Alice T1/T2 Tutorial
Site Experience
Subatech – Nantes – France
IN2P3-SUBATECH

Brief presentation

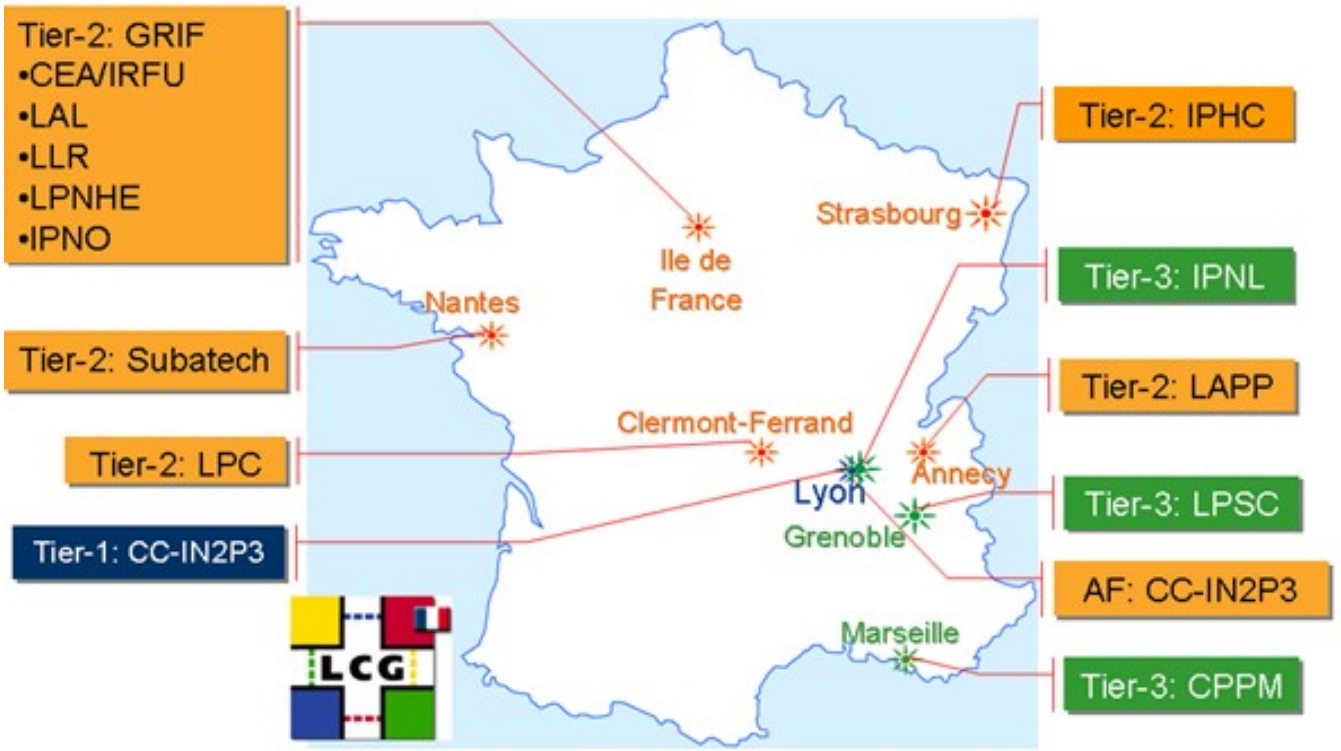
- The SUBATECH Laboratory is one of the 20 units of the French National Physics Institute : IN2P3
- It is located in Nantes, Britanny, West of France

