

SW distribution tests at Lyon

Pierre Girard

Luisa Arrabito, David Bouvet

Yannick Perret, Xavier Canehan

Suzanne Poulat, Rolf Rumler

Jamboree LHCb

March 7th, 2011

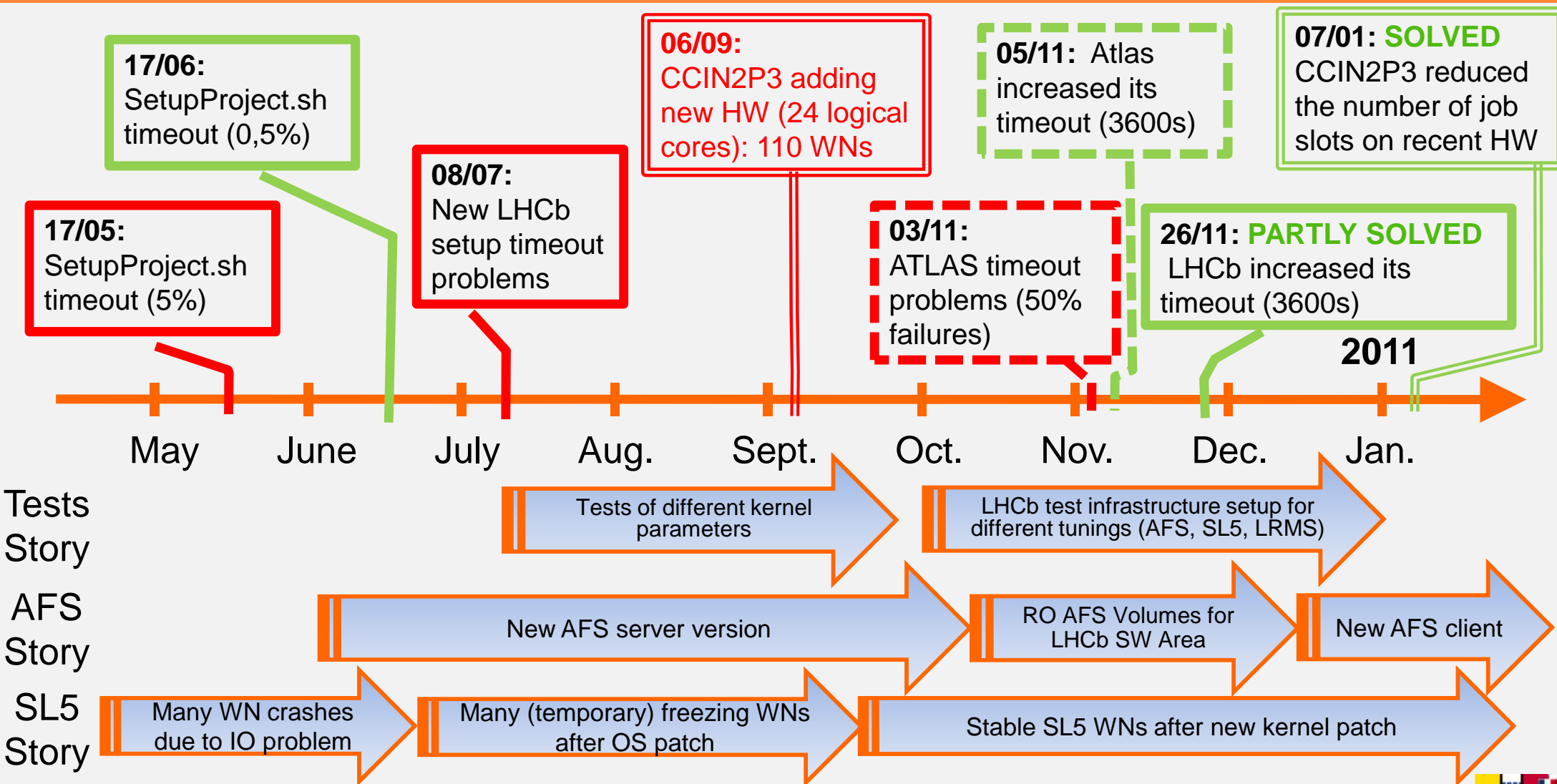


- **AFS Latency story**
 - AFS latency problem schedule
 - LHCb SetupProject tests
 - Preliminary conclusions
- **xxx-FS client stress tests**
 - Test suite results
 - CVMFS tests results
- **Conclusions**

