WLCG operations

ALICE T1-T2 Workshop
Bergen
April 18-20, 2016

Maarten Litmaath
CERN-IT

v1.0

VOMS-Admin saga

- ~1 year ago the obsolete VOMRS grid user registration service was replaced by VOMS-Admin
  - Not perfect, but actively maintained and developed
- Various issues then popped up ~1 year later
  - Unexpected VO membership “expiration” warnings
    - Should be handled better in the latest version
  - AUP re-signing request flood causing load on VO admins
    - Will be randomized to some degree
  - Errors for users of the INFN CA
    - As the CA DN had changed, they needed to re-register
  - Errors for users with 2 or 3 (historic) records
    - A cleanup has been done

➢ Things should be smoother next year…

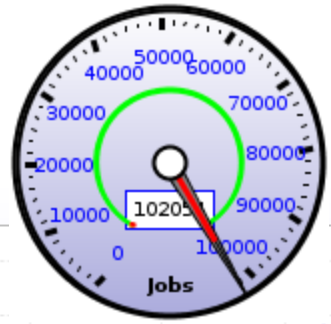
Central services

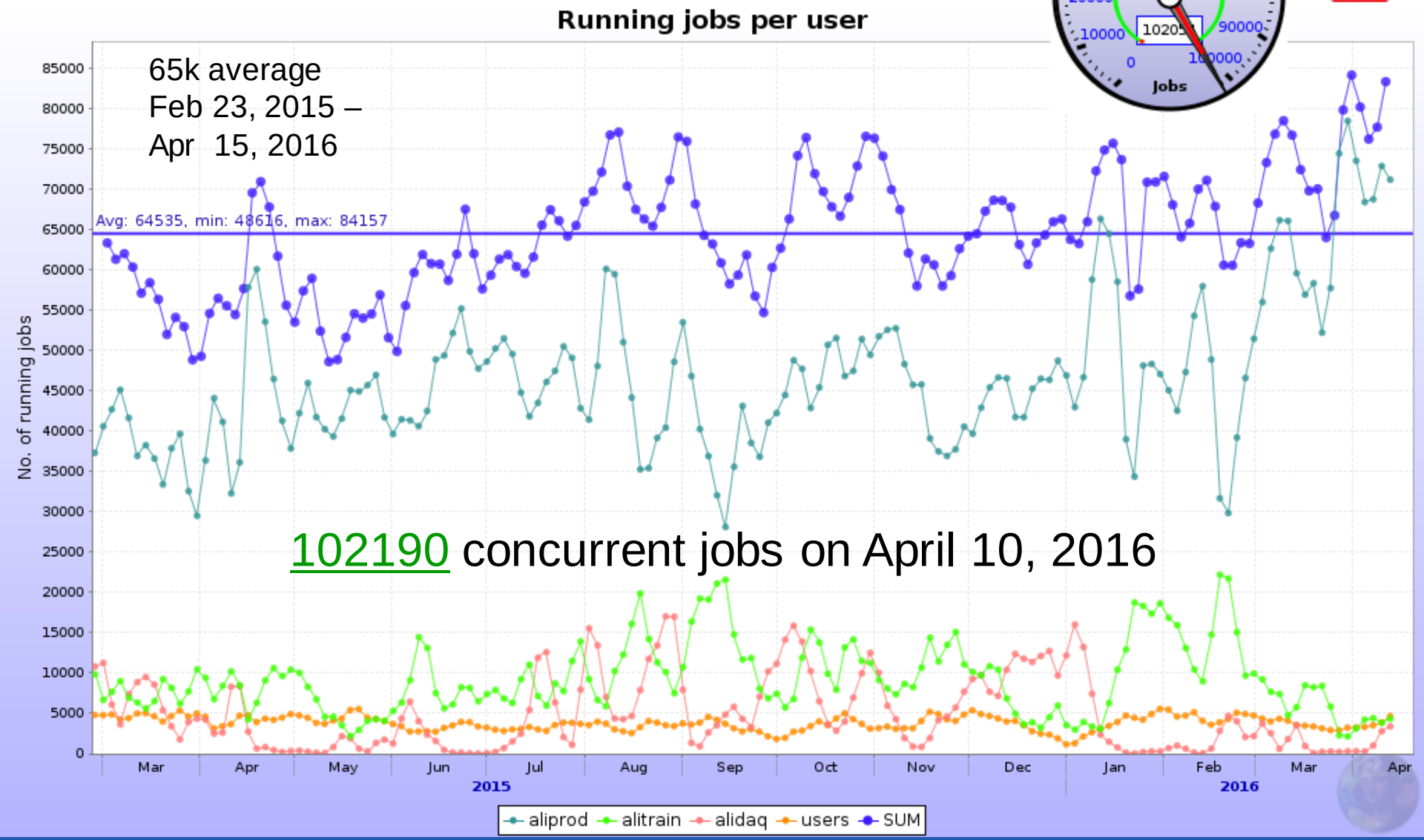
	Machine status			Machine type					Mem	
Machine	Online	Uptime	Load	Kernel	Machine model	CPU	CPUs	MHz	Total ▲	Used
8. db6c		301d 8:35	4.96	3.19.0-21...	ProLiant DL380 Gen9	Xeon E5-2687W v3 3.10GHz	40	1200	755.8 GB	298 GB
7. db6b		220d 9:22	1.01	3.19.0-26...	ProLiant DL380 Gen9	Xeon E5-2687W v3 3.10GHz	40	1200	755.8 GB	149.5 GB
13. db9		185d 2:05	0.05	3.13.0-65...	ProLiant DL380p Gen8	Xeon E5-2690 2.90GHz	32	1200	377.9 GB	20.43 GB
36. pcalimonitor		92d 4:39	4.93	3.13.0-74...	ProLiant DL380p Gen8	Xeon E5-2690 v2 3.00GHz	40	1200	377.9 GB	90.44 GB
27. alienvm1		79d 8:31	0.26	3.13.0-76...	ProLiant DL380p Gen8	Xeon E5-2690 v2 3.00GHz	40	1200	377.9 GB	29.31 GB
11. db7		545d 20:48	3.99	3.13.0-37...	ProLiant DL380p Gen8	Xeon E5-2690 2.90GHz	32	1200	377.9 GB	11.89 GB
35. pcaliweb02		496d 8:54	0.09	3.13.0-40...	ProLiant DL380p Gen8	Xeon E5-2690 2.90GHz	32	1200	377.9 GB	6.758 GB
41. alientest02		24d 2:10	14.27	4.2.0-34-...	ProLiant DL380p Gen8	Xeon	40	3000	377.9 GB	226.3 GB
39. pcalimonitor4		23d 19:37	1.39	4.2.0-34-...	ProLiant DL380p Gen8	Xeon E5-2690 2.90GHz	32	1526	377.9 GB	8.765 GB