Brief presentation

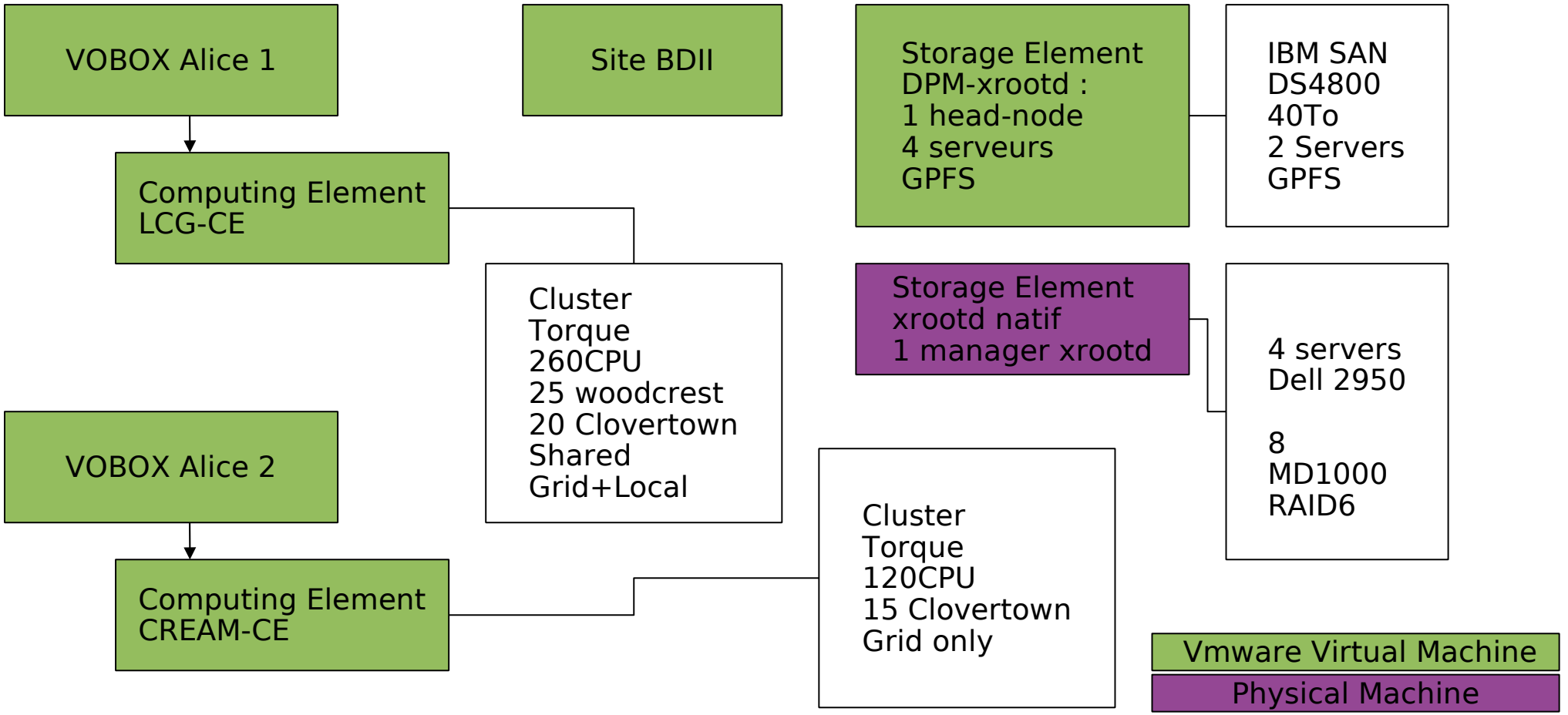
- Hosted by the Ecole des Mines de Nantes (EMN)
- Joint research unit (UMR) between EMN, Nantes University and the CNRS Institute IN2P3
- More than 150 people
- IT service : 5

LCG and Alice sites in France

- CCIN2P3 (T1+T2+AF)
- Clermont
- GRIF (IRFU,IPNO)
- Nantes SUBATECH
- Strasbourg IRES
- Lyon IPNL
- Grenoble LPSC

