GridPP

UK Computing for Particle Physics

GridPP preparing for Run2… *and the wider context*

Version 1.0

CHEP 2015 – Okinawa, Japan
14th April 2015

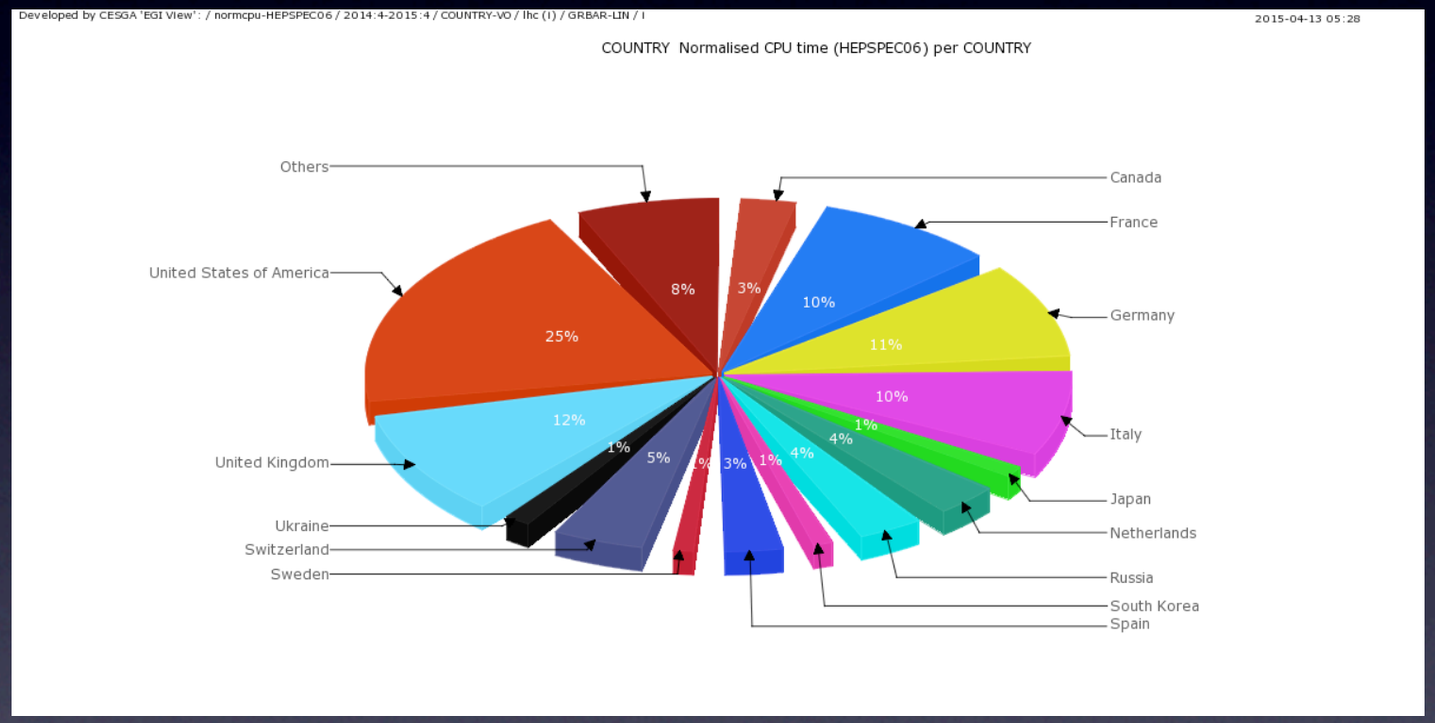
Jeremy Coles

UNIVERSITY OF CAMBRIDGE

Overview

10 minutes, 10 topics.

- Resource considerations
- Batch and CE evolution
- Multicore optimisation
- Worker node integration
- Reusing tools
- Infrastructure tasking
- IPv6 progress (may skip)
- RIPE ATLAS probes
- Wider collaboration