37. pcalimonitor2		23d 23:41	0.74	4.2.0-34-...	ProLiant DL380p Gen8	Xeon E5-2690 2.90GHz	32	1699	377.9 GB	27.67 GB
10. db6e		266d 8:53	0.54	3.19.0-23...	ProLiant DL380 G7	Xeon X5680 3.33GHz	24	1600	283.4 GB	116.5 GB
46. alinsure		150d 3:14	0.33	3.13.0-68...	ProLiant DL380p Gen8	Xeon E5-2697 v2 2.70GHz	48	1200	188.9 GB	24.62 GB
40. alitraintest2		185d 2:05	1.03	3.13.0-65...	ProLiant DL380p Gen8	Xeon E5-2690 2.90GHz	32	1200	188.9 GB	14.81 GB
17. api3		545d 19:20	0.1	3.13.0-37...	ProLiant DL380p Gen8	Xeon E5-2690 2.90GHz	32	1200	188.9 GB	73.21 GB
28. alienvm2		185d 2:04	6.37	3.19.0-30...	ProLiant DL380p Gen8	Xeon E5-2697 v2 2.70GHz	48	3254	188.9 GB	113.5 GB
44. alientest06		24d 2:53	6.45	4.2.0-34-...	ProLiant DL380p Gen8	Xeon E5-2697 v2 2.70GHz	48	3000	188.9 GB	40.01 GB
43. alientest05		13d 3:40	0.25	4.2.0-34-...	ProLiant DL380p Gen8	Xeon E5-2697 v2 2.70GHz	48	1199	188.9 GB	20.18 GB
31. alienvm5		185d 2:04	7.09	3.13.0-65...	ProLiant DL380 G7	Xeon X5690 3.47GHz	24	1596	141.7 GB	73 GB
12. db8		545d 20:13	5.41	3.13.0-37...	ProLiant DL380 G7	Xeon X5690 3.47GHz	24	3459	141.7 GB	32.03 GB
9. db6d		24d 0:08	7.05	4.2.0-34-...	ProLiant DL380 G7	Xeon X5690 3.47GHz	24	1596	141.7 GB	26.63 GB
6. db6a		323d 9:20	0.14	3.19.0-18...	ProLiant DL380 G6	Xeon X5560 2.80GHz	16	1600	141.7 GB	11.49 GB
42. alientest03		23d 9:11	2.37	4.2.0-34-...	ProLiant DL380 G6	Xeon X5560 2.80GHz	16	1596	141.7 GB	70.16 GB
5. db5		545d 19:37	1.15	3.13.0-37...	ProLiant DL380 G7	Xeon X5690 3.47GHz	24	1596	125.9 GB	9.012 GB
30. alienvm4		185d 2:05	0.58	3.13.0-65...	ProLiant DL380 G6	Xeon X5560 2.80GHz	16	1596	125.9 GB	36.34 GB
15. api1		126d 2:55	26.87	3.13.0-71...	ProLiant DL380 G6	Xeon X5560 2.80GHz	16	1596	125.9 GB	91.57 GB
21. api7		545d 19:11	4.01	3.13.0-37...	ProLiant DL380 G6	Xeon X5560 2.80GHz	16	1596	125.9 GB	89.1 GB
29. alienvm3		23d 23:26	0.07	4.2.0-34-...	ProLiant DL380 G6	Xeon X5560 2.80GHz	16	2793	118 GB	21.87 GB
1. db1		184d 10:57	5.32	3.13.0-57...	ProLiant DL380 G6	Xeon X5560 2.80GHz	16	1596	110.2 GB	15.65 GB

