GSIAF &

PoD

PROOF on Demand

Anar Manafov, Victor Penso, Carsten Preuss, and Kilian Schwarz,
GSI Darmstadt,

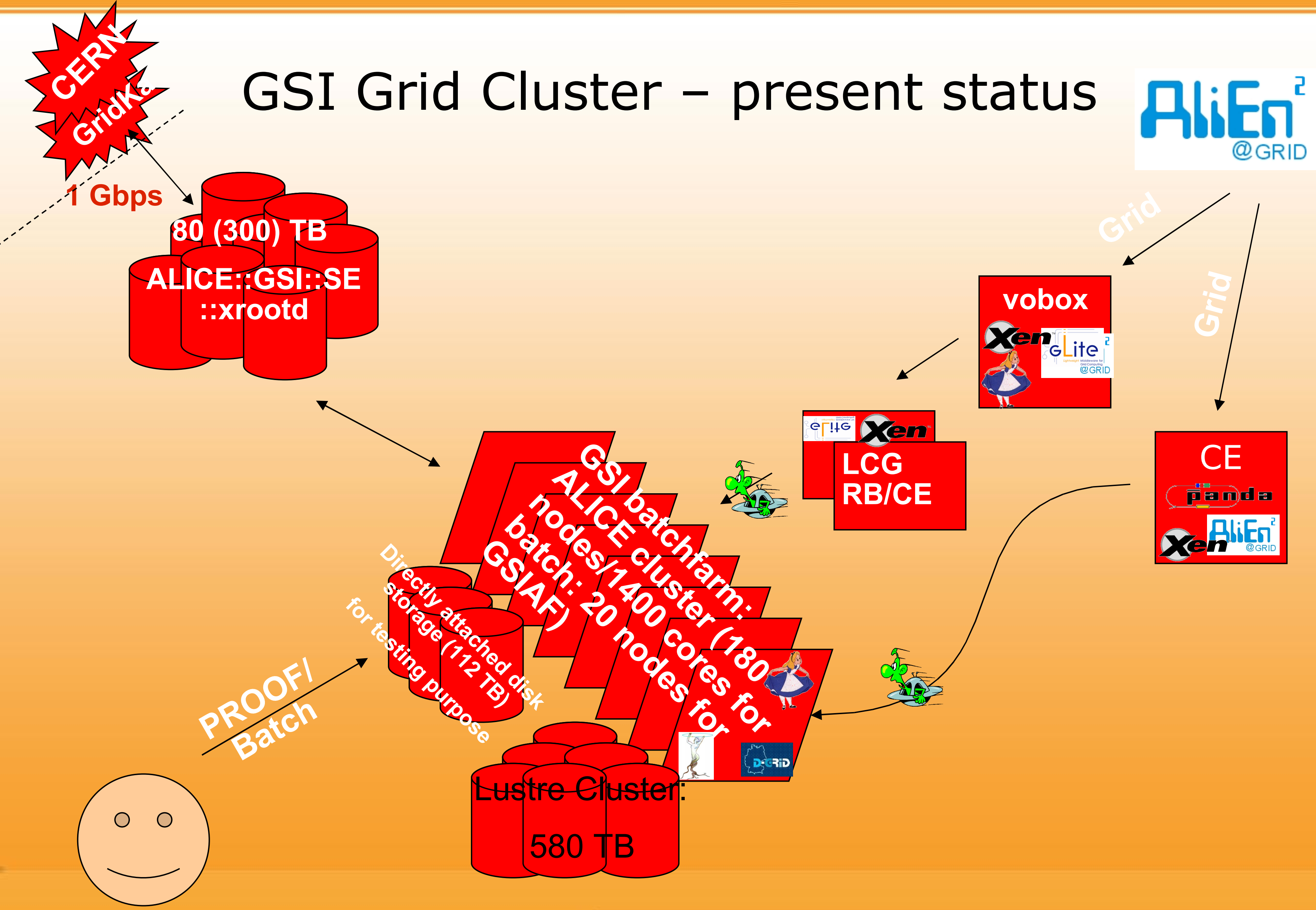
ALICE Offline week, 2009-06-25

v. 0.8

GSIAF

- Present status
- installation and configuration
- operation
- Issues and problems
- Plans and perspectives
- PoD
- summary

GSI Grid Cluster – present status



CERN

GridKa

1 Gbps

80 (300) TB

ALICE::GSI::SE ::xrootd

Grid

Grid

vobox

LCG RB/CE

CE

GSI batchfarm: ALICE cluster (180 nodes/1400 cores for batch: 20 nodes for GSIAF)

Directly attached disk storage (112 TB) for testing purpose

PROOF/ Batch

Lustre Cluster:

580 TB

GSI

Present Status

ALICE::GSI:SE::xrootd

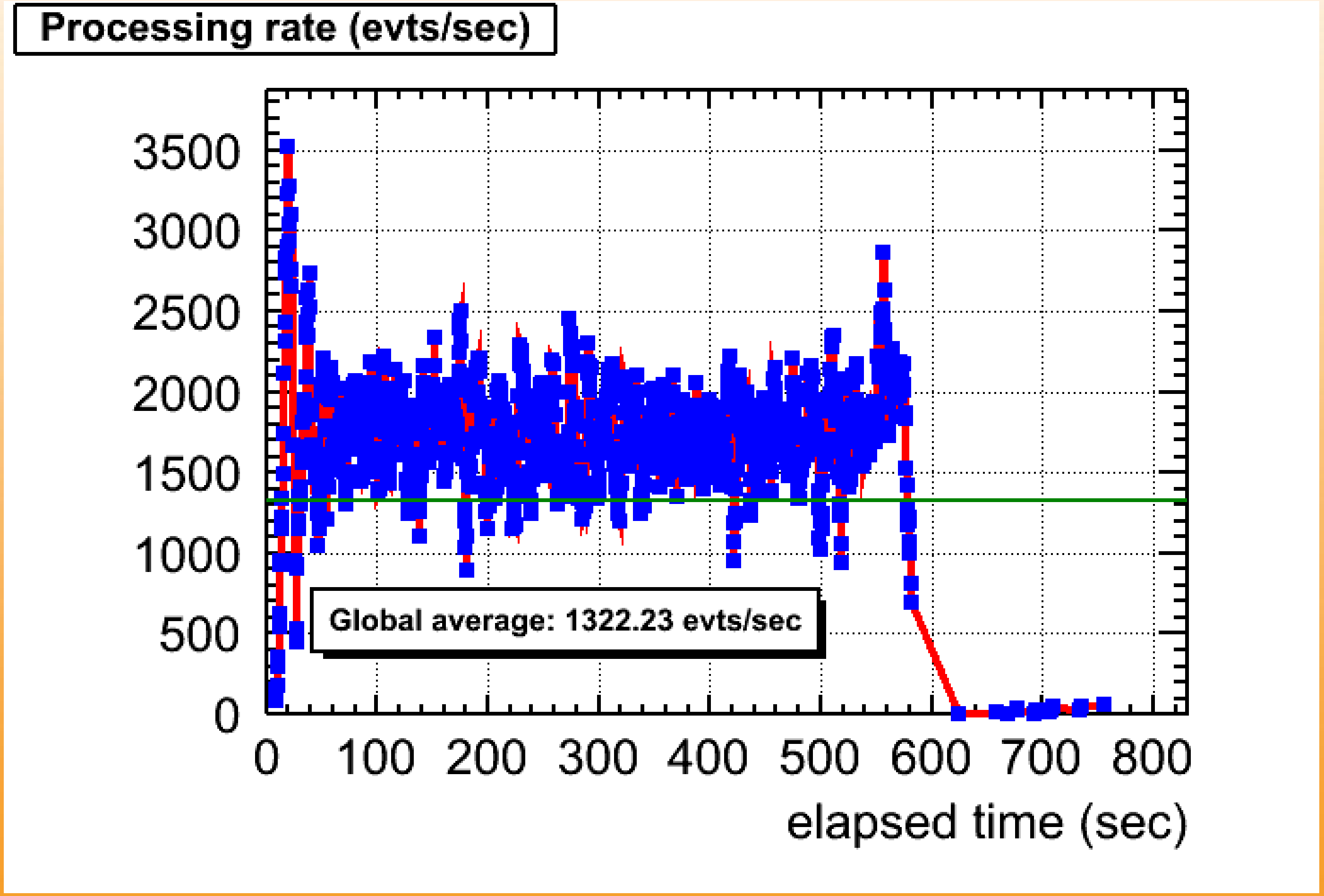
- 75 TB disk on fileserver (16 FS a 4-5 TB each)
    - currently being upgraded to 300 TB
        - 3U 12*500 GB disks RAID 5
        - 6 TB user space per server
- Lustre cluster
    - for local data storage. Directly mounted by the batch farm nodes
    - capacity: 580 TB (to be shared by all GSI experiments)

- nodes dedicated to ALICE (Grid and local)
- but used also by FAIR and Theory (less slots, lower prority)
- 15 boxes, 2*2 cores, 8 GB RAM, 1+3 disks, funded by D-Grid
- 40 boxes, 2*4 cores, 16 GB RAM, 4 disks RAID5, funded by ALICE
- 25 boxes, 2*4 cores, 32 GB RAM, 4 disks RAID5, funded by D-Grid
- 112 blades, 2*4 cores, 16 GB RAM, 2 disks RAID0, funded by ALICE
- ==> 192 computers, ca. 1500 cores
- on all nodes installed: Debian Etch64

GSIAF (static cluster)

- dedicated PROOF Cluster GSIAF
- 20 nodes (160 PROOF servers per user)
- Reads data from Lustre
- used very heavily for code debugging and development
  - needed for fast response
  - using large statistics
- Performance: see next slide

GSIAF performance (static cluster)