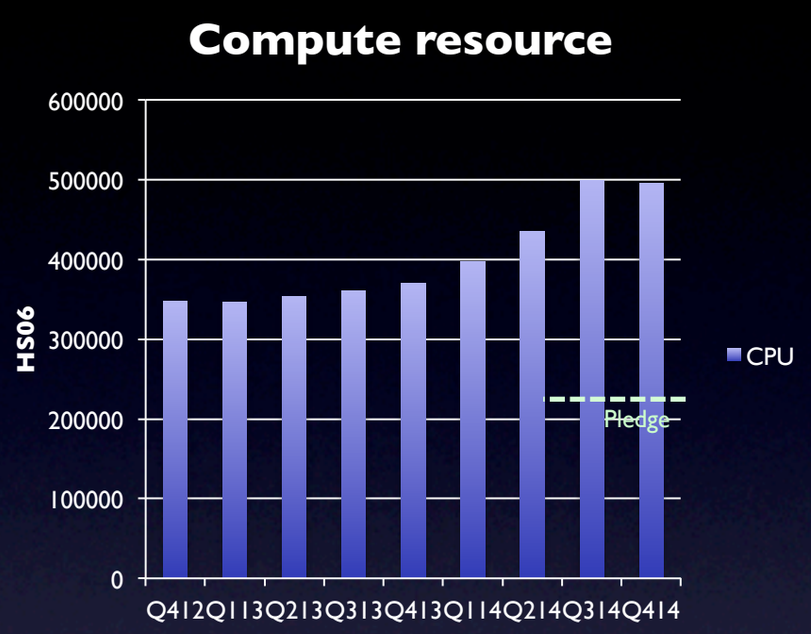


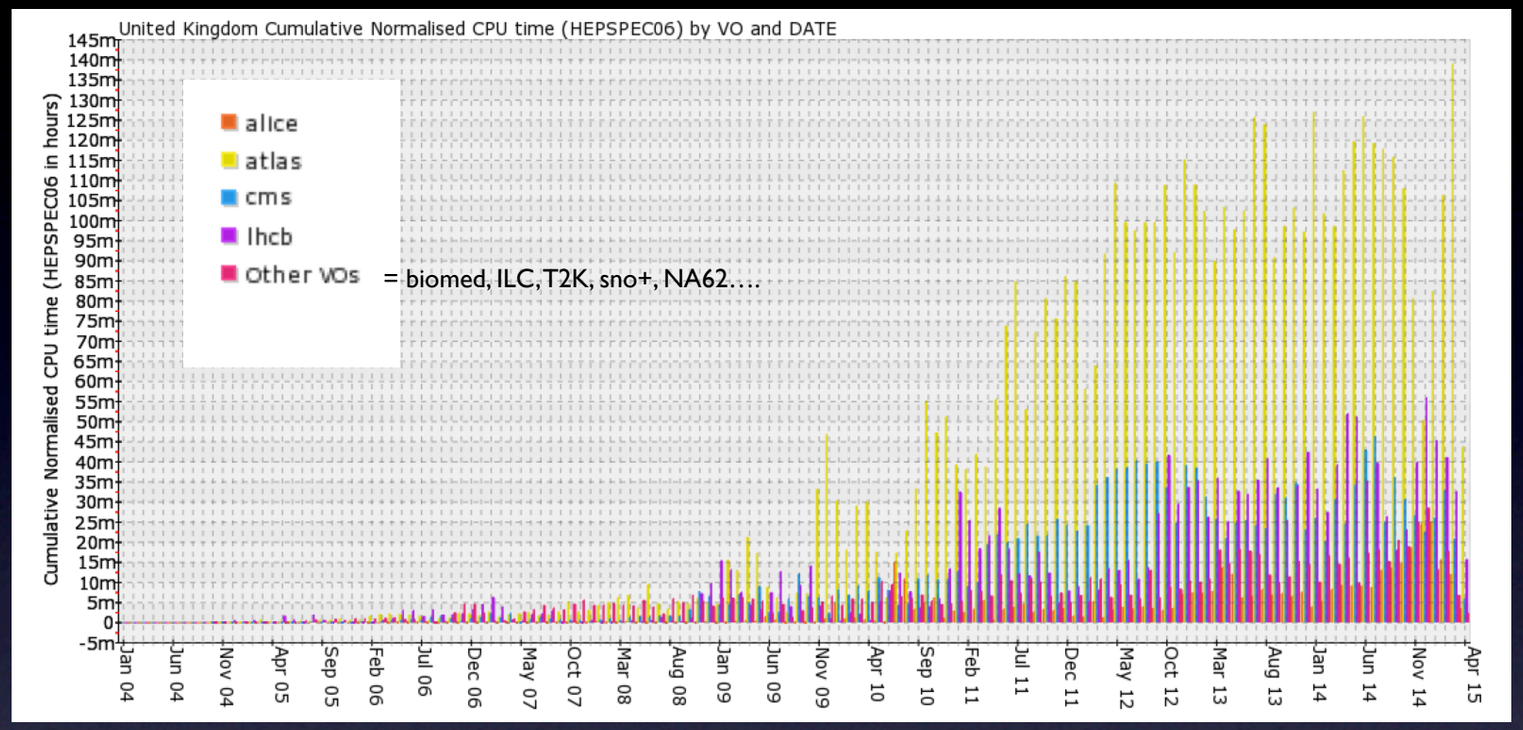
Context: GridPP supplies 10-12% of WLCG resources

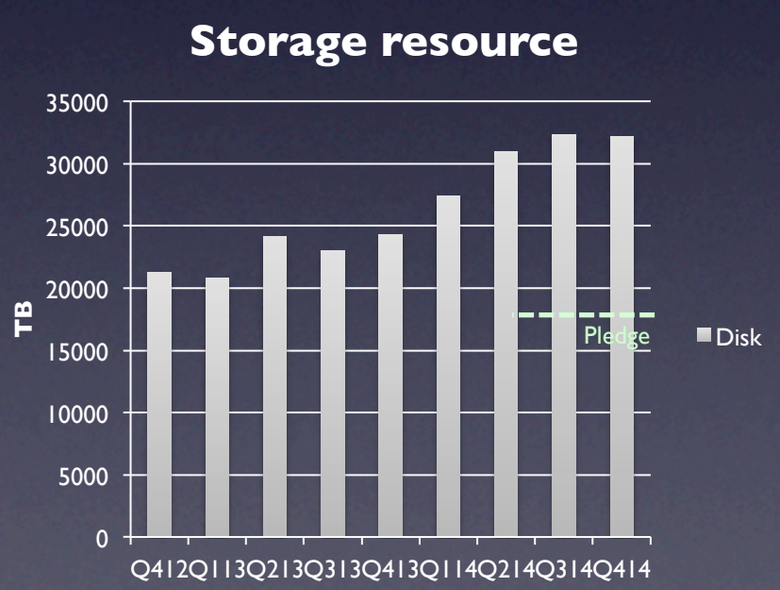
With thanks to many people who provided content….

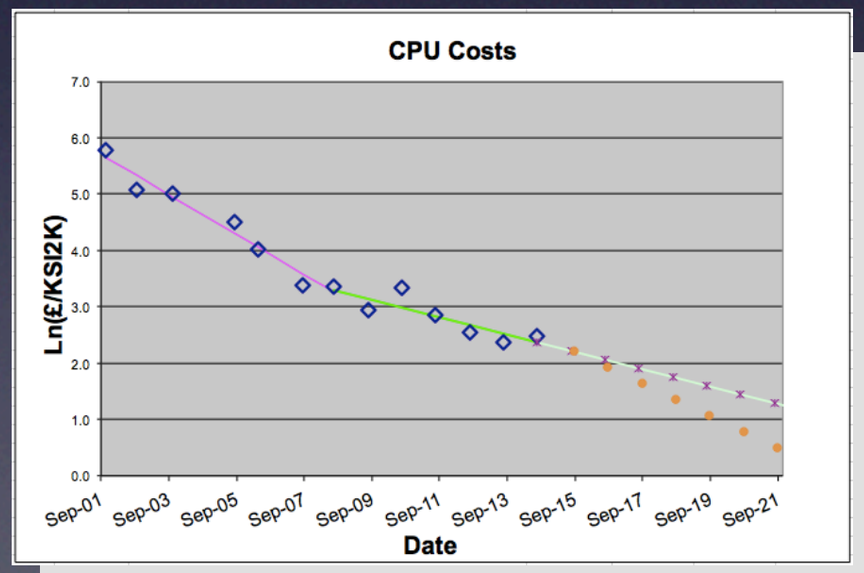
Text like this gives details of connected CHEP talks

Hardware fine as Run-2 begins









RAL T1 figures (A Sansum, D Britton)

- Historically we have leveraged additional capacity.
- The 2013 CPU cost projections start to look too optimistic.