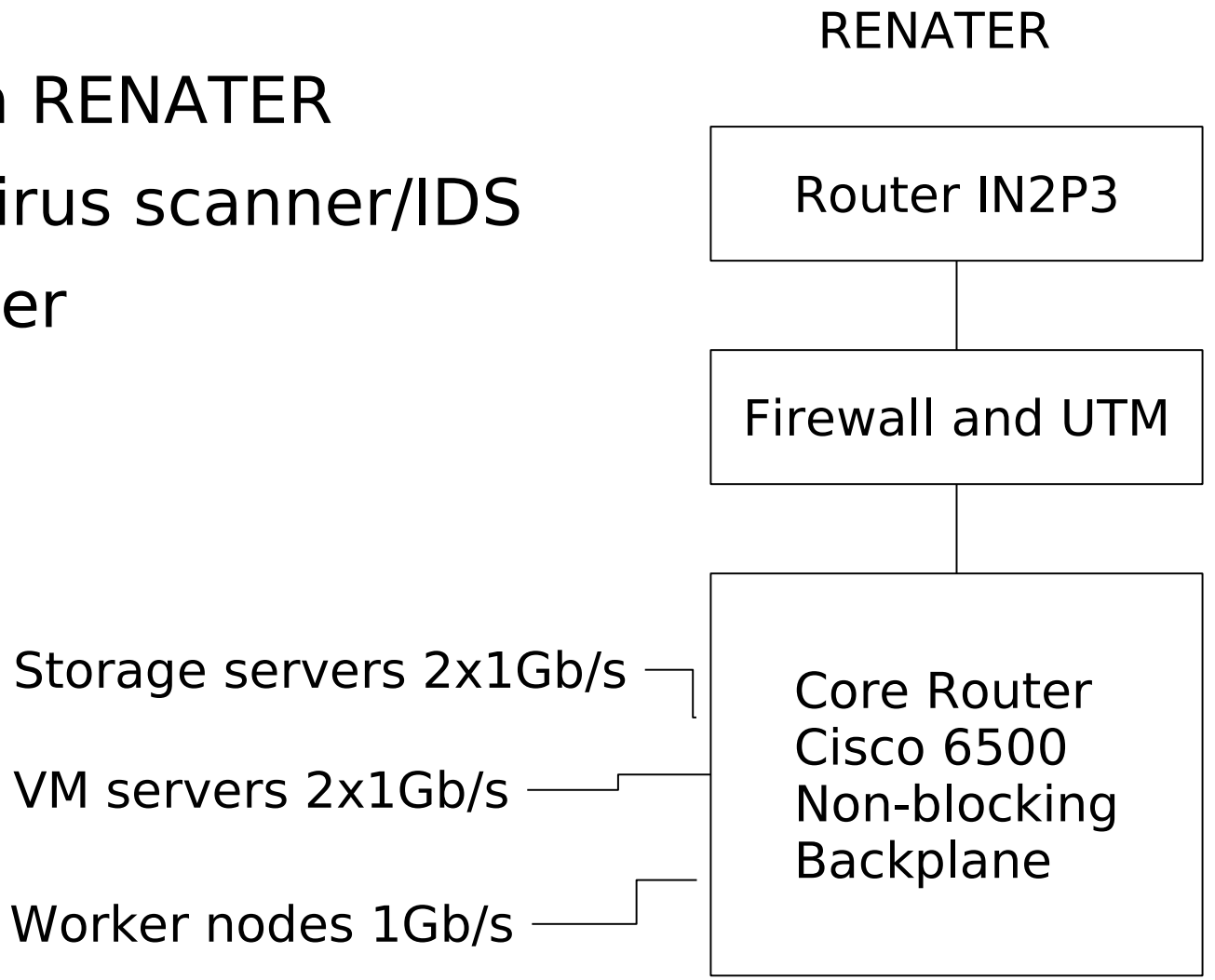


SUBATECH LCG Services and hardware



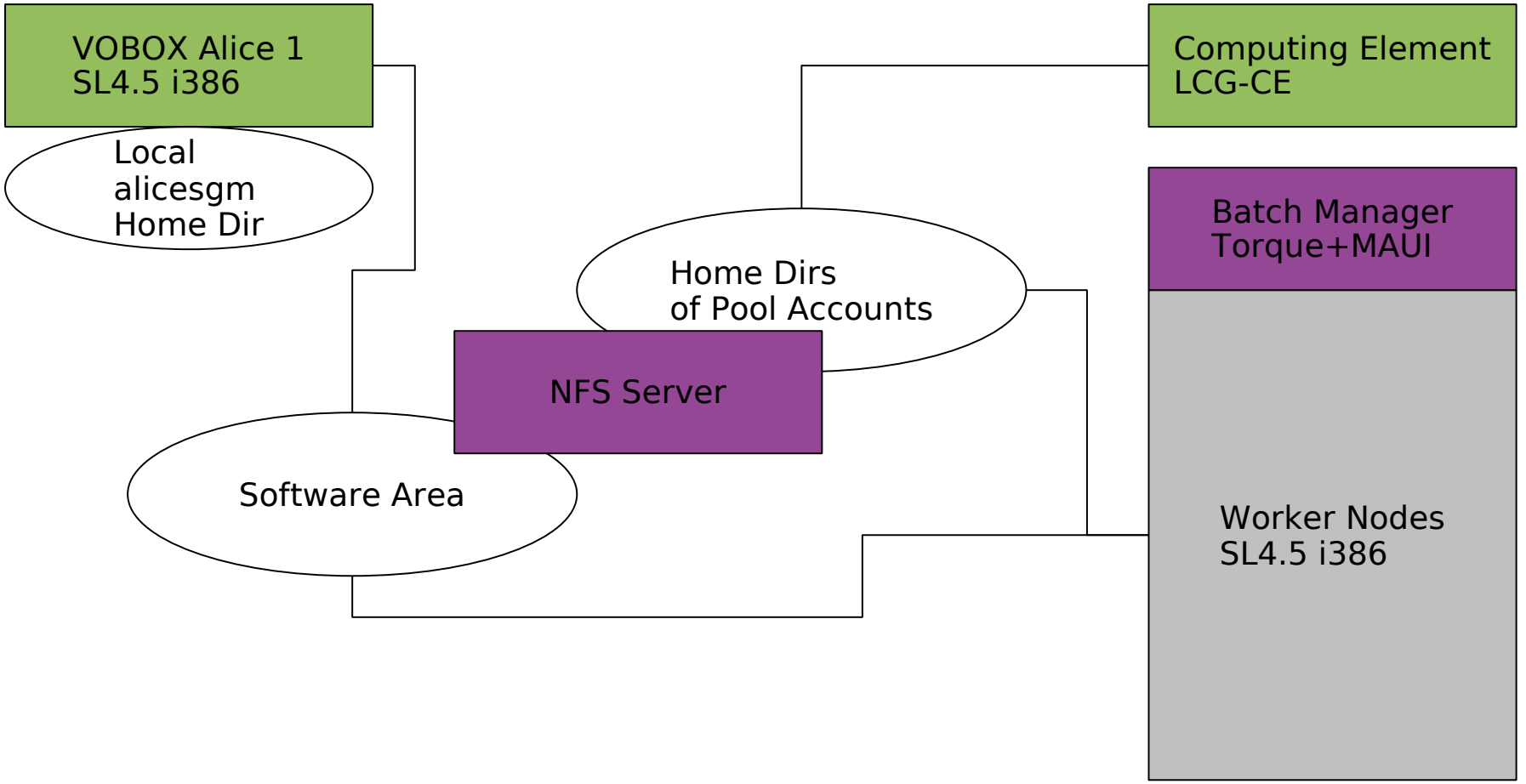