Alice first physics analysis
for pp @ 10 TeV

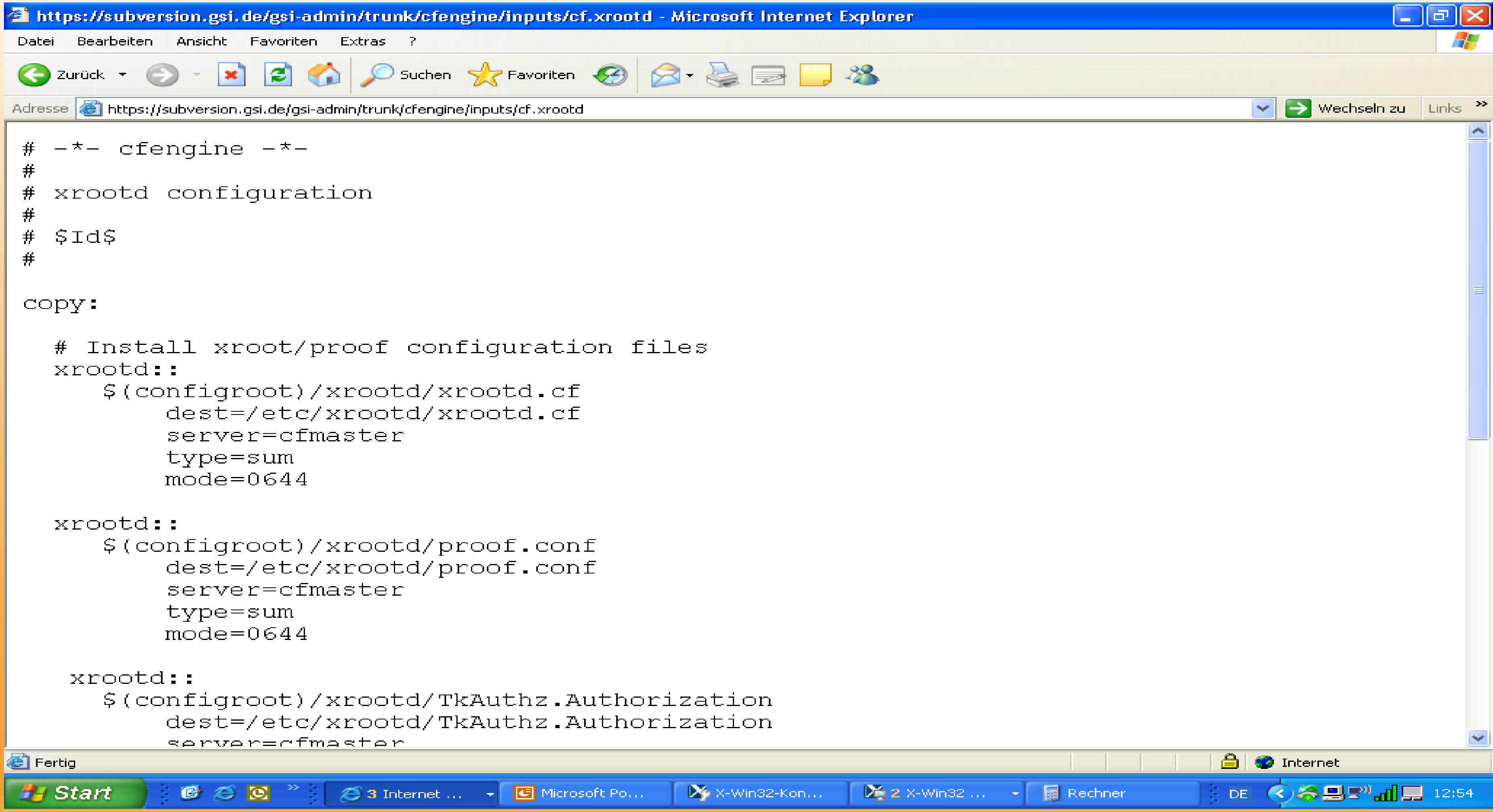
installation

- shared NFS dir, visible by all nodes
  - xrootd (version 2.9.0 build 20080621-0000)
  - ROOT (all recent versions up to 523-04)
  - AliRoot (including head)
  - all compiled for 64bit
- reason: due to fast software changes
- disadvantage: possible NFS stales
- started to build Debian packages of the used software to install locally
  - investigating also other methods (CfEngine) to distribute experiment software locally

Configuration (GSIAF)

