UKI-LT2-RHUL
ATLAS STEP09 report

Duncan Rand
on behalf of the RHUL Grid team

The Newton cluster

- 400 cores (2.4 kSI2K; 7.3 HEP-SPEC06/core)
- 2GB RAM per core
- DPM: 12 x 22TB pool nodes (total 264TB, 264 disks)
- CE/SE well resourced (similar spec as WN)
- WAN: 1 Gbit/s LAN: ~37Gbit/s
- worker nodes: 2 Gbit/s storage: 4 Gbit/s
- Located: Imperial machine room (LMAN PoP)

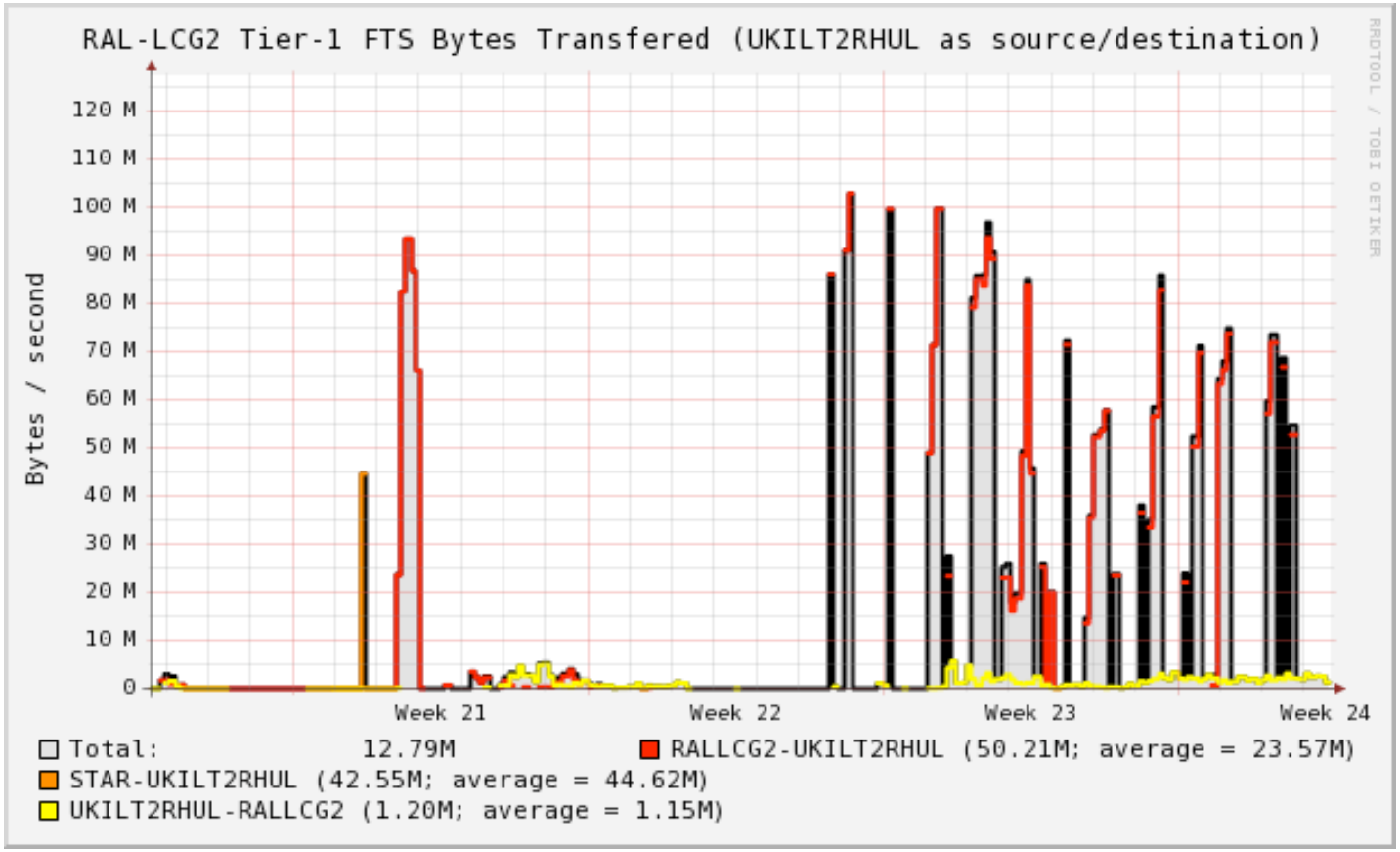
ATLAS STEP09

1. Data distribution to the site
2. Monte Carlo production