Network

- 1Gbit/s on RENATER
- Firewall/Virus scanner/IDS
- Core Router



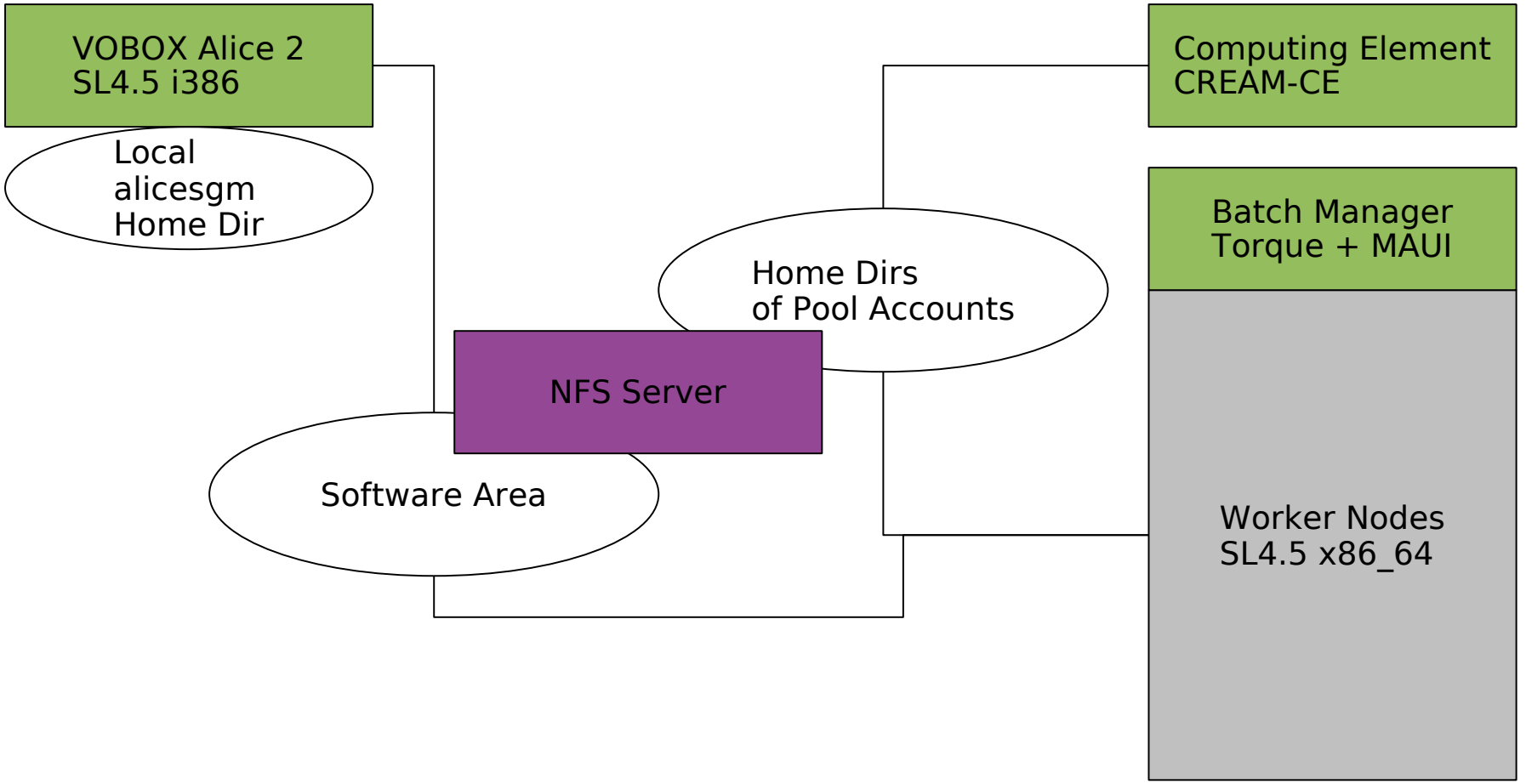
LCG-CE Computing Element

- 260 SL4.5/i386 cores for Alice