Batch and CE evolution during LS1

Site	Current product (local/shared)	Concerns and observations	Interest/Investigating/Testing	CE type(s) & plans at site	cgroups	Multi core Atlas/CMS
RAL Tier-1	HTCondor (local)	None	No reason	ARC	Yes	Yes
UKI-LT2-Brunel	Torque/Maui, Arc/Condor	No support for Torque/Maui	Slurm and HTCondor in test	Arc in test	Yes	Yes
UKI-LT2-IC-HEP	Gridengine (local)	None	No reason	CREAM, ARC	No	Yes
UKI-LT2-QMUL	Gridengine (local)	None	Son of Gridengine	CREAM	No	Yes
UKI-LT2-RHUL	Torque/Maui (local)	Torque/Maui support non-existent	Will follow the consensus	CREAM	No	Yes
UKI-LT2-UCL-HEP	Torque/Maui (local)	Torque/Maui support non-existent	HTCondor	CREAM CE	No	X
UKI-NORTHGRID-LANCS-HEP	Son of Gridengine (HEC), torque/maui (local)	Disillusioned with torque/maui.	Slurm or HTCondor.	CREAM, interested in ARC	No	Yes
UKI-NORTHGRID-LIV-HEP (Single core cluster)	Torque Maui (local)	Poor Support, Maui intrinsically broken		Cream	No	No
UKI-NORTHGRID-LIV-HEP (Multi core cluster)	HTCondor (local)	None		ARC	Loooking into it	Yes
UKI-NORTHGRID-MAN-HEP	Torque/Maui (local)	Maui patch t wrote a		Currently CREAM, investigating ARC-CE	No	Yes
UKI-NORTHGRID-SHEF-HEP	Torque/Maui (local)	Torqu		CREAM CE, ACR CE is in test	No	Yes
UKI-SCOTGRID-DURHAM	SLURM (local)			ARC CE	No	Yes
UKI-SCOTGRID-ECDF	Gridengine	None	No reason	Cream CE for standard production, ARC CE for exploratory HPC work	No	Yes
UKI-SCOTGRID-GLASGOW	HTcondor (local), Torque/Maui (local)	Becomes unresponsive at times of high load or nodes being un-contactable.	Investigating HTCondor/SoGE/SLURM as a replacement.	ARC, Cream	Yes	Yes
UKI-SOUTHGRID-BHAM-HEP	Torque/Maui	Maui sometimes fails to see new jobs and so nothing is scheduled	HTCondor	CREAM	No	No
UKI-SOUTHGRID-BRIS	HTCondor (shared), torque + maui (local)	None	No reason	ARC & CREAM CEs	No	X
UKI-SOUTHGRID-CAM-HEP	Torque/Maui (local)	Torque/Maui support non-existent	Will follow the consensus	CREAM CE	No	Yes
UKI-SOUTHGRID-OX-HEP	HTCondor (local)	None	No reason	ARC CE in production	Yes	Yes
UKI-SOUTHGRID-RALPP	HTCondor	None	No reason	ARC CE	Yes	Yes
UKI-SOUTHGRID-SUSX	(Shared) Gridengine - (Univa Grid Engine)	None	No reason	CREAMCE	Looking into it	Yes

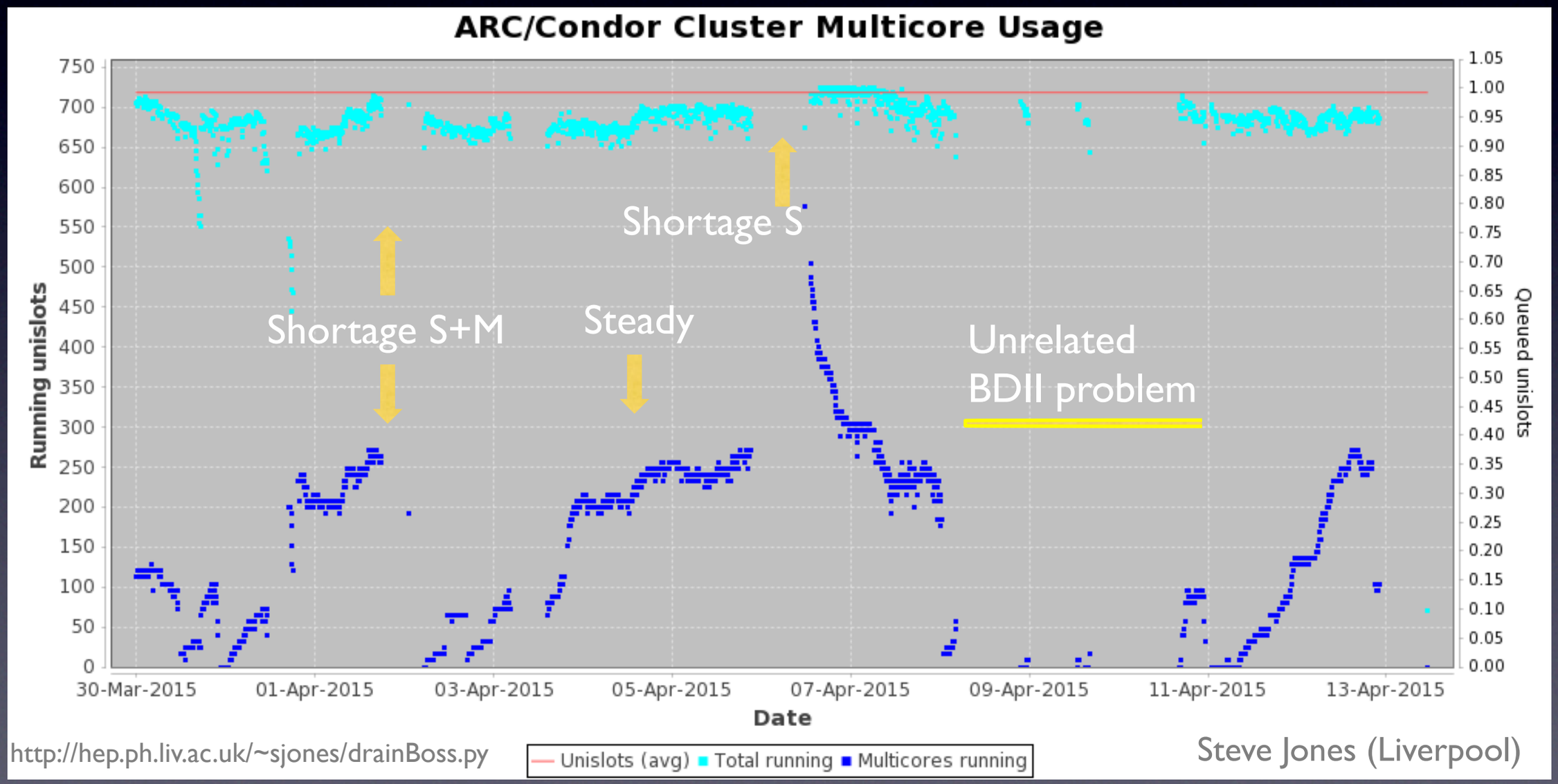
HTC – 7
GE – 5
Torque/Maui - 9
SLURM - 1

ARC CE – 10
CREAM CE -12

Drivers: Simplification. Stability. Features.