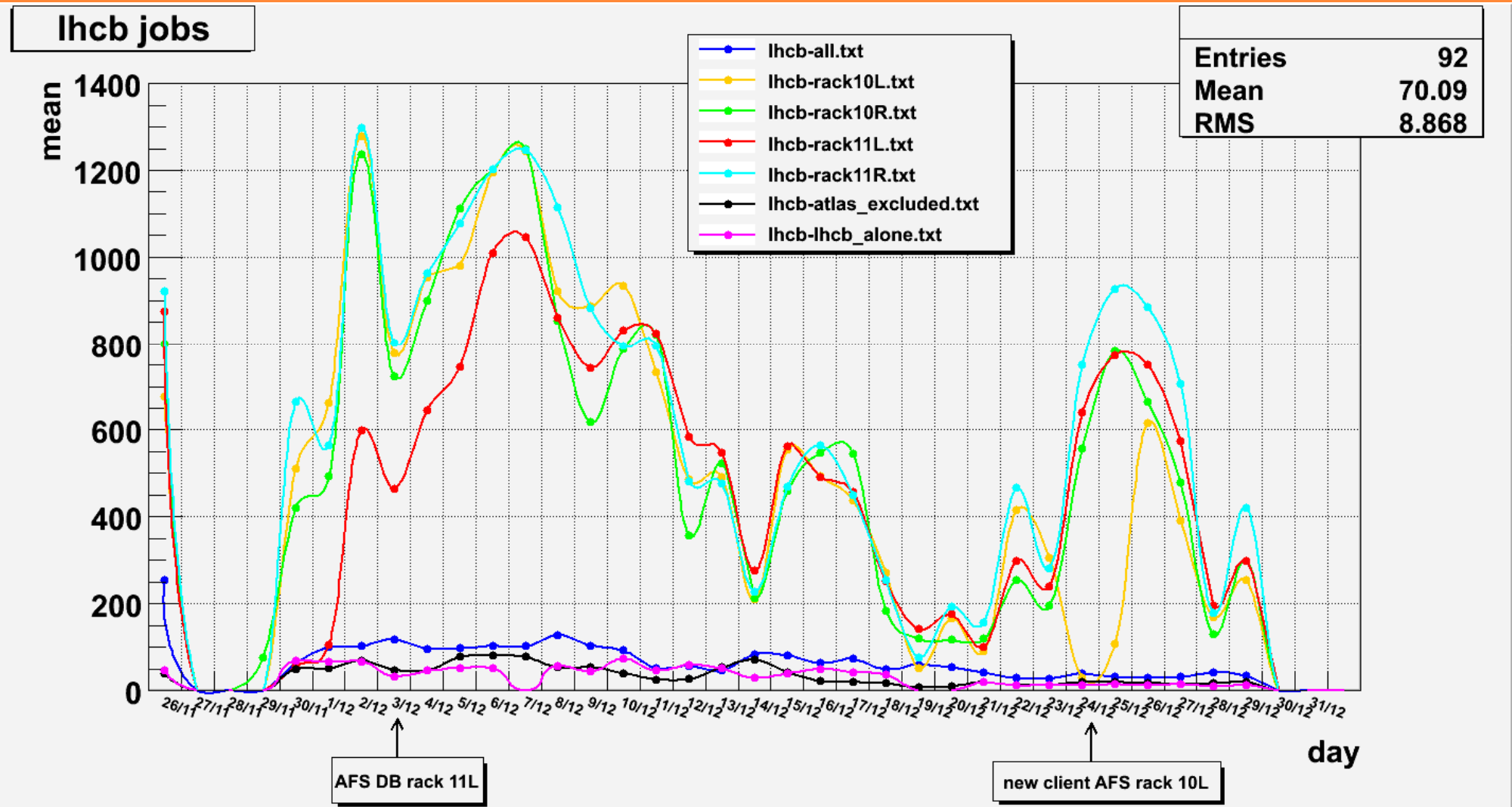
AFS latency story



AFS latency problem schedule



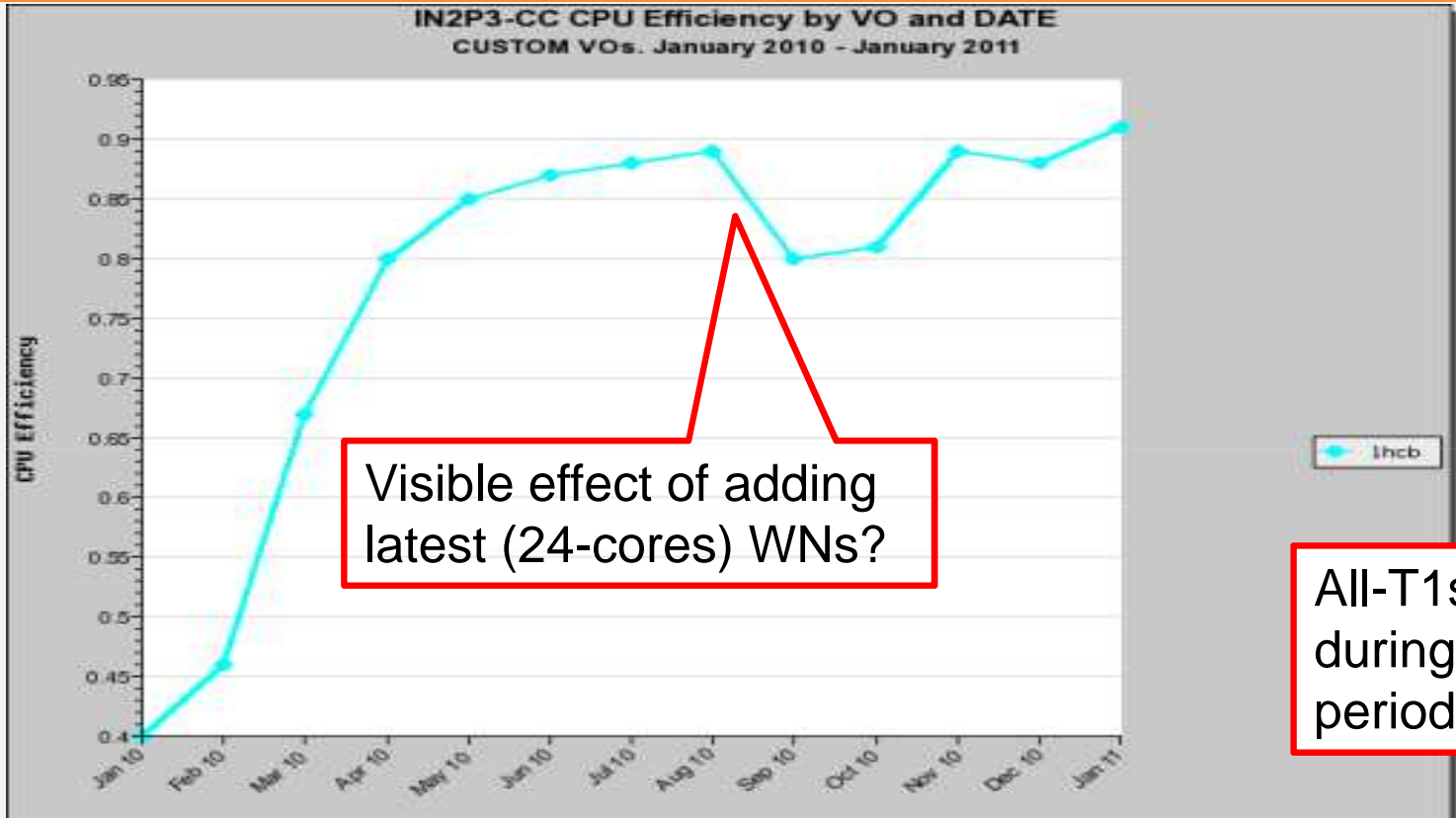
LHCb job setup tests



Source: L. Arrabito

AFS Latency / LHCb job efficiency

Source: http://www3.egee.cesga.es/gridsite/accounting/CESGA/tier1_view.html



Site	Jan 10	Feb 10	Mar 10	Apr 10	May 10	Jun 10	Jul 10	Aug 10	Sep 10	Oct 10	Nov 10	Dec 10	Jan 11	Total
CC	39.9	46.3	67.1	80.3	84.7	87.3	88.5	88.6	79.8	80.7	88.6	88.5	90.9	86.0
All T1s	42.7	58.4	74.6	82.4	88.2	88.4	88.3	73.5	79.4	85.1	81.7	89.8	92.8	82.5



Preliminary conclusions

- LHCb/ATLAS environment setup is very (too much) FS-intensive
 - By stracing SetupProject.sh
 - ◆ 17 868 open()
 - ◆ 110 765 stat()
- Investigate on job distribution strategy to avoid too many similar jobs on the same WN
 - According to “lhcb-alone” and “atlas-excluded” tests results
- AFS latency problem is now a AFS client scalability problem
 - Temporary solved by decreasing the number of job slots on the most recent machines, but ...
 - Is that a major concern in the near future ?
- Have the other sites already experienced the same problem ?