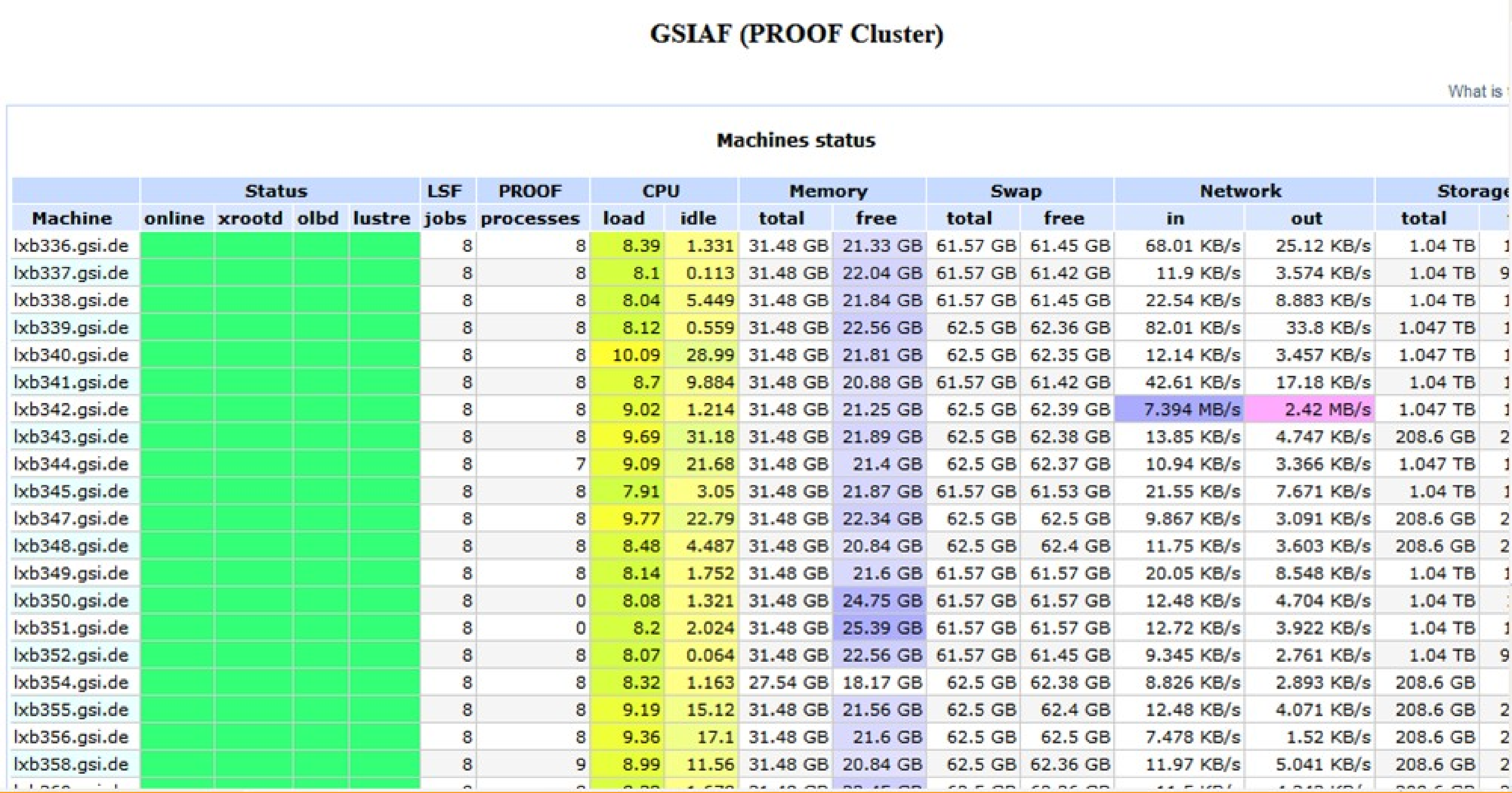
- setup: 1 standalone, high end 8 GB machine for xrd redirector and proof master, Cluster: LSF and proof workers
- so far no authentification/authorization
- via Cfengine
  - platform independent computer administration system (main functionality: automatic configuration).
- xrootd.cf, proof.conf, access control, Debian specific init scripts for start/stop of daemons (for the latter also Capistrano for fast prototyping)
- all configuration files are under version control (SVN)

Cfengine – config files in subversion



```
# -*- cfengine -*-
#
# xrootd configuration
#
# $Id$
#

copy:

  # Install xroot/proof configuration files
  xrootd::
     $(configroot)/xrootd/xrootd.cf
         dest=/etc/xrootd/xrootd.cf
         server=cfmaster
         type=sum
         mode=0644

  xrootd::
     $(configroot)/xrootd/proof.conf
         dest=/etc/xrootd/proof.conf
         server=cfmaster
         type=sum
         mode=0644

   xrootd::
     $(configroot)/xrootd/TkAuthz.Authorization
         dest=/etc/xrootd/TkAuthz.Authorization
         server=cfmaster
```

Monitoring via MonaLisa



GSIAF (PROOF Cluster)