Efficient multicore - drainBoss

- Drain rate controller for ARC/HTCondor. Recent performance looks good.
- Senses condition of cluster and adjusts how nodes are drained (to allow 8-core slot)
- See HEPiX talk: https://indico.cern.ch/event/346931/session/5/contribution/55
- Note gaps (below) are due to job shortages. Prototype is a python script implementation.
- Target = 250 unislots when possible. (S=Single core jobs. M=Multicore jobs.)



http://hep.ph.liv.ac.uk/~sjones/drainBoss.py

Steve Jones (Liverpool)

Multicore in WLCG: 14/04 B250 15:30 – 15:35

Evolving integration approaches during LS1

Site	VM Provider	VM Lifecycle Manager	Experiment/VO					
			GridPP DIRAC	ALICE	ATLAS	CMS	LHCb	Ot
Birmingham	Vac	Vac	X					
Bristol								
Brunel	OpenVZ							
Cambridge								
CERN	OpenStack	Vcycle	X		X	X	X	
Durham								
Edinburgh								
Glasgow								
Imperial	OpenStack	CloudScheduler			X			
Imperial	OpenStack	Vcycle	X		X	X	X	
Lancaster	OpenStack (offsite)				X			
Lancaster	Vac	Vac	X		X	X	X	
Liverpool	Looking at Vac							
Manchester	Vac	Vac	X		X	X	X	
Oxford	OpenStack	CloudScheduler						
Oxford	Vac	Vac	X		X	X	X	
QMUL	CloudStack							
RAL PPD								
RAL Tier-1	HTCondor	Condor Vacuum	X		X	X	X	
RAL Tier-1	OpenNebula							
RHUL								

Drivers:

- Demonstrators
- Gain experience
- Expectations (Cloud)
- Flexibility (Cloud)
- Support (Cloud)
- Simplicity (VAC, HTC)

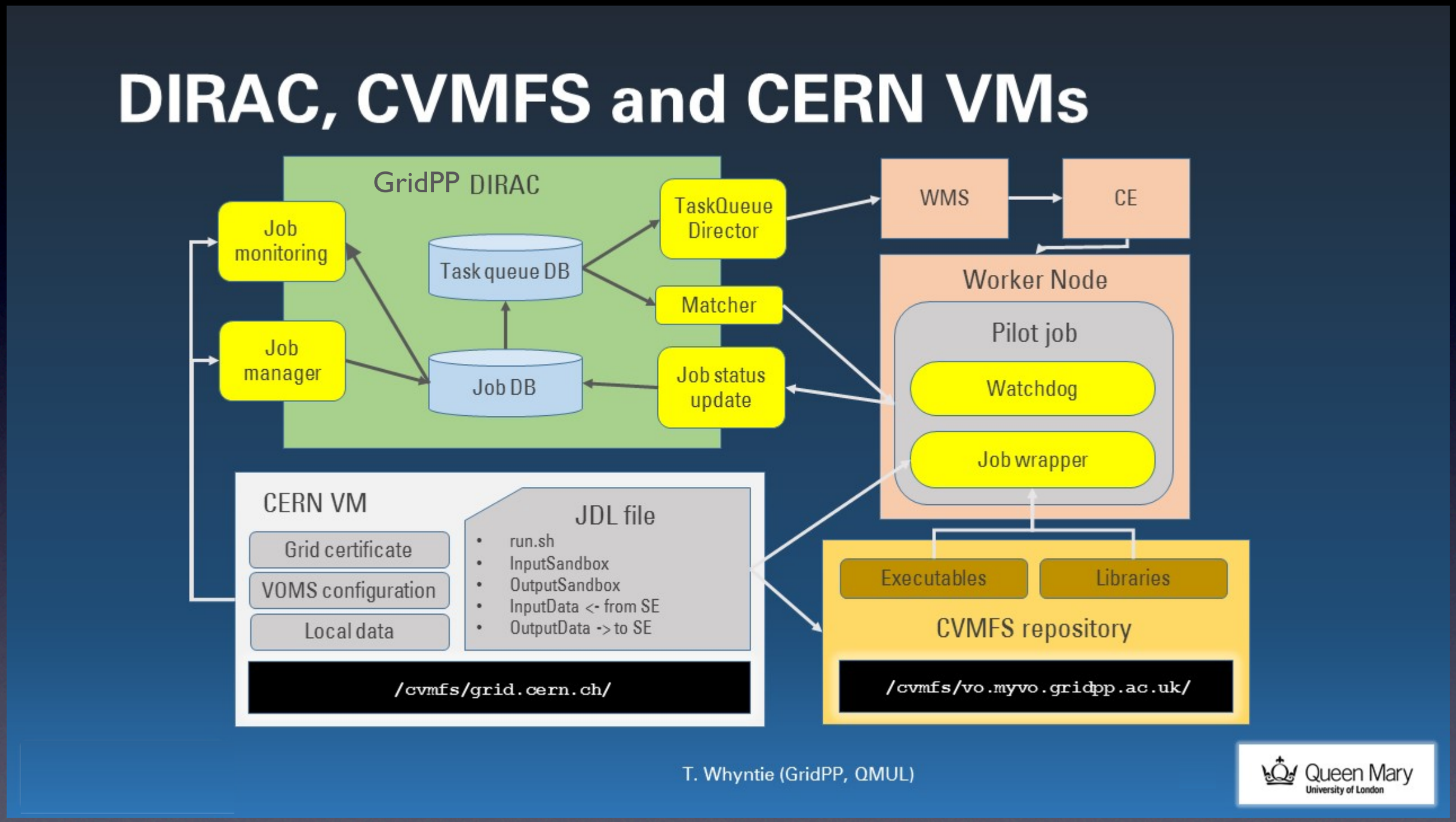
Outcomes:

- Work effectively. Optimisation.
- Ready to reposition capacity but this needs policy/strategic discussion.
- Containers look more efficient and similar management tools apply.

For VAC as a resource see this example: https://www.gridpp.ac.uk/wiki/Vac_configuration_for_GridPP_DIRAC