xxx-FS client stress tests

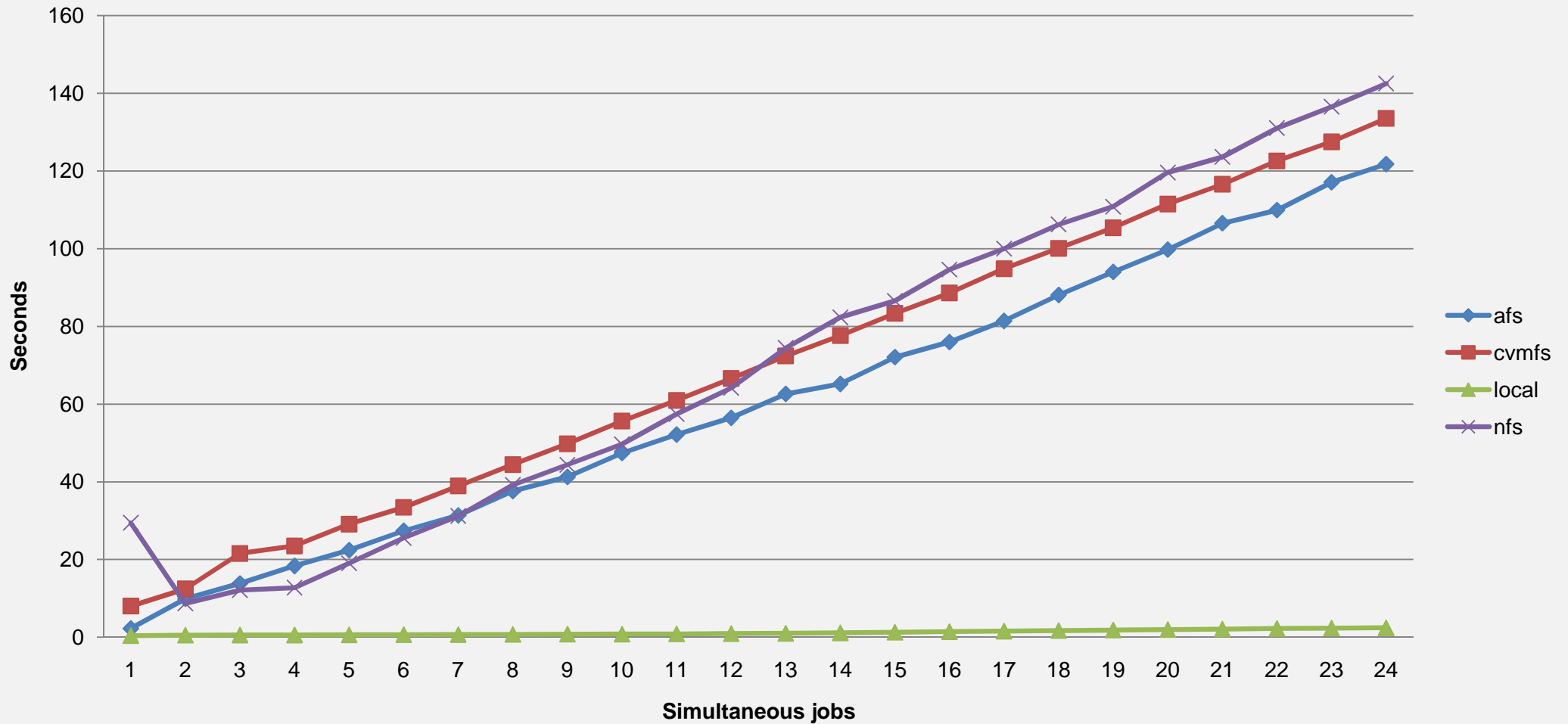
Test suite conditions

- For each test
 - Walk of the same directory arborescence (DAVINCI)
 - Same actions are achieved in the same order
 - ◆ LHCb ProjectSetup-like
 - ◆ 100 000 stat()
 - ◆ 7 000 open()
 - First block is read to ensure the open() is effective
- Pre-loading the cache (if any) by pre-executing the test once
- Averaged results are taken from 4 executions