Site evolution

- Growth thanks to increased pledges and new sites
- Thanks & bye-bye: LLNL, CSC, Madrid, NECTEC
- Welcome to ORNL, HIP, CBPF, Vienna, Tsukuba, Subatech_CCIPL
- Welcome back Bandung, Cibinong, Wuhan
- Welcome to the HLT clusters
  - Up to 8.5k (HLT) + 800 (HLTDEV) job slots
- Welcome to CERN-SIRIUS
  - Shared HTCondor cluster with 15k cores and growing
- Welcome to Altaria
  - Virtual site in front of cloud deployments
- New EOS instances at Kolkata, LBL
- EOS-ALICE head nodes now have 512 GB RAM
  - That should be sufficient for a long time, thanks!
- New VOBOXes at CNAF, KISTI, KIT

High activity





102190 concurrent jobs on April 10, 2016

Sep 27 – Oct 3: QM 2015

CASTOR (1/2)

- Various incidents led to a reorganization for heavy ion data taking:
  - 1 “fast” pool for DAQ + writing to tape
  - 1 “slow” pool for reco, replication to T1 sites, recalls from tape
  - New files in pool 1 automatically get replicated to pool 2
  - Many disk servers that were prepared for retirement were temporarily added
  - Enough space to keep all 2015 HI data on disk for a number of months

CASTOR (2/2)

- Mostly stable operation during and after the heavy ion run
  - A few incidents were quickly resolved
- Plans for this year were discussed with the CASTOR team
- The old extra HW was already extended until ~mid April
  - It may remain available yet for a few more months
- New HW will be added as of late April
- The resources seem sufficient
  - 2015 Pb-Pb and pp reco and reprocessing
  - 2016 data taking
- Thanks to the CASTOR and Fabric teams!

Grid storage

- EOS-ALICE was almost completely full
  - Led to service instabilities
  - And failures for users and jobs around the grid
  - Mitigated by the admins, thanks!
  - Cured through cleanups
- About 2.5 PB were recovered on the grid thanks to ad-hoc cleanups
- Further cleanups are expected this spring
  - Pending agreement on data usage and retention policy changes

Heavy ion reconstruction (1/2)

- Important changes in the code and workflow were implemented to reduce the memory usage
  - Tested with 2011 heavy ion reference data
  - If all goes well, the reco jobs will only need ~2.5 GB RAM
- To be on the safe side, special arrangements were made with the sites that received 2015 HI raw data
  - CNAF, FZK, KISTI and SARA set up dedicated high memory queues
  - At CERN the jobs can request 2 cores and hence have twice the memory
  - All setups were tested with normal jobs
  - Thanks to the sites for their good support !
- Normal queues will be used wherever possible, depending on the actual memory needs

Heavy ion reconstruction (2/2)

- So far the jobs needed at most ~2.5 GB RSS
- The high-memory arrangements at CERN and the T1 sites were undone
  - Could be reinstated per site if needed
- Waiting for the big campaign to start…