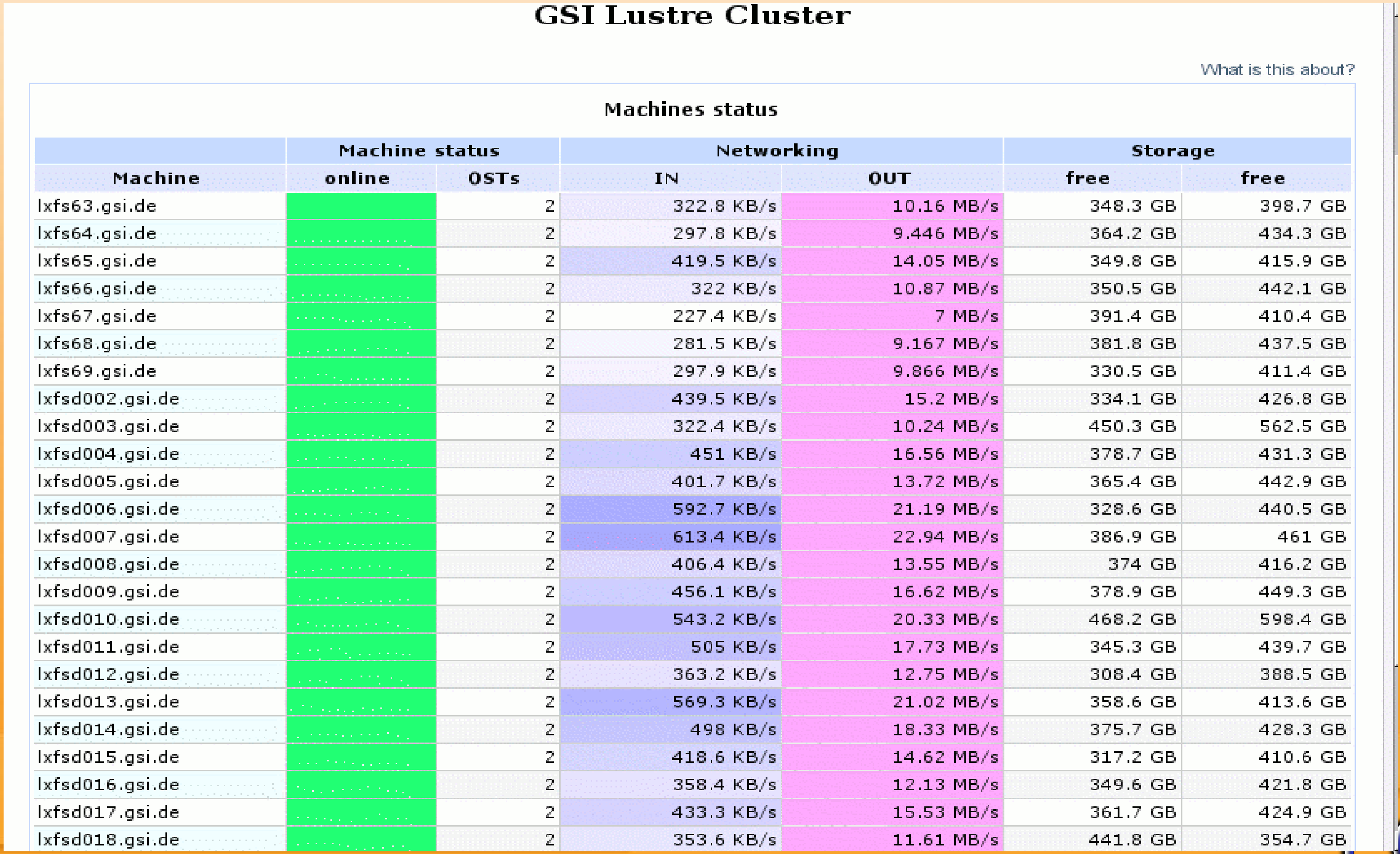
What is

Machines status

	Status				LSF	PROOF	CPU		Memory		Swap		Network		Storage	
Machine	online	xrootd	olbd	lustre	jobs	processes	load	idle	total	free	total	free	in	out	total	
lxb336.gsi.de	🟩	🟩	🟩	🟩	8	8	8.39	1.331	31.48 GB	21.33 GB	61.57 GB	61.45 GB	68.01 KB/s	25.12 KB/s	1.04 TB	1
lxb337.gsi.de	🟩	🟩	🟩	🟩	8	8	8.1	0.113	31.48 GB	22.04 GB	61.57 GB	61.42 GB	11.9 KB/s	3.574 KB/s	1.04 TB	9
lxb338.gsi.de	🟩	🟩	🟩	🟩	8	8	8.04	5.449	31.48 GB	21.84 GB	61.57 GB	61.45 GB	22.54 KB/s	8.883 KB/s	1.04 TB	1
lxb339.gsi.de	🟩	🟩	🟩	🟩	8	8	8.12	0.559	31.48 GB	22.56 GB	62.5 GB	62.36 GB	82.01 KB/s	33.8 KB/s	1.047 TB	1
lxb340.gsi.de	🟩	🟩	🟩	🟩	8	8	10.09	28.99	31.48 GB	21.81 GB	62.5 GB	62.35 GB	12.14 KB/s	3.457 KB/s	1.047 TB	1
lxb341.gsi.de	🟩	🟩	🟩	🟩	8	8	8.7	9.884	31.48 GB	20.88 GB	61.57 GB	61.42 GB	42.61 KB/s	17.18 KB/s	1.04 TB	1
lxb342.gsi.de	🟩	🟩	🟩	🟩	8	8	9.02	1.214	31.48 GB	21.25 GB	62.5 GB	62.39 GB	7.394 MB/s	2.42 MB/s	1.047 TB	1
lxb343.gsi.de	🟩	🟩	🟩	🟩	8	8	9.69	31.18	31.48 GB	21.89 GB	62.5 GB	62.38 GB	13.85 KB/s	4.747 KB/s	208.6 GB	2
lxb344.gsi.de	🟩	🟩	🟩	🟩	8	7	9.09	21.68	31.48 GB	21.4 GB	62.5 GB	62.37 GB	10.94 KB/s	3.366 KB/s	1.047 TB	1