FS Test Results

FS results comparison



CVMFS test suite conditions

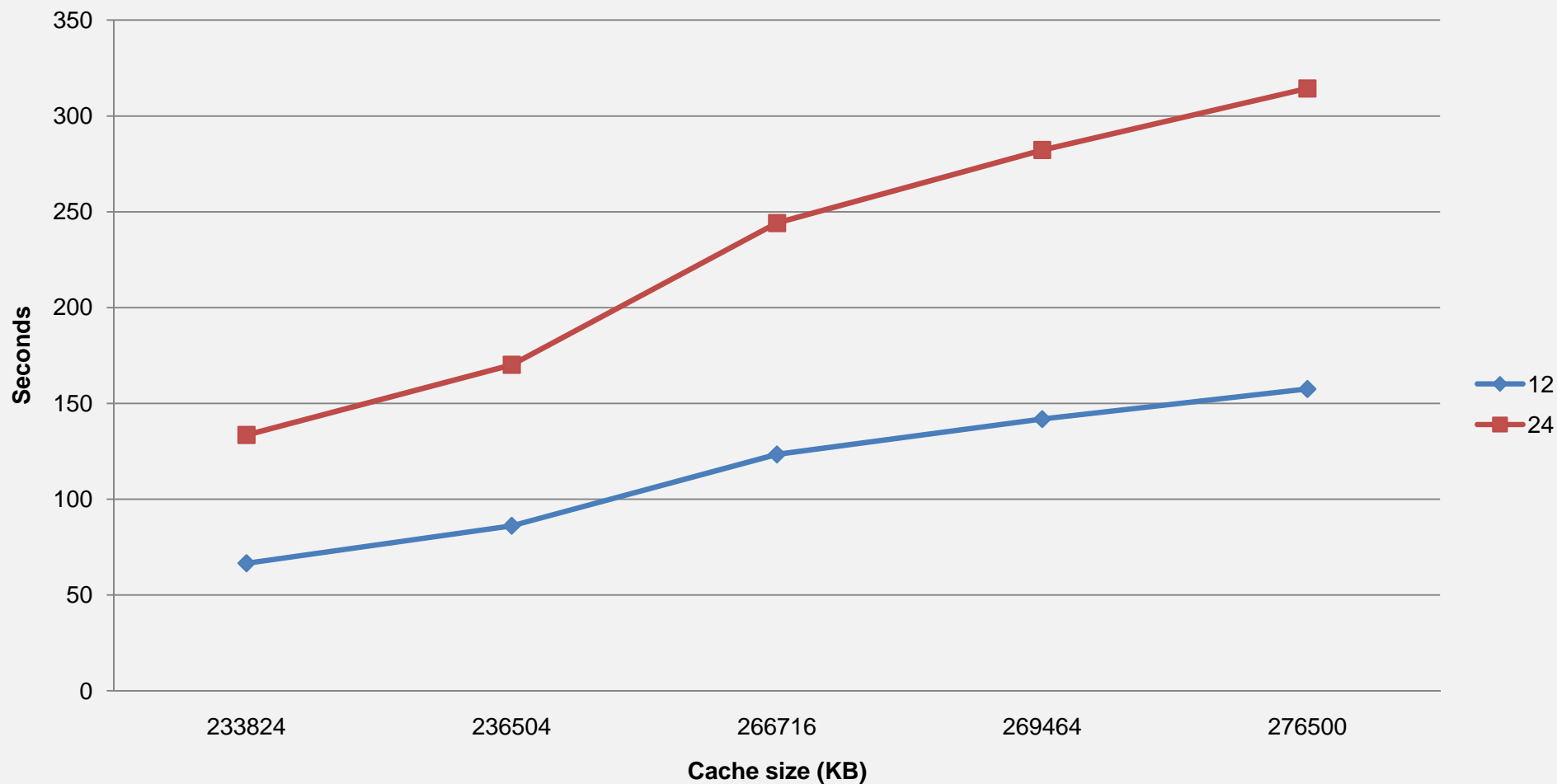
Different cache sizes

- **Dedicated SQUID**
 - Used by the tested WN only
 - With pre-loaded LHCb cache
- **On CVMFS client (0.2.53-1), before each test**
 - Cache was removed
 - Service was restarted
 - Different cache sizes
 - ◆ « ls -IR » on sibling directories to make grow up the cache