Vac/Vcycle VMs: 13/04 B210 17:00 - 17:15
Containers: 13/04 B210 17:15 – 17:35

Benefits for non-LHC VOs

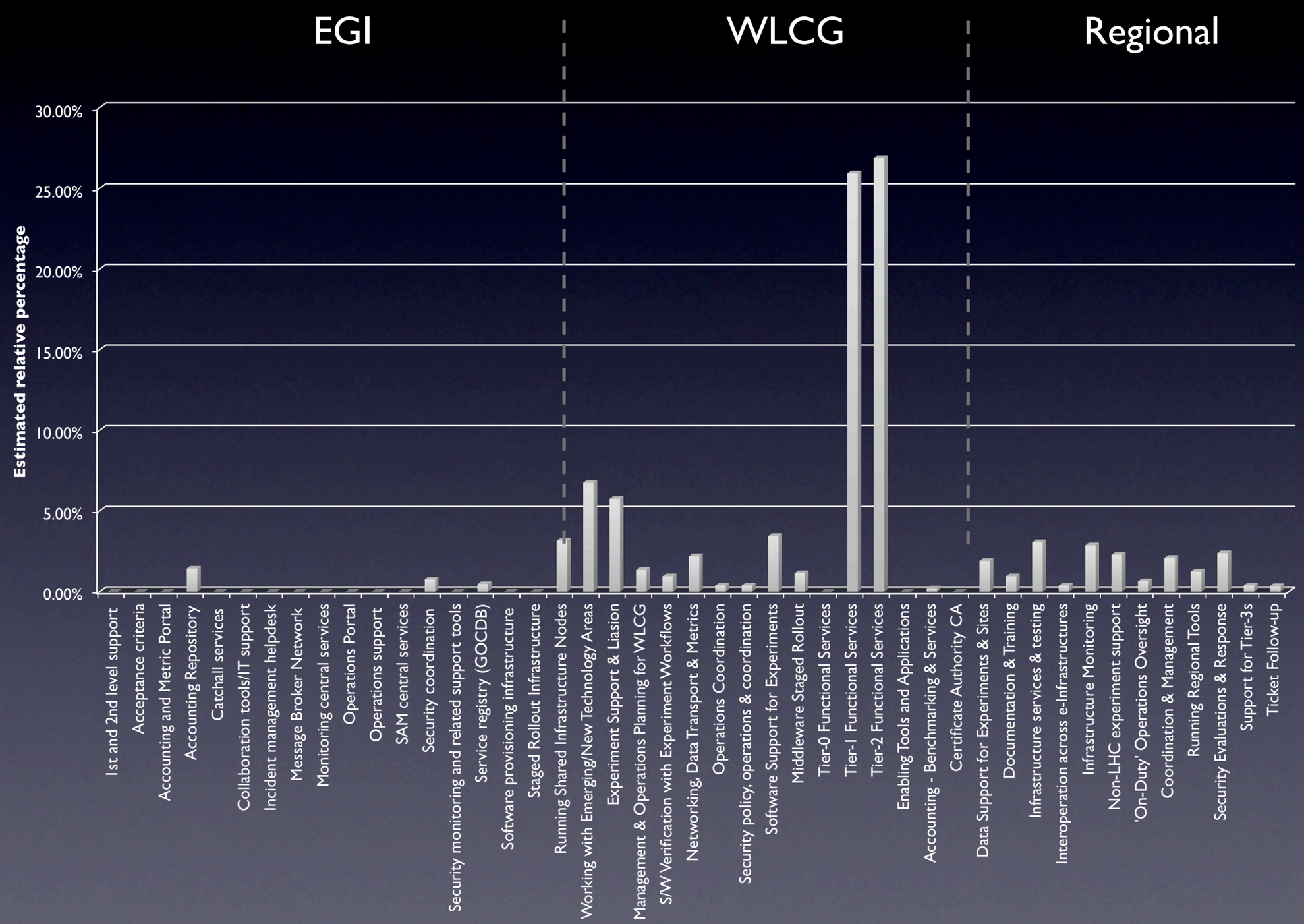


Case study cern@school: https://indico.cern.ch/event/350917/session/0/contribution/11/material/slides/0.pdf

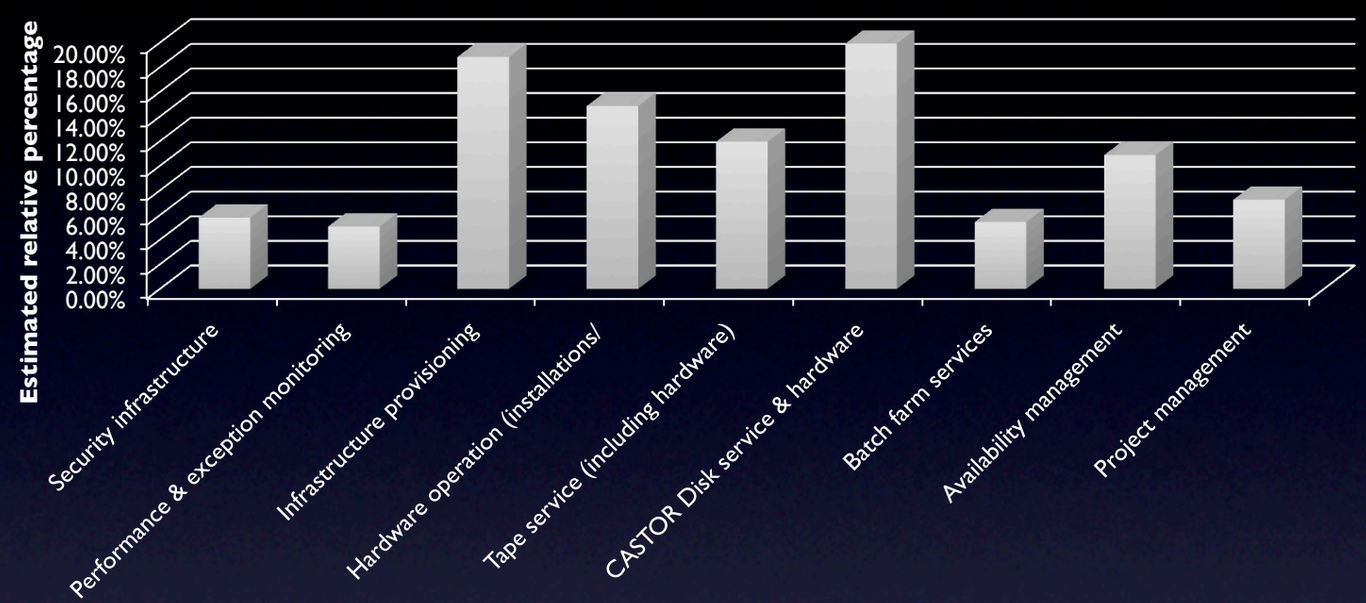
GridPP DIRAC project: 16/04 B250 12:00 - 12:15

Infrastructure provider tasks

Estimated effort distribution across tasks. More than just hardware!