3. Hammercloud user analysis
   - Ganga WMS
     - RFIO and filestaging
   - Panda -
     - filestaging only

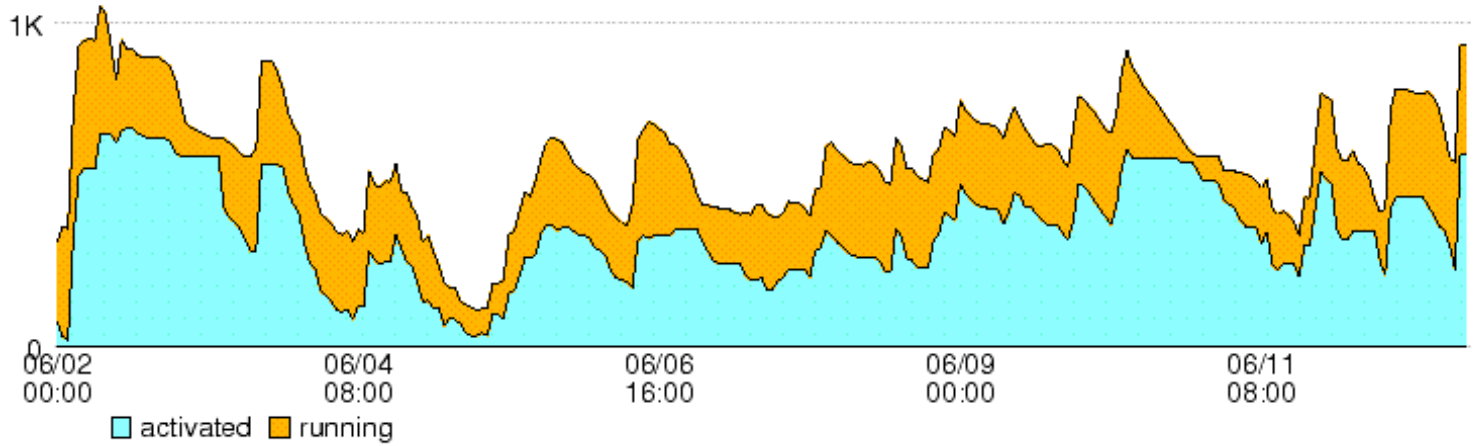
Data distribution

- 23% of the UK data i.e. 26TB
- No obvious problems were encountered in this section of the exercise; data was absorbed by the site as and when required, apparently without backlogs
- Disk sufficient - no filling up of space tokens
- Bursty transfers
- Cluster moving to Egham over summer
- Support move to 2 Gbit/s WAN

