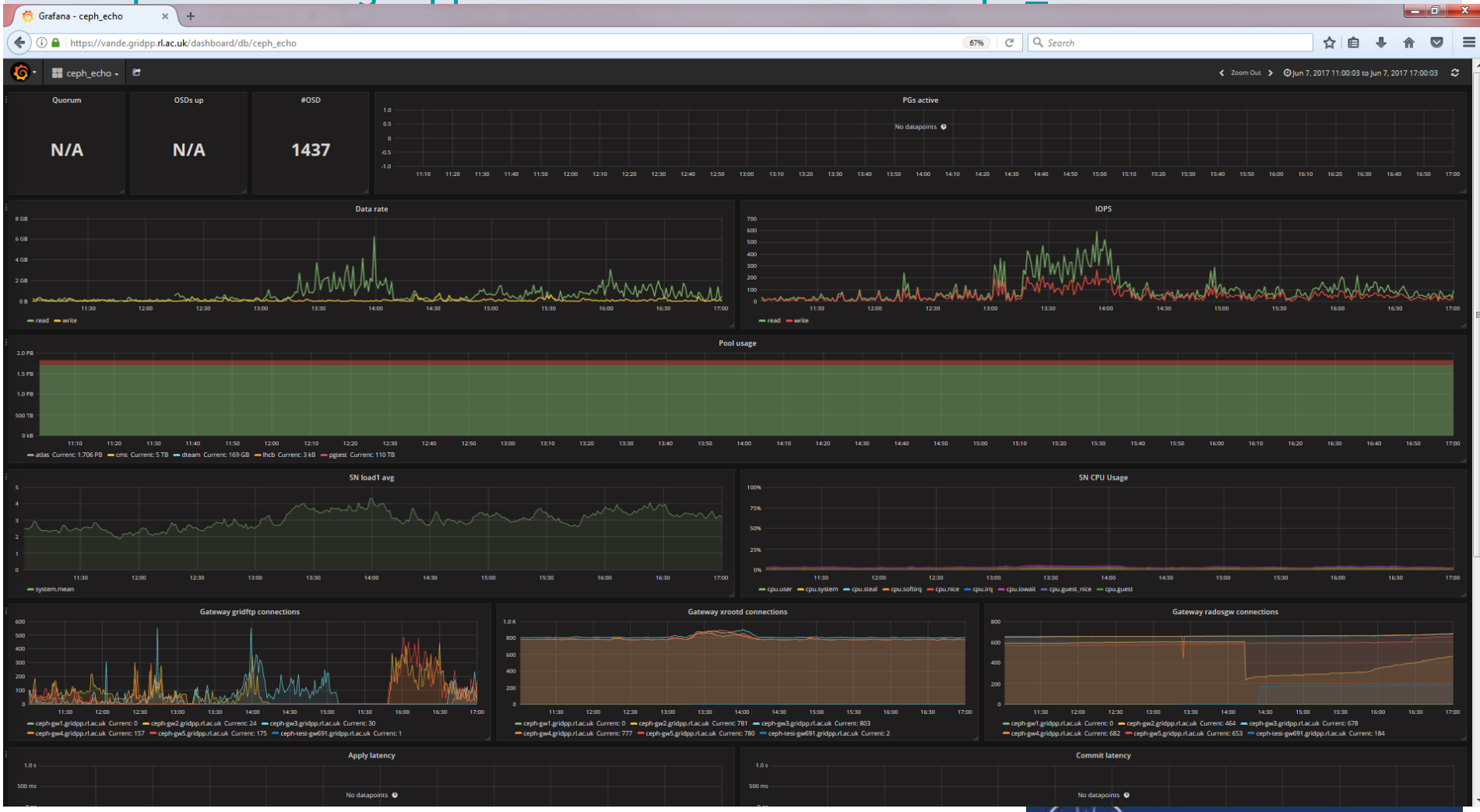




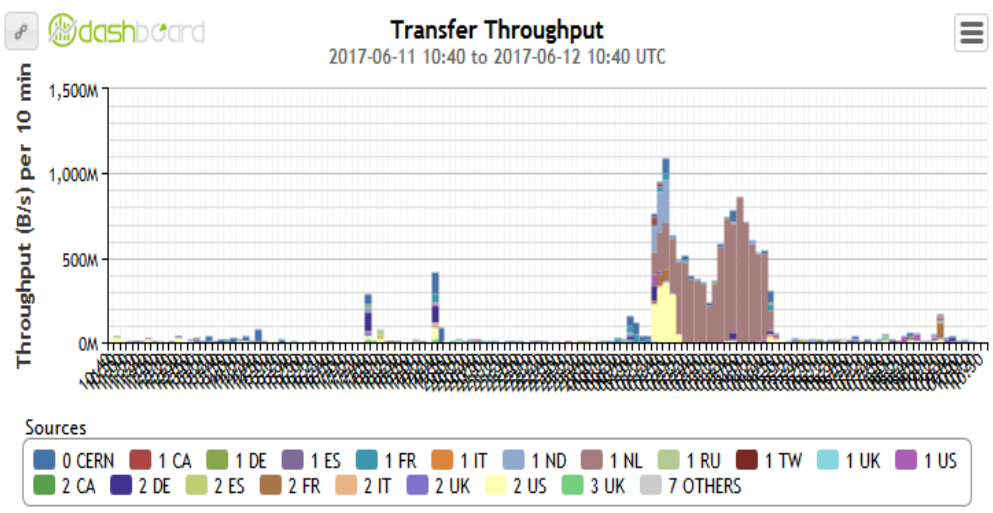


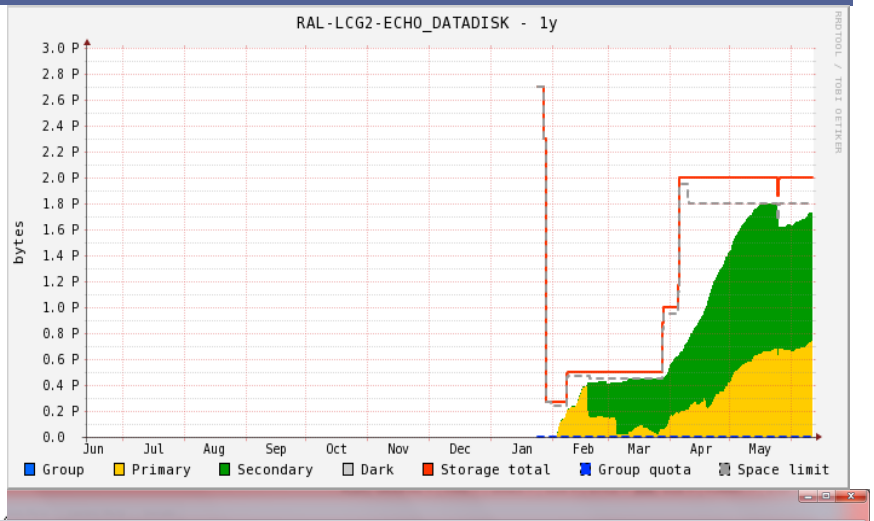