Site functional service tasks – looking for savings



T1

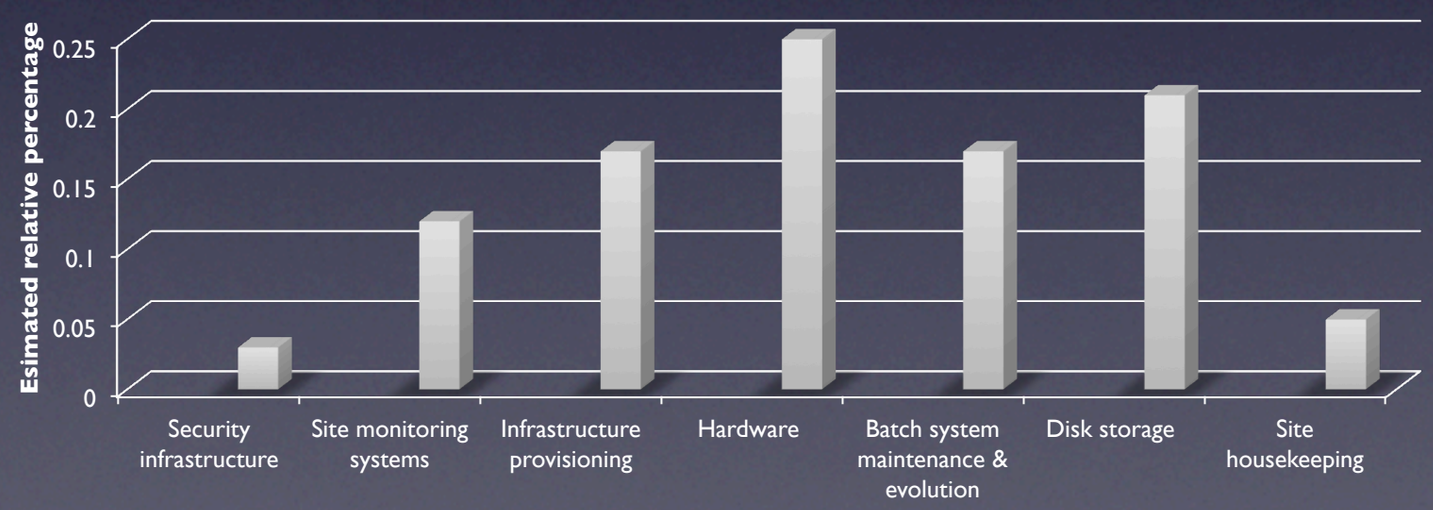
- Already improved batch
- Moved to VM based service setup
- Investing effort in CEPH and VM infrastructure
- Exploring sharing of costs (e.g. DiRAC - HPC)

Batch sharing experiences: Track 7. Poster Session B
Cloud sharing experiences: Track 7. Poster Session A

T2

- Individual effort is distributed across many tasks
- Average over T2 sites much less than 2.8FTE
- Cutting sites means additional leverage lost

- Looking at improving batch maintenance
- Options for remote disk for smallest sites
- Remote management so far not taken off but…
- Exploring VAC/Vcycle as lightweight model
- E.g. Vac-in-a-box!



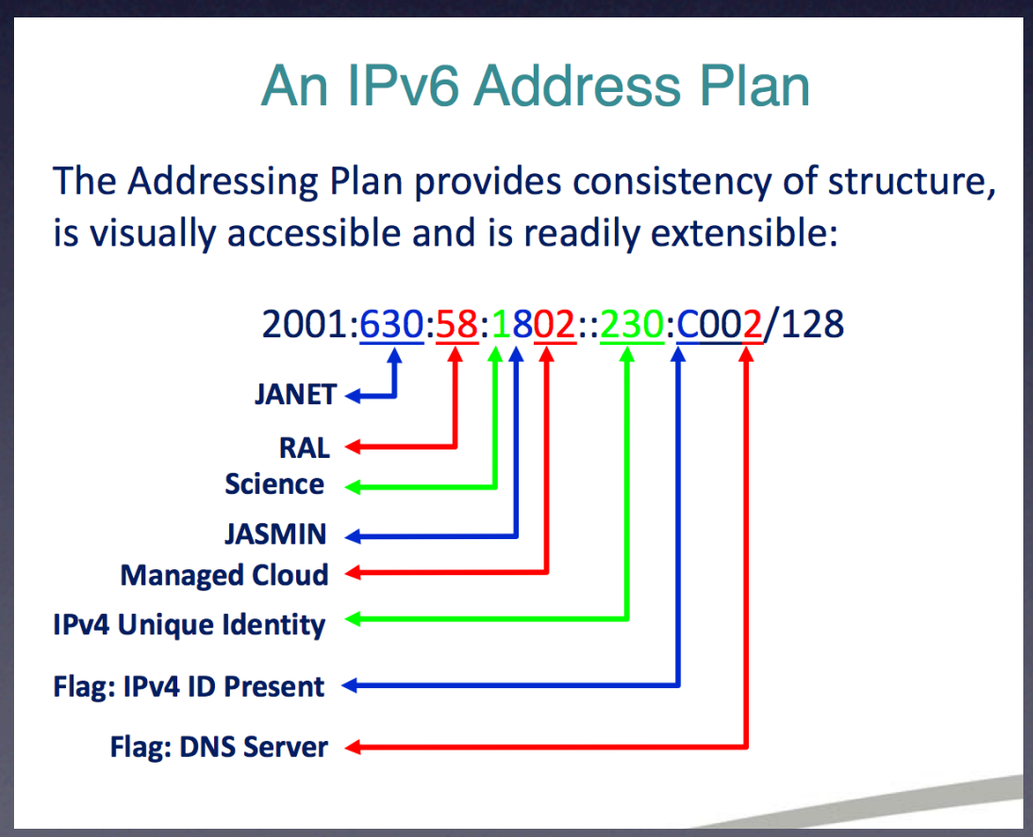
WLCG federation technologies: Track 3. Poster Session B

IPv6 – A core infrastructure concern

- Tracking institute requests/allocations
- Encouraging sites to try dual-stack
- Raising awareness of how address plans need careful thinking….

Supporting the testing of components

	Latest Jobs										Time of Last Job	
Site	10	9	8	7	6	5	4	3	2	1	Submitted	Successful
UKI-LT2-Brunel	S	S	S	S	S	S	S	S	C	C	15/03/15 03:00	15/03/15 01:00
UKI-LT2-IC-HEP	S	S	S	S	S	S	S	S	S	C	15/03/15 03:00	15/03/15 02:00
UKI-LT2-IC-HEP-V6-TESTBED	S	S	S	S	S	S	S	S	S	C	15/03/15 03:00	15/03/15 02:00
UKI-LT2-QMUL	F	F	F	F	F	F	F	F	F	C	15/03/15 03:00	14/09/14 14:20
UKI-SOUTHGRID-OX-HEP	S	S	S	S	S	S	S	S	S	C	15/03/15 03:00	15/03/15 02:00



	Storage Element						
CE	Imp_V6	Imp	QMUL	Ox	Brun	Ave	CE
Ox	69.5	62.0	75.1	104.9		78.4	Ox
Imp	95.3	82.5	62.6	53.4		75.5	Imp
Imp_V6	56.5	66.7	71.4	50.9		59.5	Imp_V6
Brun	53.2	40.4	39.4	17.5		38.0	Brun
QMUL							QMUL
Ave	69.7	64.2	64.2	59.7		64.5	Ave
	Imp_V6	Imp	QMUL	Ox	Brun	Ave	