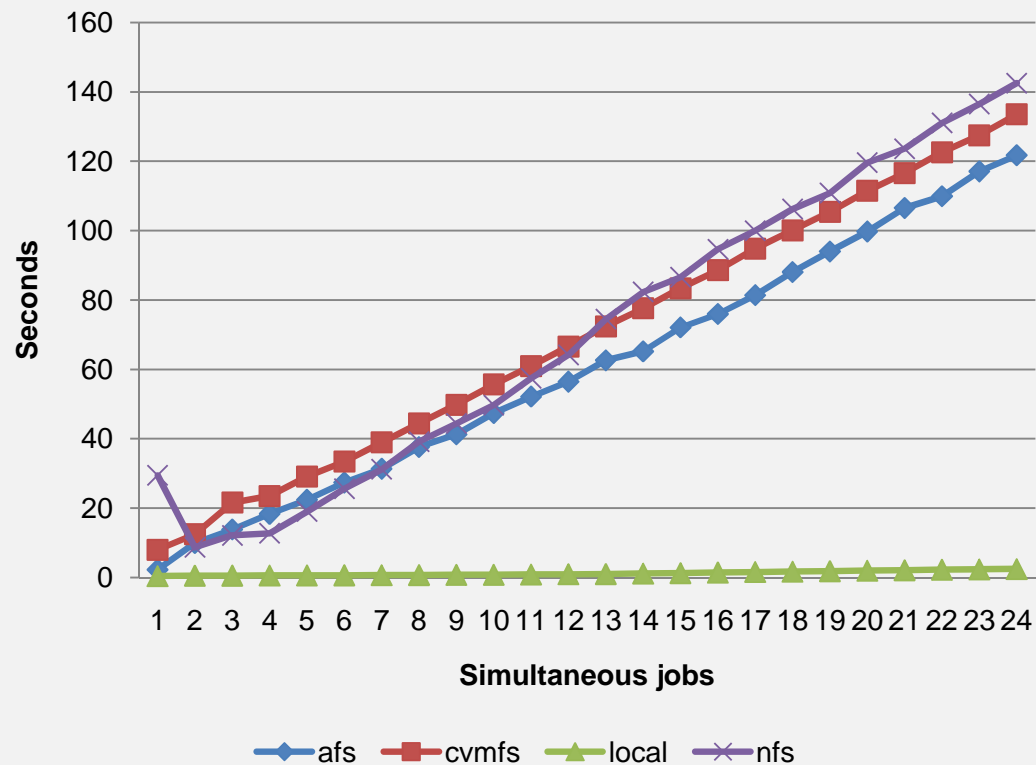
CVMFS Cache Size Tests Results

CVMFS cache size tests

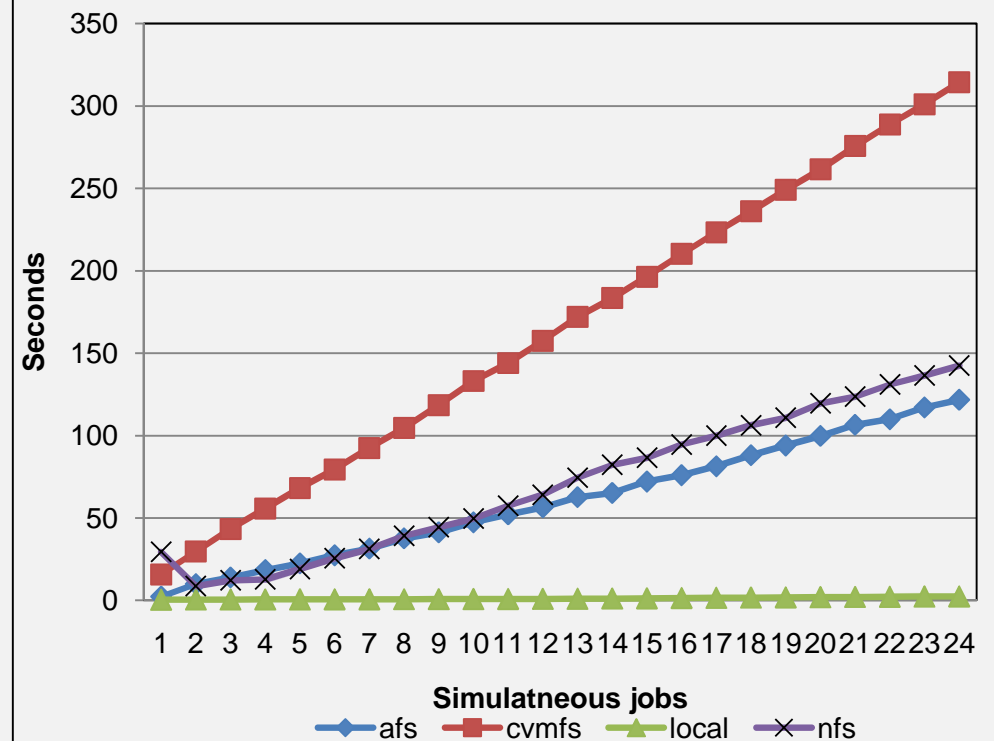


CVMFS Min/Max results

FS results comparison CVMFS 233824 KB



FS results comparison CVMFS 276500 KB



Conclusions



Conclusions

- LHCb/Atlas job setup should be optimized
- Multi-VOs sites should try to implement a fair distribution of VO' jobs over the cluster WNs
 - Restrict the number of similar jobs on a WN
- Issue of shared FS client scalability (Most likely)
 - Checked with AFS, NFS4.0, and CVMFS
 - Tests must go on
 - ◆ NFS4.1 (pNFS) still to be tested
 - ◆ With other HWs (for now, only Dell D6100)
 - ◆ By virtualizing the WNs (“divide and rule” principle)
 - First attempt was achieved by basically splitting 24-cores WN into 2x(12-cores VM-WN)
 - Must be further investigated
- CVMFS
 - Interesting for VO SW distribution (without installation job)
 - But, take care that latency increases with cache size