Storage: Echo Results

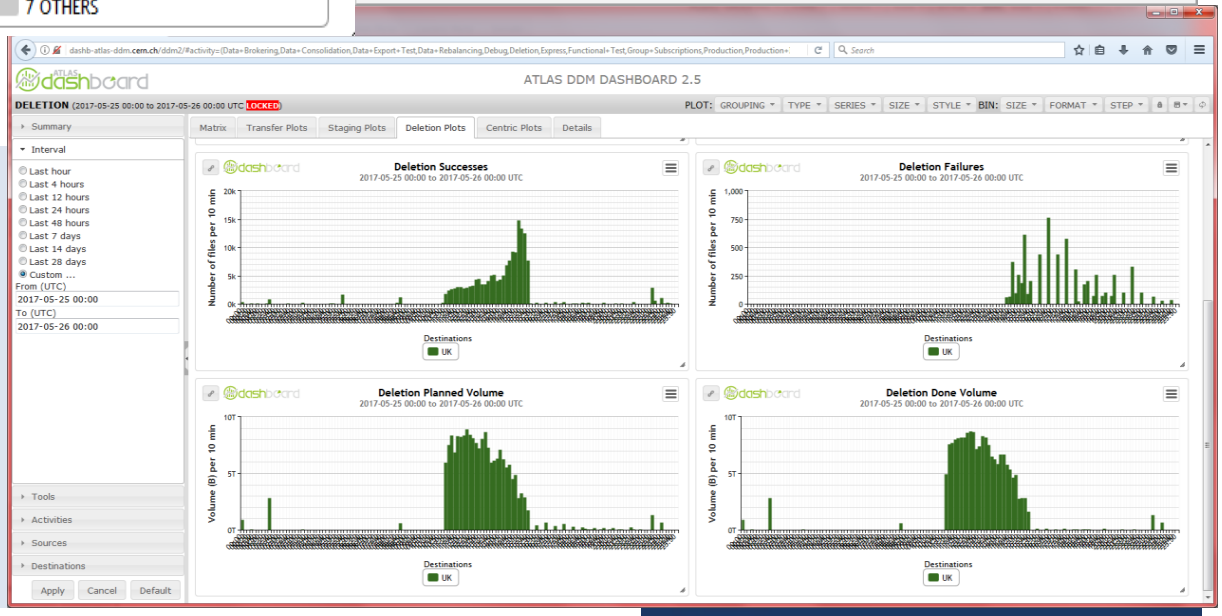
- https://vande.gridpp.rl.ac.uk/dashboard/db/ceph_echo