RFC proxies

- All WLCG VOBOXes are running with RFC proxies since 1 year
- AliEn v2-19.276 is the default for “alienv” since many months
  - It does not support legacy proxies
- WLCG is steadily moving away from legacy proxies
  - The 4 LHC experiments already rely on RFC proxies
- The default proxy type is foreseen to become RFC later this year
  - New version of VOMS clients
  - Updated UI configuration for MyProxy

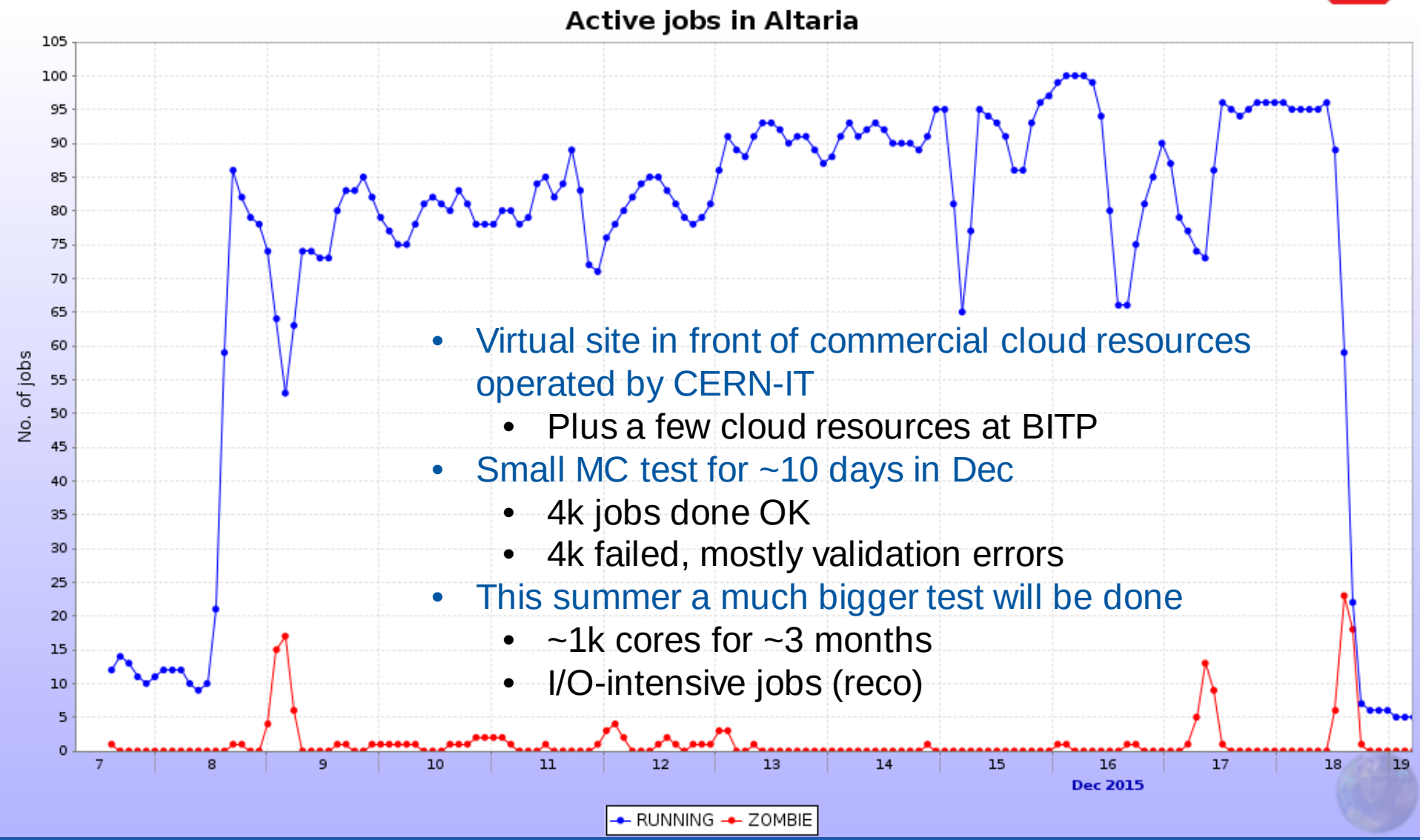
Middleware

- SL5 will be deprecated for MW this spring
  - EGI have set a deadline of April 30
  - Already there have been openssl compatibility issues
- CentOS/EL7 will become more important
  - Some services already available
  - SL6 remains the default
- HTCondor batch system remains on the rise
  - It comes with its own CE
    - AliEn can use it (currently still via a prototype)
    - SAM tests to be added this spring
  - The ARC CE remains an alternative