lxb345.gsi.de	🟩	🟩	🟩	🟩	8	8	7.91	3.05	31.48 GB	21.87 GB	61.57 GB	61.53 GB	21.55 KB/s	7.671 KB/s	1.04 TB	1
lxb347.gsi.de	🟩	🟩	🟩	🟩	8	8	9.77	22.79	31.48 GB	22.34 GB	62.5 GB	62.5 GB	9.867 KB/s	3.091 KB/s	208.6 GB	2
lxb348.gsi.de	🟩	🟩	🟩	🟩	8	8	8.48	4.487	31.48 GB	20.84 GB	62.5 GB	62.4 GB	11.75 KB/s	3.603 KB/s	208.6 GB	2
lxb349.gsi.de	🟩	🟩	🟩	🟩	8	8	8.14	1.752	31.48 GB	21.6 GB	61.57 GB	61.57 GB	20.05 KB/s	8.548 KB/s	1.04 TB	1
lxb350.gsi.de	🟩	🟩	🟩	🟩	8	0	8.08	1.321	31.48 GB	24.75 GB	61.57 GB	61.57 GB	12.48 KB/s	4.704 KB/s	1.04 TB	[illegible]
lxb351.gsi.de	🟩	🟩	🟩	🟩	8	0	8.2	2.024	31.48 GB	25.39 GB	61.57 GB	61.57 GB	12.72 KB/s	3.922 KB/s	1.04 TB	1
lxb352.gsi.de	🟩	🟩	🟩	🟩	8	8	8.07	0.064	31.48 GB	22.56 GB	61.57 GB	61.45 GB	9.345 KB/s	2.761 KB/s	1.04 TB	9
lxb354.gsi.de	🟩	🟩	🟩	🟩	8	8	8.32	1.163	27.54 GB	18.17 GB	62.5 GB	62.38 GB	8.826 KB/s	2.893 KB/s	208.6 GB	
lxb355.gsi.de	🟩	🟩	🟩	🟩	8	8	9.19	15.12	31.48 GB	21.56 GB	62.5 GB	62.4 GB	12.48 KB/s	4.071 KB/s	208.6 GB	2
lxb356.gsi.de	🟩	🟩	🟩	🟩	8	8	9.36	17.1	31.48 GB	21.6 GB	62.5 GB	62.5 GB	7.478 KB/s	1.52 KB/s	208.6 GB	2
lxb358.gsi.de	🟩	🟩	🟩	🟩	8	9	8.99	11.56	31.48 GB	20.84 GB	62.5 GB	62.36 GB	11.97 KB/s	5.041 KB/s	208.6 GB	2