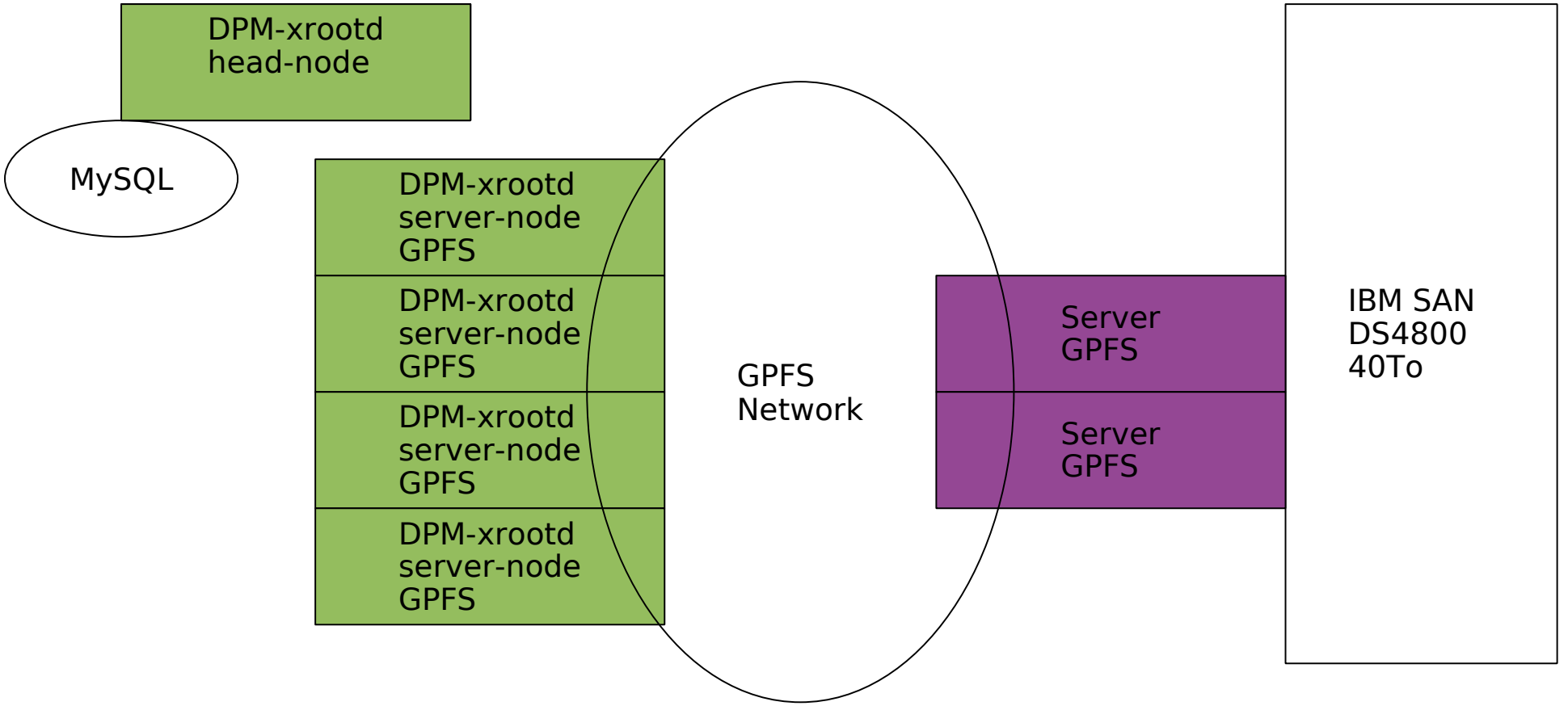
CREAM-CE Computing Element

- 120 cores SL4.5/x86_64 for Alice (separate cluster)