Birds and clouds

- A new AliEn job submission module for the HTCondor CE has been worked on by Pavlo
  - In collaboration with CERN HTCondor service manager Iain Steers
- To be used at CERN and other sites moving to such a setup
  - Some code improvements are still in progress
  - Meanwhile a patched prototype is being used in production
- Also used for job submissions to a special CE in front of commercial cloud resource projects from CERN-IT
  - This way AliEn just sees a virtual "site" with a standard CE
  - The cloud management aspects are handled behind the scenes by IT department experts
- The *Altaria* site is being used for that
  - Located on Eyjafjallajökull ! ☺

Altaria



- Virtual site in front of commercial cloud resources operated by CERN-IT
  - Plus a few cloud resources at BITP
- Small MC test for ~10 days in Dec
  - 4k jobs done OK
  - 4k failed, mostly validation errors
- This summer a much bigger test will be done
  - ~1k cores for ~3 months
  - I/O-intensive jobs (reco)

SAM - ETF

- ETF = Experiments Test Framework
  - New version of the previously unnamed SAM-Nagios test framework
- Changes
  - SL5 → SL6
  - Core rewritten and simplified, still based on Nagios
  - WN test results no longer sent directly via the message bus, but taken from the job output
  - All services in the experiment's VO feed can be tested
    - Presence in GOCDB is no longer required
  - RFC instead of legacy proxies
  - Classic Nagios GUI plus new Check_MK GUI
- The probes essentially are unchanged → the Availability / Reliability computations are unaffected
- Already tested and compared with production since many weeks
- Planned date for the switch: April 25
- Further details:
  - https://indico.cern.ch/event/466818/contribution/1141679/attachments/1230650/1804584/etf_intro_wlcg.pdf

SAM - MonALISA

- New Availability / Reliability profile based on selected MonALISA results remains not yet critical
  - http://wlcg-sam-alice.cern.ch/templates/ember/#/plot?profile=ALICE_MON_CRITICAL
- Its introduction was delayed because of issues with the Dashboard GUI that needed to be fixed
- We intend to give the new profile a try this spring
- Sites will first be alerted to compare old and new profile results
  - SE test failures will reduce the A / R!
- Test job submission to the HTCondor CE still to be added
- Direct submission to ARC CE in production since ~1 year (also for LHCb) thanks to Pavlo!

New A/R computation

- Planned new formula as of May:
  - Computing = (any CE) || (!CE && any VOBOX)
  - Storage = all SE || !SE
  - A/R = Computing && Storage
- Meaning
  - If any CE is working → Computing OK
  - If no CE is used and the any VOBOX is working → Computing OK
  - If all SE at the site are working for reading → Storage OK
    - A T1 has multiple (logical) SE
    - Test SEs will not be considered
  - If the site has no SE → Storage OK (!)
    - For now…
- NDGF-T1 is a virtual site with 4 VOBOXes

Xrootd reminder

- Sites should start planning upgrades to Xrootd >= 4.1
  - Stable and well supported
  - Required for IPv6 support
- T1 sites are particularly concerned
  - They receive raw data exported from T0
  - Latest CASTOR incompatible with old xrd3cp
    - Gateway hosts are being used as a temporary workaround
- Communication via LCG Task Force list as usual for expert advice
- SL6 hosts now can have xrootd for ALICE installed through rpms!
  - http://linuxsoft.cern.ch/wlcg/
  - Thanks to Adrian Sevcenco!

Networks

- KISTI OPN link at 10 Gbps since Apr 24
- Successful workshop on network evolution in Asia Sep 22-24 at KISTI
- All ALICE sites in Asia were represented
- LHCONE peering was agreed between network infrastructure providers TEIN, KREONET2 and ASGC

Tips for sites – thanks!

- Possible issues on VOBOX, CE, WN
  - CVMFS problem, CE not ready for jobs, myproxy running low, myproxy type wrong, …
  - Absence of “system” library
    - HEP_OSlibs rpm helps avoid that
- Jobs may fail due to SE problem
- Admins please check site issues page
  - http://alimonitor.cern.ch/siteinfo/issues.jsp
- Subscribe to relevant notifications
  - http://alimonitor.cern.ch/xml.jsp