GSI Luster Clustre

directly attached to PROOF workers
580 TB (to be shared by all GSI experiments)
used for local data storage



GSI Lustre Cluster

What is this about?

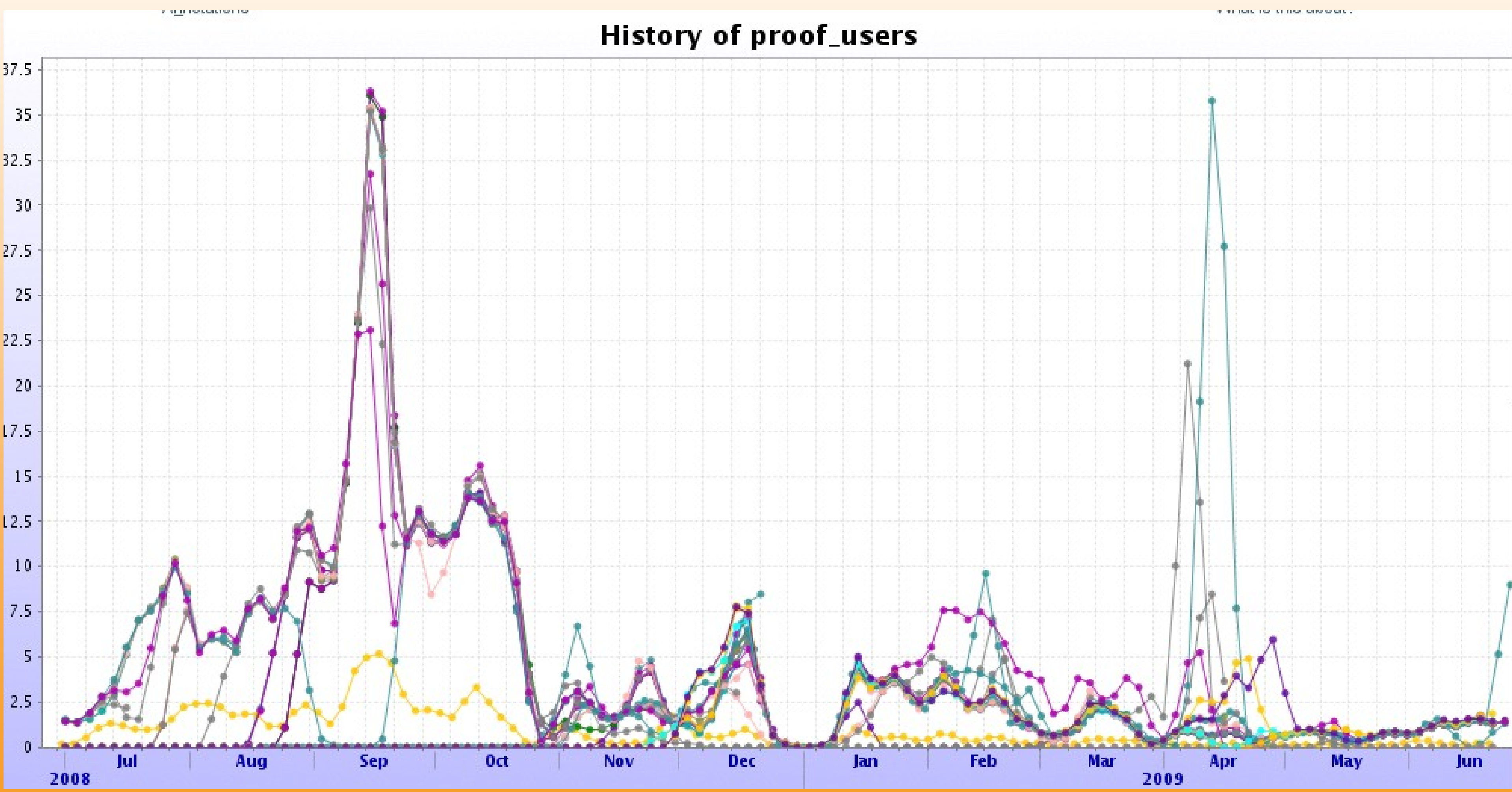
Machines status

	Machine status		Networking		Storage	
Machine	online	OSTs	IN	OUT	free	free
lxfs63.gsi.de	🟩	2	322.8 KB/s	10.16 MB/s	348.3 GB	398.7 GB
lxfs64.gsi.de	🟩	2	297.8 KB/s	9.446 MB/s	364.2 GB	434.3 GB
lxfs65.gsi.de	🟩	2	419.5 KB/s	14.05 MB/s	349.8 GB	415.9 GB
lxfs66.gsi.de	🟩	2	322 KB/s	10.87 MB/s	350.5 GB	442.1 GB
lxfs67.gsi.de	🟩	2	227.4 KB/s	7 MB/s	391.4 GB	410.4 GB
lxfs68.gsi.de	🟩	2	281.5 KB/s	9.167 MB/s	381.8 GB	437.5 GB
lxfs69.gsi.de	🟩	2	297.9 KB/s	9.866 MB/s	330.5 GB	411.4 GB
lxfsd002.gsi.de	🟩	2	439.5 KB/s	15.2 MB/s	334.1 GB	426.8 GB
lxfsd003.gsi.de	🟩	2	322.4 KB/s	10.24 MB/s	450.3 GB	562.5 GB
lxfsd004.gsi.de	🟩	2	451 KB/s	16.56 MB/s	378.7 GB	431.3 GB
lxfsd005.gsi.de	🟩	2	401.7 KB/s	13.72 MB/s	365.4 GB	442.9 GB
lxfsd006.gsi.de	🟩	2	592.7 KB/s	21.19 MB/s	328.6 GB	440.5 GB
lxfsd007.gsi.de	🟩	2	613.4 KB/s	22.94 MB/s	386.9 GB	461 GB
lxfsd008.gsi.de	🟩	2	406.4 KB/s	13.55 MB/s	374 GB	416.2 GB
lxfsd009.gsi.de	🟩	2	456.1 KB/s	16.62 MB/s	378.9 GB	449.3 GB
lxfsd010.gsi.de	🟩	2	543.2 KB/s	20.33 MB/s	468.2 GB	598.4 GB
lxfsd011.gsi.de	🟩	2	505 KB/s	17.73 MB/s	345.3 GB	439.7 GB
lxfsd012.gsi.de	🟩	2	363.2 KB/s	12.75 MB/s	308.4 GB	388.5 GB
lxfsd013.gsi.de	🟩	2	569.3 KB/s	21.02 MB/s	358.6 GB	413.6 GB
lxfsd014.gsi.de	🟩	2	498 KB/s	18.33 MB/s	375.7 GB	428.3 GB
lxfsd015.gsi.de	🟩	2	418.6 KB/s	14.62 MB/s	317.2 GB	410.6 GB
lxfsd016.gsi.de	🟩	2	358.4 KB/s	12.13 MB/s	349.6 GB	421.8 GB
lxfsd017.gsi.de	🟩	2	433.3 KB/s	15.53 MB/s	361.7 GB	424.9 GB
lxfsd018.gsi.de	🟩	2	353.6 KB/s	11.61 MB/s	441.8 GB	354.7 GB

GSIAF usage experience

- real life analysis work of staged data by GSI ALICE group (1-4 concurrent users)
- 2 user tutorials for GSI ALICE users (10 students each training)
- GSIAF and CAF were used as PROOF clusters during PROOF course at GridKa School 2007 and 2008