Questions & Comments

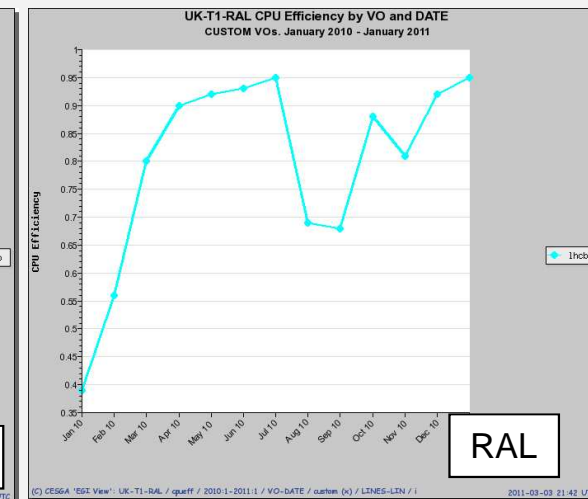
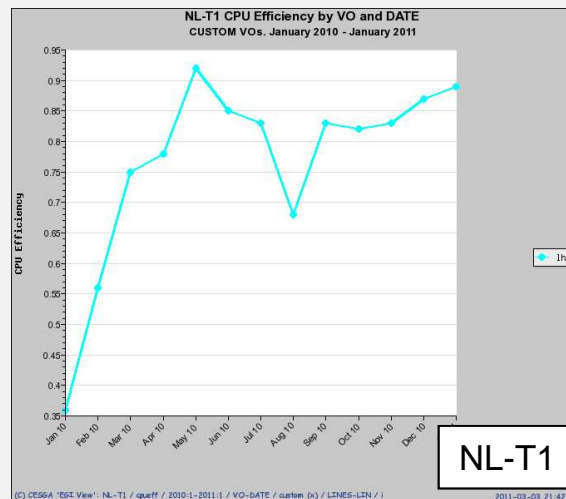
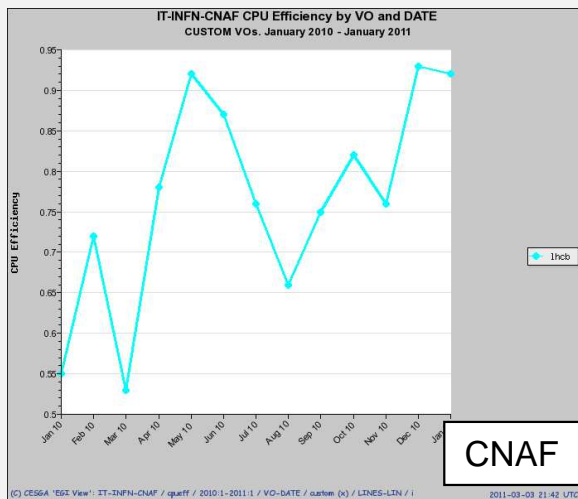
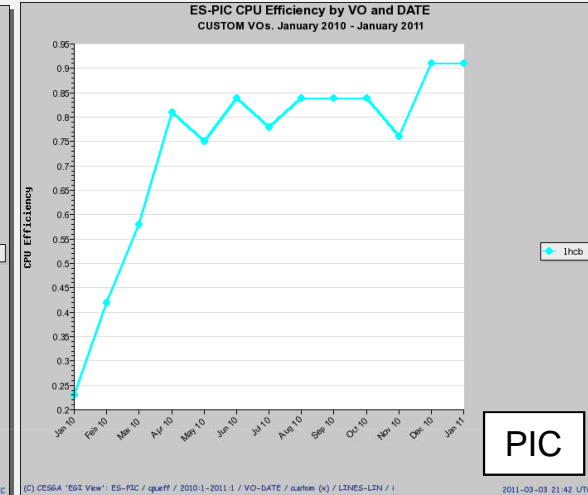
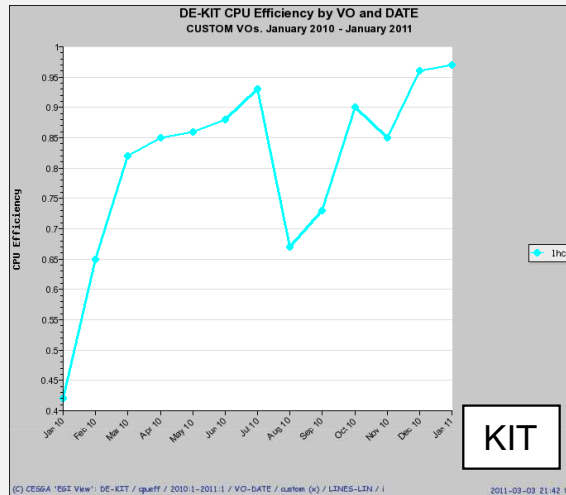
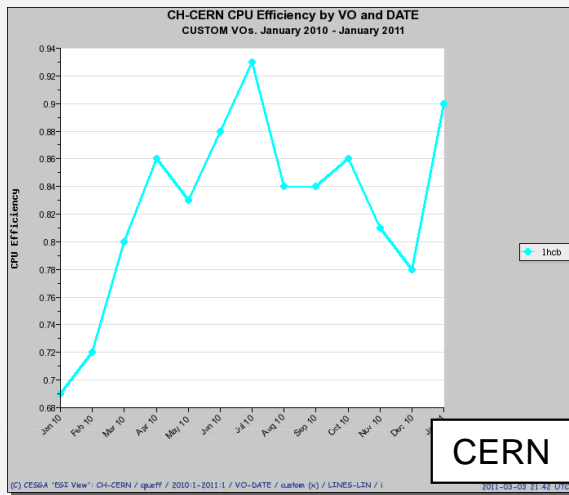