Data transfer to and from RHUL

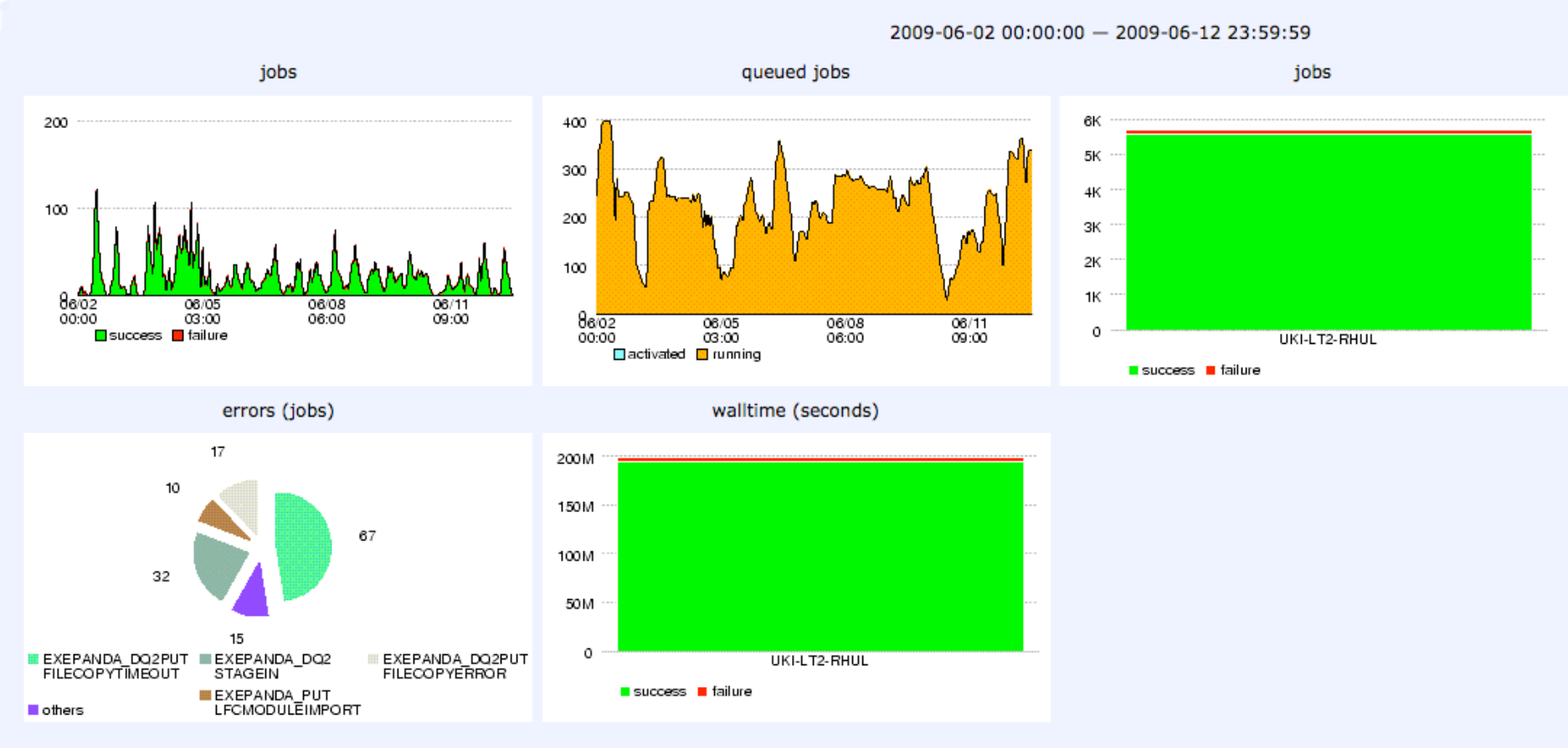


Monte Carlo production



- Steady rate of Monte Carlo production maintained throughout STEP09
- A total of 5578 jobs were completed with an efficiency of 97.5%