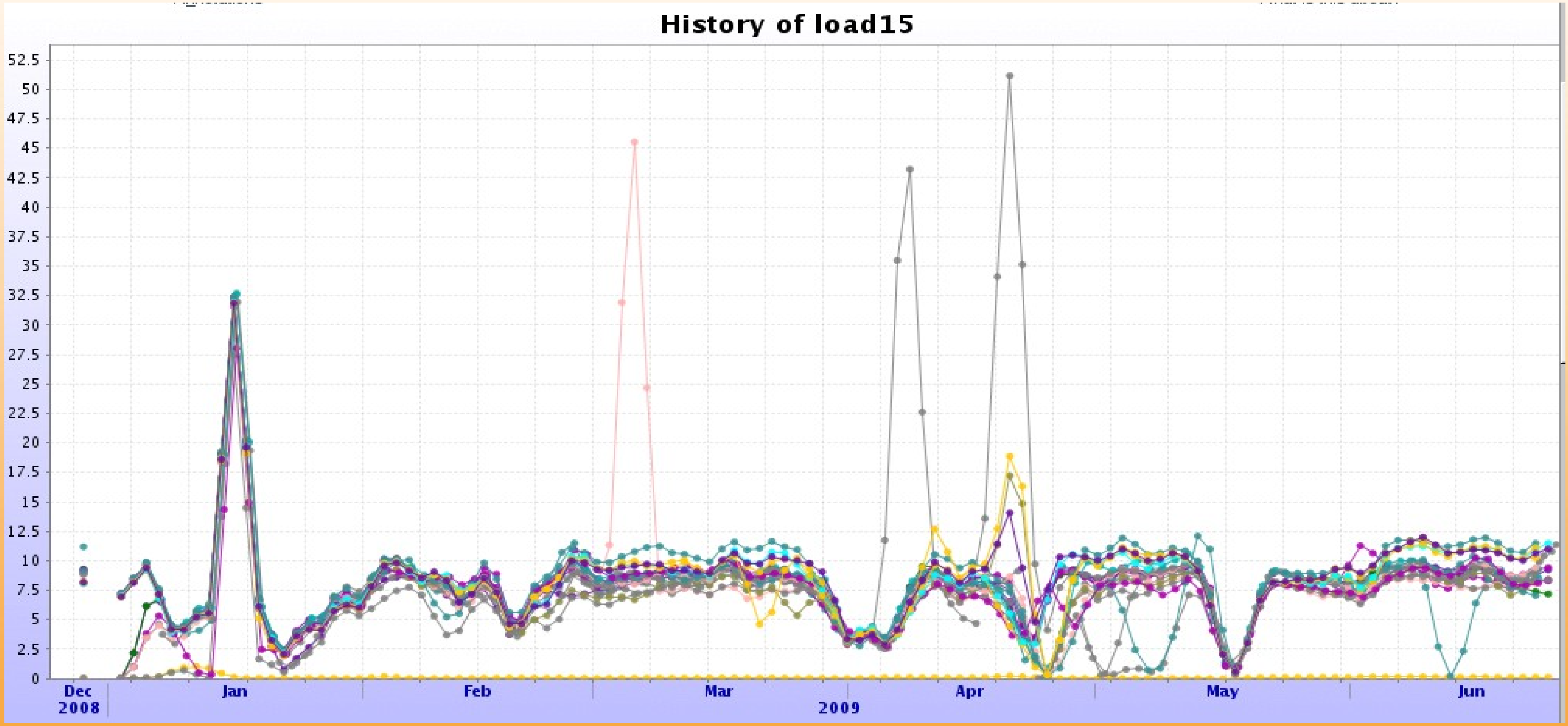
PROOF users at GSIAF (sc)



ideas came true …

- Coexistence of interactive and batch processes (PROOF analysis on staged data and Grid user/production jobs) on the same machines can be handled !!!
  - re”nice” LSF batch processes to give PROOF processes a higher priority (LSF parameter)
  - number of jobs per queue can be increased/decreased
  - queues can be enabled/disabled
  - jobs can be moved from one queue to other queues
- Currently at GSI each PROOF worker is an LSF batch node
- optimised I/O. Various methods of data access (local disk, file servers via xrd, mounted lustre cluster) have been investigated systematically. Method of choice: Lustre and xrd based SE. Local disks are not used for PROOF anymore at GSIAF.
  - PROOF nodes can be added/removed easily
  - local disks have no good surviving rate
- extend GSI T2 and GSIAF according to promised ramp up plan
  - PoD

cluster load



Cluster is sufficiently and homogenously loaded.

data staging and user support

- how to bring data to GSIAF ?
- used method:
  - GSI users copy data from AliEn individually to Lustre using standard tools
- user support at GSIAF:
  - private communication or help yourself via established procedures.

general remarks

- it is not planned to use more than 160 cores for GSIAF static cluster
- for less than 160 cores per PROOF session users use PoD

issues and problems

- currently the cluster is automatically restarted every night. Otherwise the system behaves stable. Manual intervention rarely needed.

wishlist, summary and outview

Overall: PROOF users at GSI are happy !!!
PROOF is definitely usable, current PROOF capacity is fine (recently no request for more PROOF nodes), performance is doing well
...

future plans and perspectives

- no major developments, upgrades or enlargements planned at GSIAF static cluster
- static cluster will slowly phase out
- users will be encouraged to use dynamic PROOF setup on GSI's batch system (PoD)
- PoD will improve and mature during a certain time of parallel existence with static cluster
- without local data on individual machines a static cluster is no must
- dynamic PROOF cluster is easier to handle, saves administration time and is fully under the control of the users
- PoD will be the system of choice and future of GSIAF !!! (a dedicated master machine to handle PoD PROOF sessions has been set up)

GSIAF -----> PoD

„a PROOF cluster on the fly"

Summary

- GSI T2 resources have been extended according to the plan
- GSIAF is heavily used and behaves stable
- static cluster will phase out
- PROOF on Demand shall provide the possibility to create dynamic and private/individual PROOF clusters „on the fly"
  - First official release of LSF plugin April 09
  - First user training in same month