Storage: Echo Results



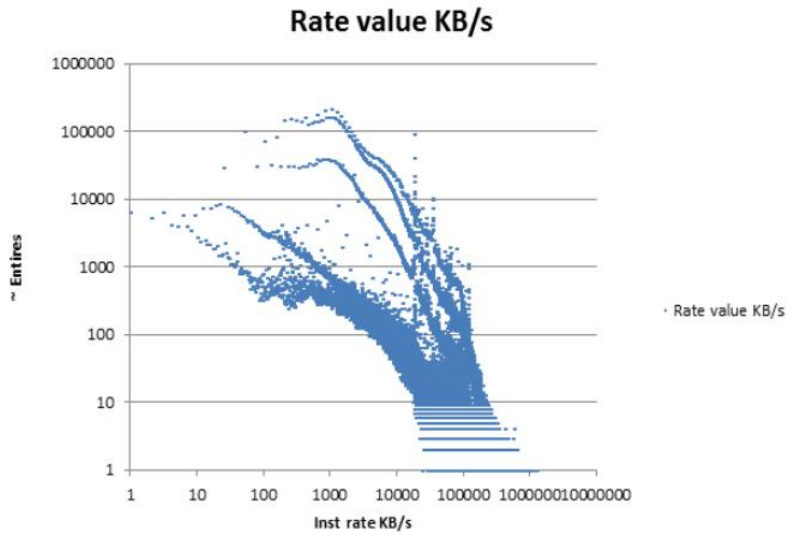


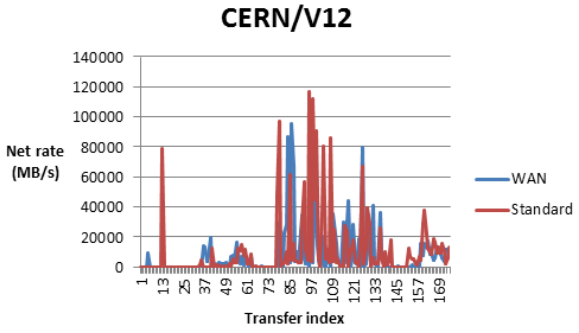


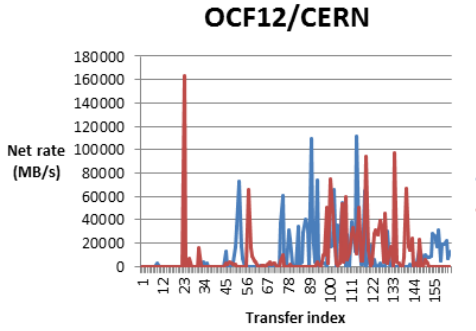
- Object stores the Future?
- Gridftp/S3/xrootd
- Ipv6/xroot cache/ GWs on WNs
- Write Rate/ Deletion rates:

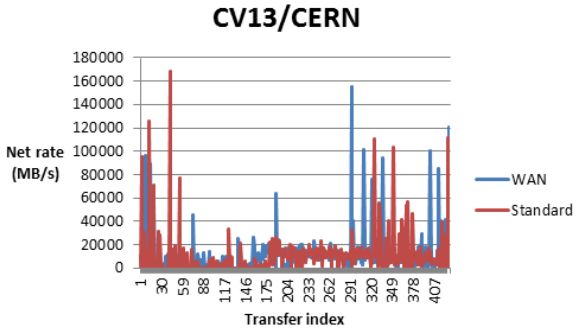
The Kraken and Disney:

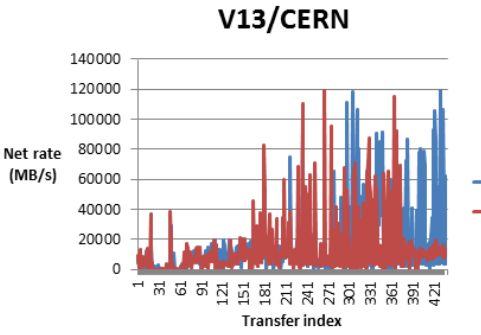
- Instantaneous rate inside gridftp within FTS transfers.
- For all transfers and those for RAL
- Difference in WAN tunings
  - To be investigated

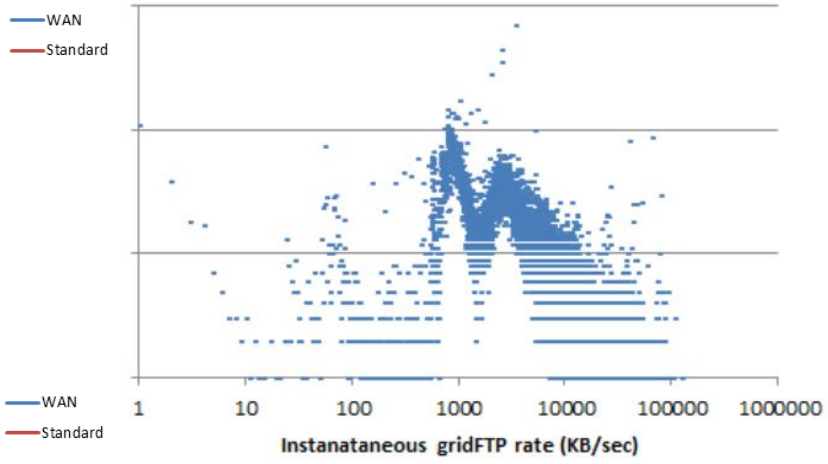










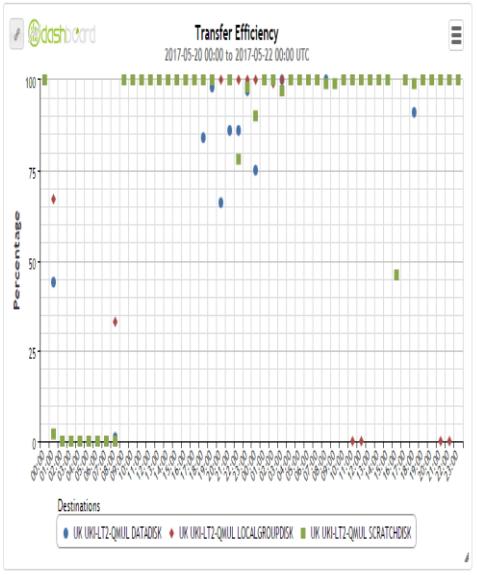


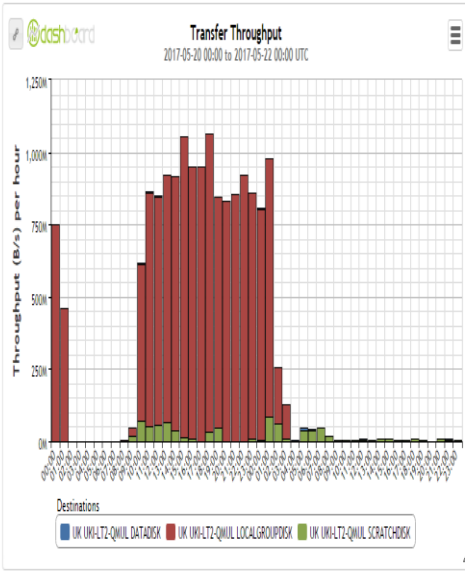
Storage: gfal-* commands

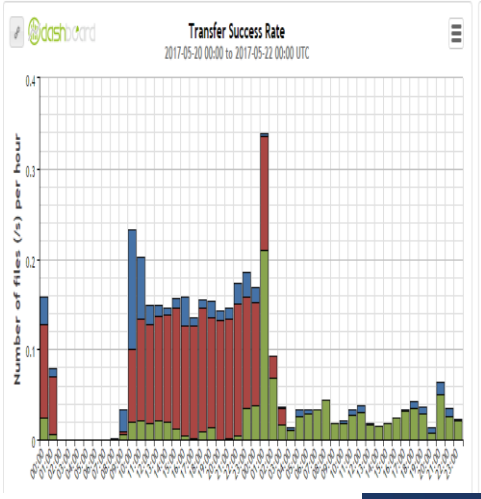
- Storage
  - Gfal-* Commands
  - gfal-copy, gfal-ls, gfal-rm, gfal-cat, gfal-mkdir, gfal-chmod, gfal-rename
    - do what you expect.
  - gfal-legacy-replicas, gfal-legacy-unregister, gfal-legacy-bringonline, gfal-legacy-register
  - gfal-save could be trouble command….
  - gfal-stat
  - gfal-sum (checksum of file but is it shallow or deep?)
  - gfal-xattr
  - What version do you have installed?
  - Does the VO , user even use this version?
    - ATLAS import from cvmfs their own version
- What alternative options for copying data could/are people using.
  - rsync, wget, curl, GlobusOnline, cp

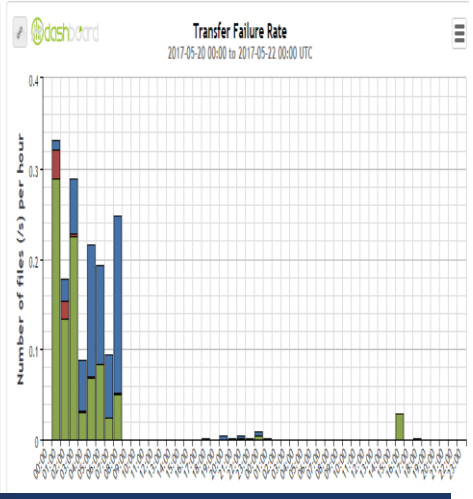
Site Evolutions

- T2s can now get their network filled.
  - Protection needed for rest of site/university?
    - More services outside university firewalls?
  - Is there a need to increase bandwidth?
  - How much needed for Wide Area Worker Nodes (WAWN)
  - How many connections?
    - Hardware Choice CPU/RAM
    - Protocol Dependency









Site Evolution

- How do our sites fir into ADs list
  - Full—lots of cpu and disk more than adequate for caching needs, possibly able to support WAN access as well as LAN