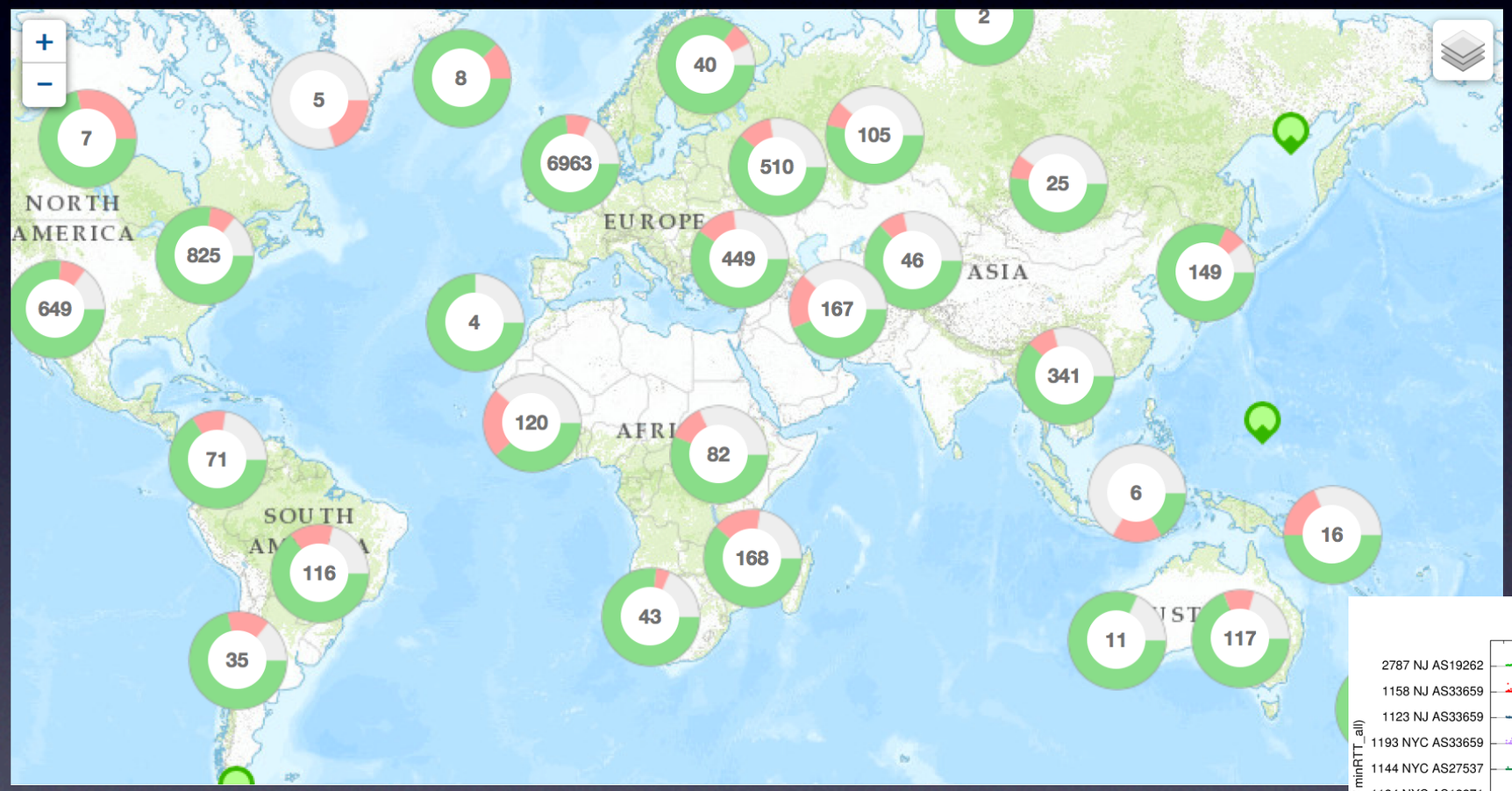
Code:	0.0	5.0	10.0	15.0	20.0	25.0	30.0	35.0	40.0	45.0	50.0	MB/s

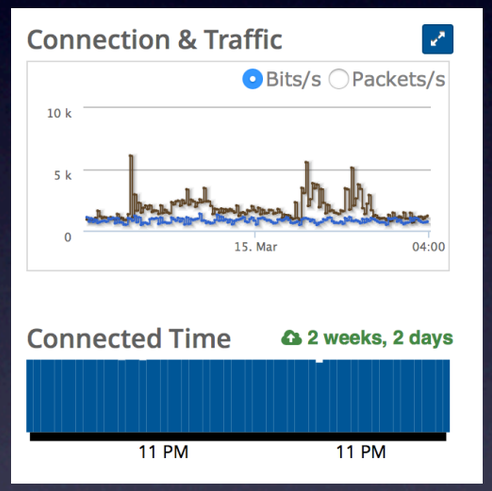
http://pprc.qmul.ac.uk/~lloyd/gridpp/nettest_v6.html

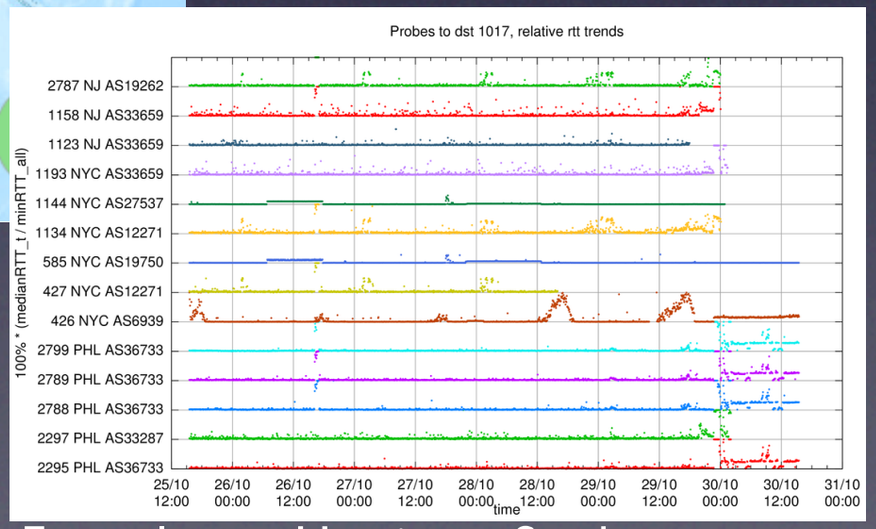
Production deployment in WLCG: 16/04 B503 11:45 – 12:00

RIPE ATLAS – Global measurements

- We work with RIPE NCC to build a global internet measurement network. https://atlas.ripe.net/