MC production at RHUL

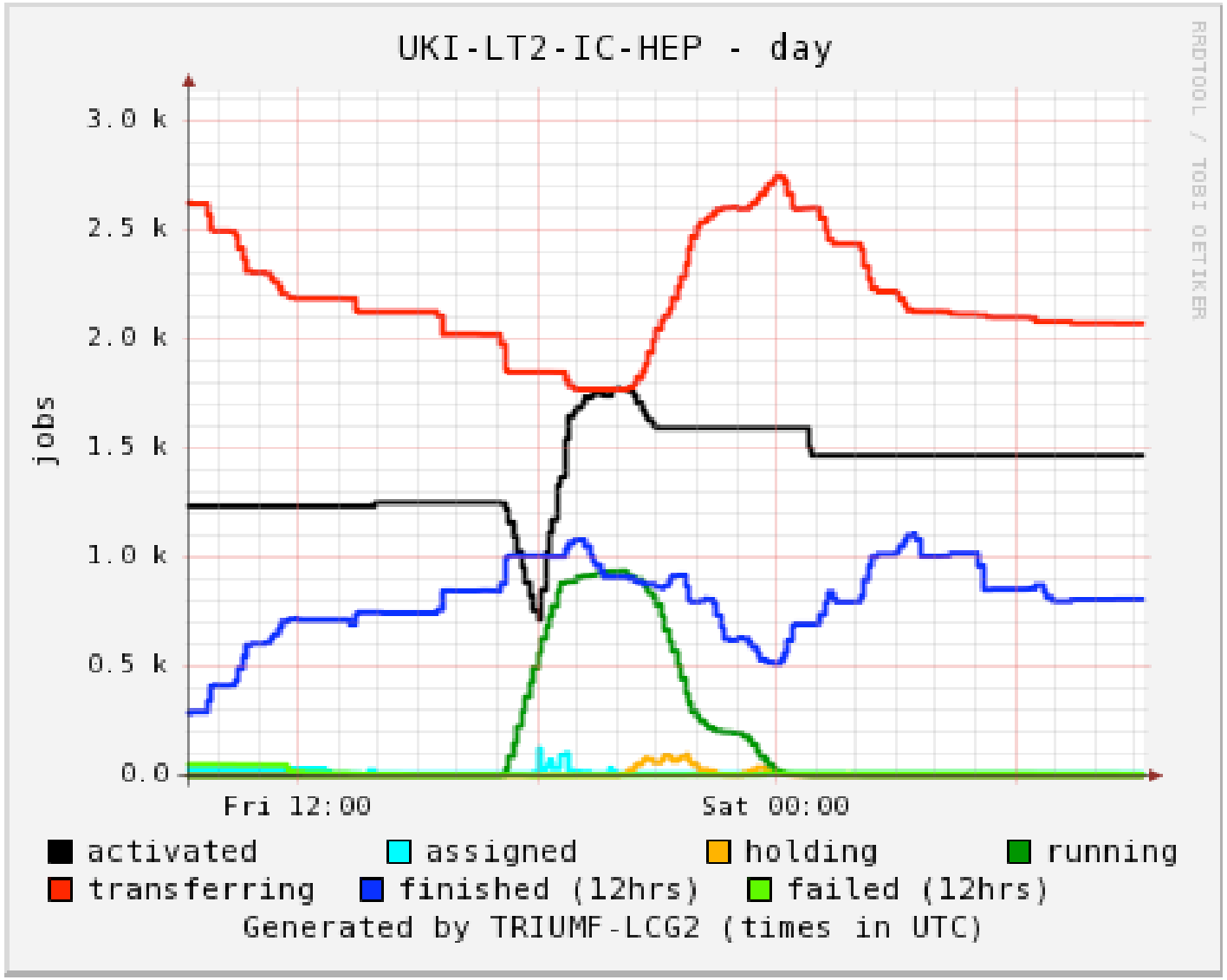


site	defined	assigned	waiting	activated	running	holding	transferring	success	failure	efficiency
× UKI-LT2-RHUL	0	0	0	0	1	5	0	5578	141	97.5%
total	*0*	*0*	*0*	*0*	*1*	*5*	*0*	*5578*	*141*	*97.5%*