Backup slides



Other T1s CPU Efficiency

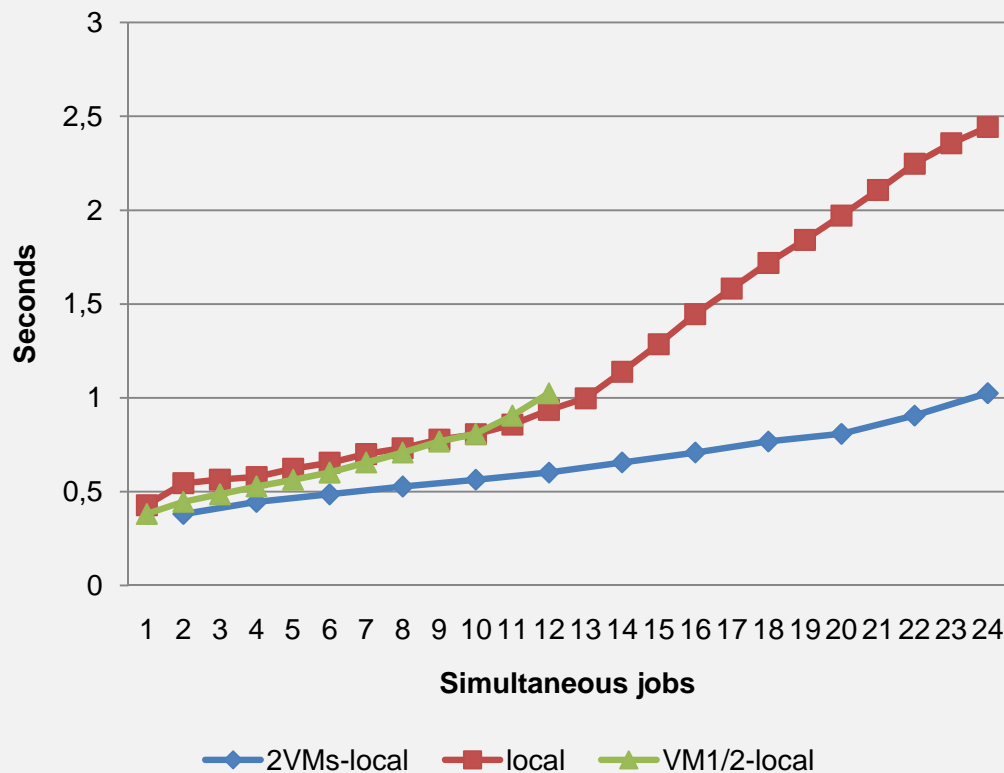
From January 2010 to January 2011



Source: http://www3.egee.cesga.es/gridsite/accounting/CESGA/tier1_view.html

Virtualized WN: divide and rule ?

2 Virt. WNs vs 1 Phys. WN Local FS



2 Virt. WNs vs 1 Phys. WN AFS