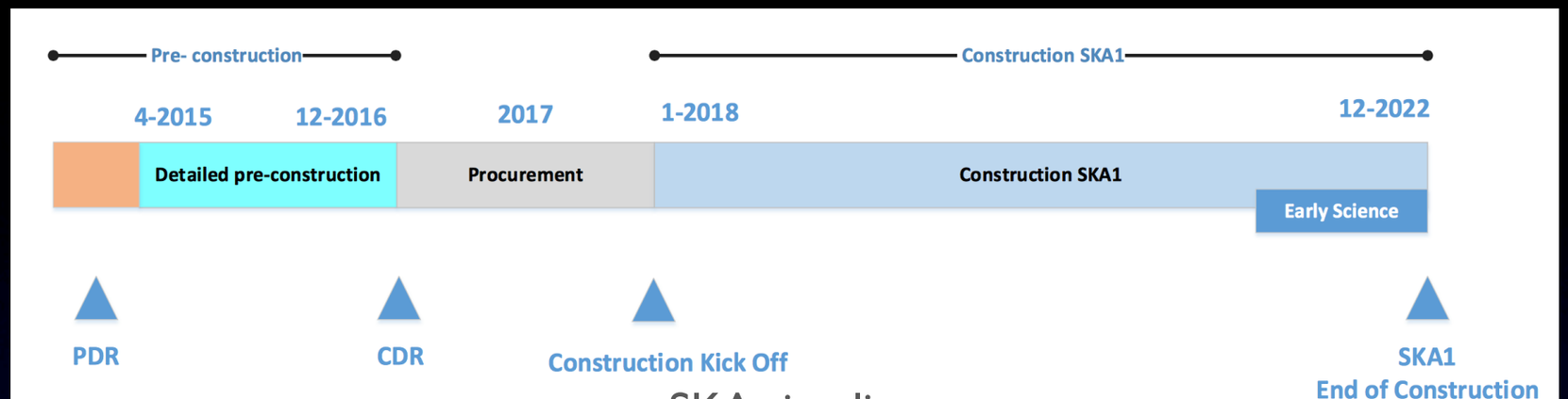




Example use: Hurricane Sandy

For a probe contact me or apply via https://atlas.ripe.net/.
Probes are sponsored. There are also requests to enable more anchor nodes. ATLAS probes compliment perfSONAR with better coverage.

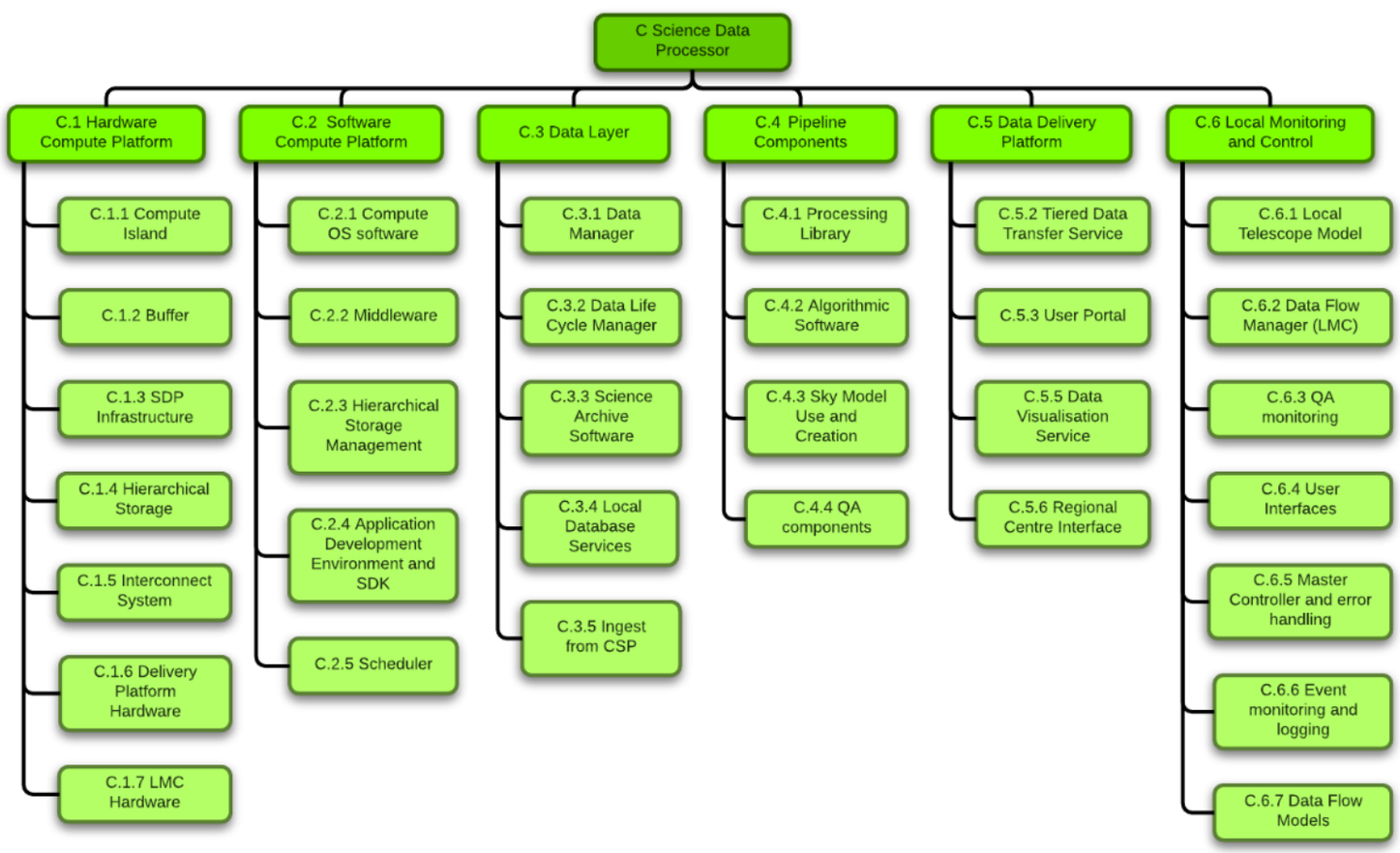
Wider collaboration



SKA timeline

- Look for commonalities
- HEP +… LSST, LOFAR, LIGO..

- SKA pre-cursors
- Power budget limits



SDP sub-system product tree

- SDP = Science Data Processor

"the design of the computing hardware platforms, software, and algorithms needed to process science data from the correlator or non-imaging processor into science data products".

- Several areas of potentially interesting overlap

Summary

- Resource considerations: Fine for now. CPU cost projections a concern.
- Batch and CE evolution: We diversified. HTC+ARC looking good.
- Multicore optimisation: drainBoss results promising.
- Worker node integration: Working implementations. VAC is lightweight option.
- Reusing tools: Community DIRAC service.
- Infrastructure tasking: More than just running hardware.
- IPv6 progress: Takes time for local institutes. Keep pushing!
- RIPE ATLAS probes: Global internet measurement network. Useful.
- Wider collaboration: Important in coming years.