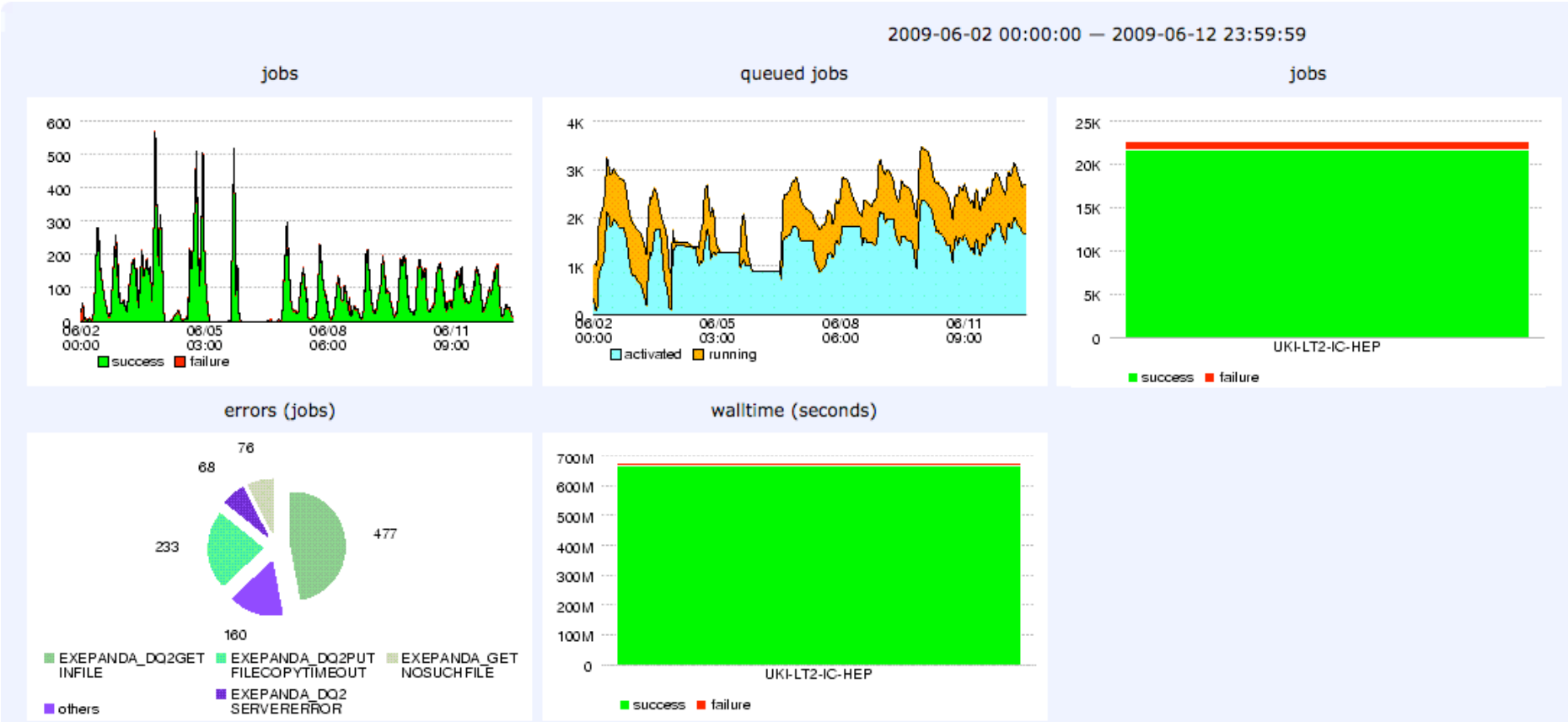
IC-HEP: MC data upload problems

- UKI-LT2-IC-HEP stages out to RHUL SE
- Problems copying data back to RAL over weekend of 5-6th June
- Gridftp transfers stalled, until time-out (3600s), blocking one or more of the 4 FTS file shares RALLCG2-UKILT2RHUL
- Result: IC-HEP transferring jobs > 2000 and consequent throttling of jobs
- Obviously need to understand original problem
- *Recomendation: Improve FTS channel configuration to mitigate against similar such problems in the future*

MC production at Imperial-HEP



MC production at Imperial-HEP



site	defined	assigned	waiting	activated	running	holding	transferring	success	failure	efficiency
UKI-LT2-IC-HEP	0	0	0	60	0	10	0	21589	1014	95.5%
total	*0*	*0*	*0*	*60*	*0*	*10*	*0*	*21589*	*1014*	*95.5%*