  - CPU rich—i.e. some disk, but probably only enough to act as a local cache
  - Disk rich—if we imagine disk is still seen as hard to manage then maybe some T2s will specialise in this and maybe have less CPU than the average or way more disk. This could also be a configuration for an individual site in a distributed T2.
  - Disk poor—lower disk/cpu ratio than the CPU rich site
  - Diskless—is this feasible for a standalone T2? For a site that is part of a distributed T2
- Similar criteria for other VOs?

Site Evolution relative components

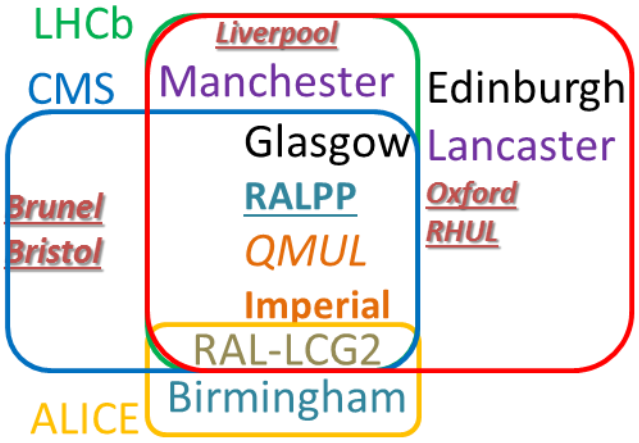
- Ratio of resources VO needs:
- CMS ~2GB/s per 1k cores
- How much do each component cost:
  - £/TB £/Gbps £/HepSpec
- Equivalent capacities
  - 10Gbps~ 30PB/yr ~ 500Cores~10 Machines
- How much storage is needed per CPU
- Can we find out real use values for our sites?

Site Evolution

- Who actually wants to run storage?
  - And if so just WLCG VOs or also for/just other communities local to SE.
    - Funding issues.
  - Or just be Diskless site
- If So, will WLCG VOs use it.
  - Who to go SRMless. Who needs to push IPV6, Caching.
- Are large sites ready for extra WAN traffic from WAWNs
  - Which of the plethora of options to go for.
    - Site Choice
    - VO Choice
- What method to setup storage.
  - ARC/xrootd cache ,federated storage systems

Site Evolution

- Where we get to (V1.1)
  - My vision if not others!!



Summary/Discussion output:

- Storage Summary
- Networking Summary