Batch system shares

- Appropriate fair shares between and within VO's were set in Maui (after idea from RAL)
- ACCOUNTCFG option used to set percentages for three LHC VO's and 'others'
- GROUPCFG option (i.e. unix group) used to set intra-LHC VO fairshares and 'others' share

  e.g. 50% /atlas/Role=production, 25% /atlas/Role=pilot, 25% /atlas
- Successful - allowed tight control of resource
- In practice capping of ATLAS analysis jobs prevented fairshare operating fully during STEP09

```
FSUSERWEIGHT           10
FSGROUPWEIGHT          100
FSACCOUNTWEIGHT        1000
# ATLAS 89% overall
ACCOUNTCFG[atlas] FSTARGET=89
# Within ATLAS: 50% /atlas/Role=production, 25% /atlas/Role=pilot, 25% /atlas
GROUPCFG[atlas]      FSTARGET=25 ADEF=atlas MAXJOB=400
GROUPCFG[atlaspil]   FSTARGET=25 ADEF=atlas MAXJOB=400
GROUPCFG[atlasprd]   FSTARGET=50 ADEF=atlas MAXJOB=400
GROUPCFG[atlassgm]   FSTARGET=1  ADEF=atlas PRIORITY=500000 MAXJOB=4
# CMS 5% overall
ACCOUNTCFG[cms] FSTARGET=5
# Within CMS: 50% /CMS/Role=production, 50% /CMS
GROUPCFG[cms]        FSTARGET=50 ADEF=cms
GROUPCFG[cmsprd]     FSTARGET=50 ADEF=cms
GROUPCFG[cmssgm]     FSTARGET=1  ADEF=cms PRIORITY=500000 MAXJOB=6
# LHCb 5% overall
ACCOUNTCFG[lhcb] FSTARGET=5
# 90% Role=pilot; 5% Role=user; 4% Role=production; 1% Role=lcgadmin
GROUPCFG[lhcb]       FSTARGET=5  ADEF=lhcb
GROUPCFG[lhcbprd]    FSTARGET=4  ADEF=lhcb
GROUPCFG[lhcbpil]    FSTARGET=90 ADEF=lhcb
# non-LHC VO's get 1% of cluster, shared equally (GROUPCFG FSTARGET= 1 for all)
ACCOUNTCFG[others] FSTARGET=1
GROUPCFG[biomd]      FSTARGET=1  ADEF=others
GROUPCFG[hone]       FSTARGET=1  ADEF=others
GROUPCFG[ilc]        FSTARGET=1  ADEF=others
```

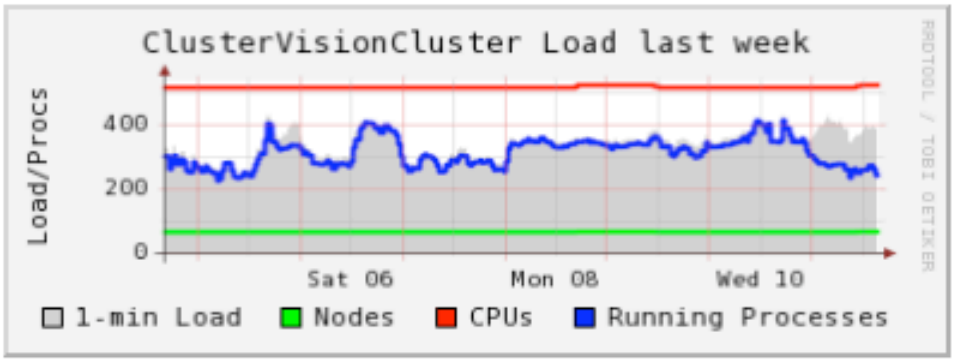
Ganga WMS based analysis

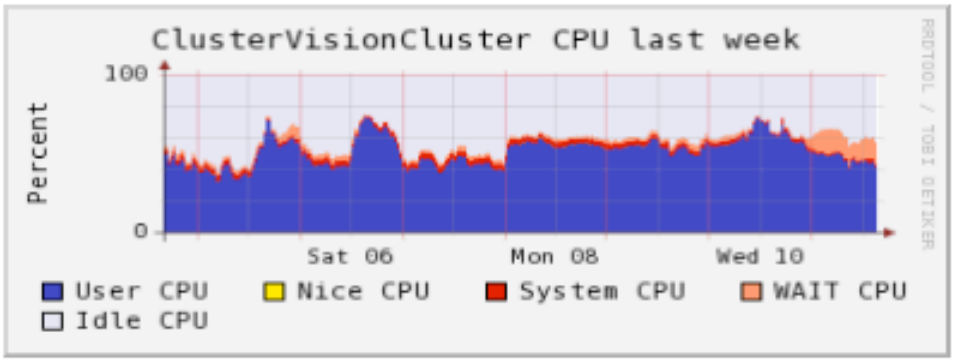
STEP09 Overall Site Results

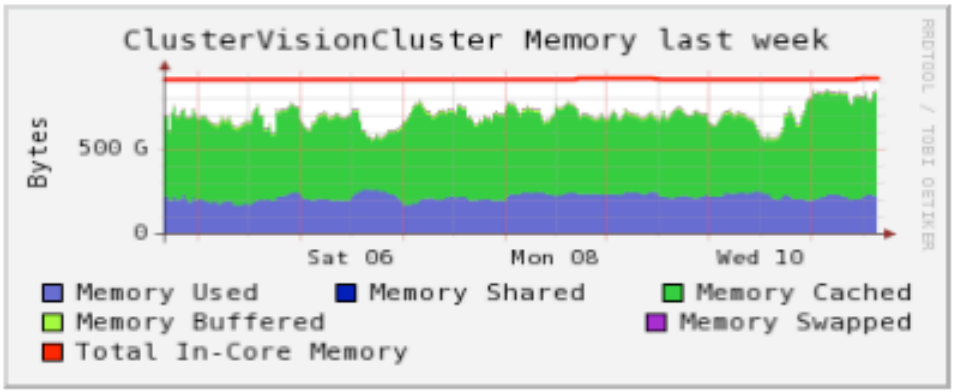
Cloud	Site ▲	# Jobs	# Successful	# Failed	Efficiency	#files	#events	Hz	CPU/Wall
UK	UKI-LT2-RHUL_MCDISK	12381	6571	5810	0.531	22272	178983220	5.8	46.3

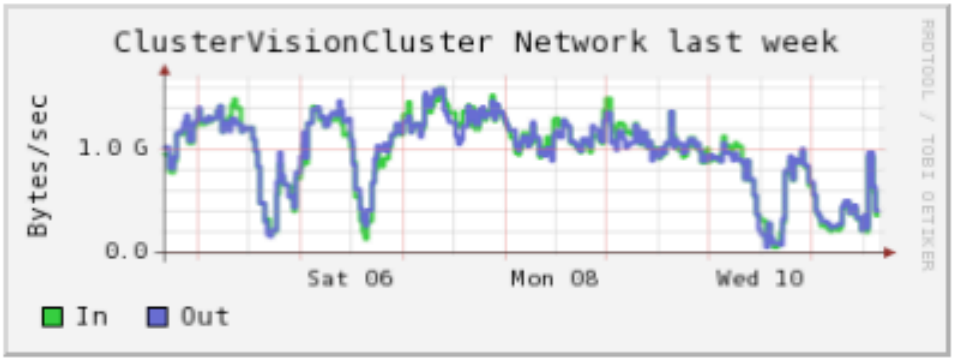
- Setpoint 100 slots within HammerCloud
- One pool node down - lots of ‘no such file’ errors
- RFIO IOBUFSIZE set to default 128KB throughout
- High LAN bandwidth (ganglia plots overestimate by factor 2)
- Hard to work out what was going on

LAN ganglia plots









Panda based analysis

STEP09 Overall Site Results

Cloud	Site ▲	# Jobs	# Successful	# Failed	Efficiency	#files	#events	Hz	CPU/Wall
UK	ANALY_RHUL	17877	14657	3220	0.820	36759	326268373	6.4	33.0

- Set point was 100 slots
- Any more than this and rfcp commands stalled
- Recent test (test_482) with hard cap of 100 got 11Hz event rate (59% efficiency)

Conclusion

- RHUL took full part in STEP09
- Data transfer into site successful
- Monte Carlo production was steady apart from some temporary problems transferring IC-HEP data back to RAL
- Analysis didn’t scale well - network performance needs careful examination
- Nevertheless 500 million events analysed