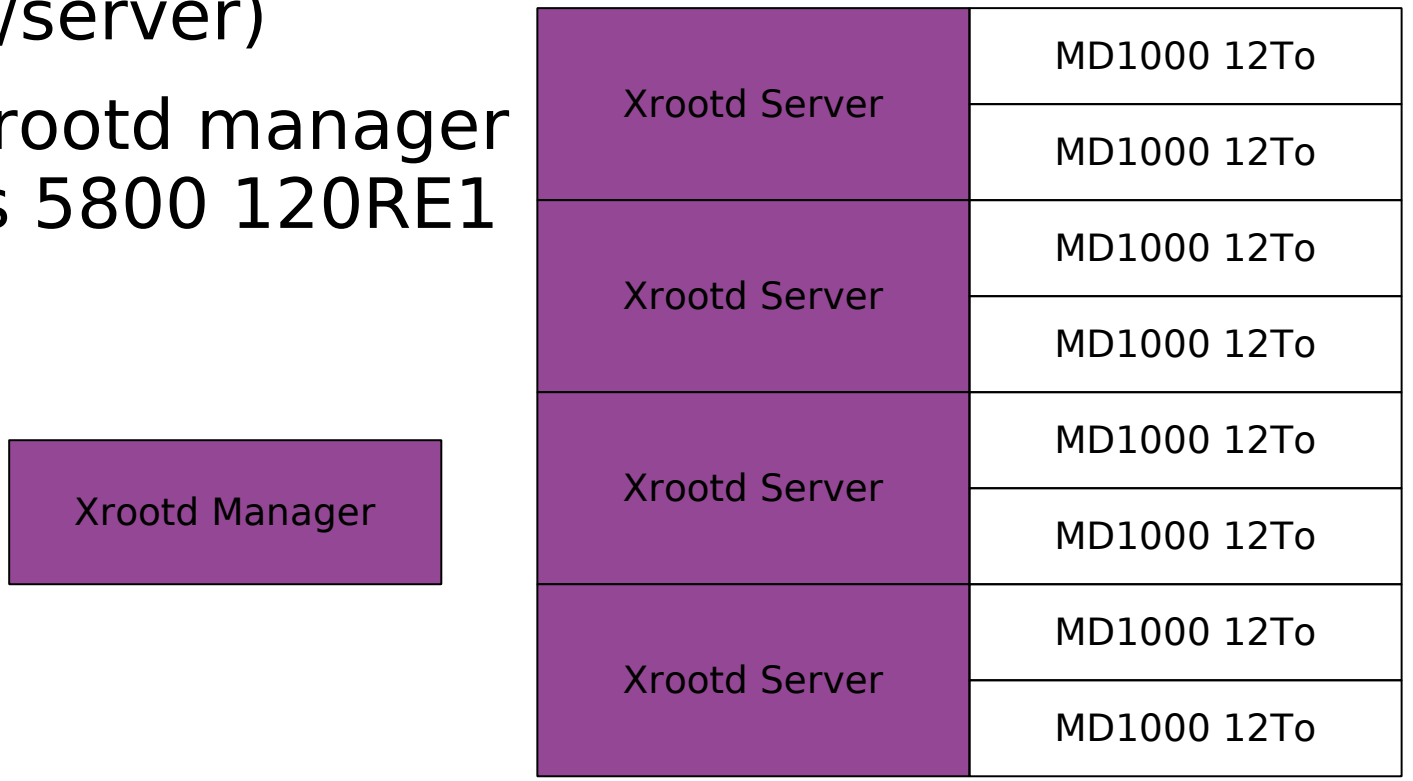
DPM-xrootd Storage

- 40To SAN storage, 4 GPFS partitions



Native xrootd Storage

- 100To in 4 DAS cells each composed of 1 Dell 2950 server attached to 2 MD1000 in RAID6 (4 partitions/server)
- Dedicated xrootd manager NEC Express 5800 120RE1



Site Experience

- DPM-xrootd
- Native xrootd
- CREAM-CE
- AliEn Torrent

DPM-xrootd

- Took time to be available and to set up
- Integration xrootd with DPM not complete :
  - Xrootd usage gets mapped to the DPM root principal
  - Used/free space incorrectly reported by the GIP
  - Used/free space not automatically reported by MonaLisa
- Sometimes break
  - The xrootd daemon stalls on servers and have to be restarted (automatic cron script may help)
  - High load involving the MySQL daemon in case of high access rate on the manager (a virtual machine here is probably a bad choice and it may require better designed hardware)

Native xrootd

- Easy to setup
- How to deploy ?
  - Manually ?
  - RPM is OK but have to understand dependencies on OS, architecture and other software in order to share RPMs
  - RPM requires install under the root account while install under dedicated”xrootd” account is usually preferred
  - Can be deployed using Quattor if RPM is used
- Storage under native xrootd not accounted by EGEE/LCG
- Seems rock solid but not enough experience yet
  - For example : how data is spread over 4 filesystems

CREAM-CE

- Relatively easy to setup
  - YUM install
  - Had to extract the BLAH server from RPMs to install on the batch manager (which is separate)
  - Had to manually setup APEL publication
- Still new
  - Glite update broke the CREAM-CE in january 2009, a bug in gLexec configuration was discovered
  - Need to master TomCat ? (security, management, debugging)
- Not yet tested by SAM
- Same Job profile between LCG-CE and CREAM-CE

AliEn Torrent

- Successfully tested but needed adjustments :
  - The port 6881 originally used is filtered for outgoing connections on most IN2P3 sites (old P2P fighting)
  - Our UTM firewall have a policy to block it, have to put an exception
- May raise questions and concerns
  - A poll on french sites about their feelings on the usage of this tool resulted in comments that still have to be summarized (mostly on efficiency and usage of network resources, not on security)
- What about distributing software for users this way ?

Building Quality

- Infrastructure
- Resilience/Redundancy
- Incident detection
- Monitoring
- Security
- Change management

Infrastructure

- Room : ~60m2
- Air Conditioning
  - 2 units, 1 can keep the room cool enough depending on outside temperature
- Power
  - 2 different 50Kw lines with UPS on each used at 50%
  - Almost all hardware except worker nodes have dual power, one on each line
  - Add specific device for boxes with only one power supply
  - Worker nodes have only one power supply and are balanced on the 2 lines

Resilience - Redundancy

- In case of problems, have the service survive (Resilience) or have a replacement ready (Redundancy)
- Resilience
  - Dual power on different lines help the machine stay powered in case of problem on one line. But for this you need each line to be able to power all hardware on which vital services rely
  - Vmware vMotion : if one of the VM server goes down, virtual machines can be moved to the other (here also capacity matters)
- Redundancy
  - Have a second machine ready, on-line or not

Incident detection

- Better know before things really break
- Infrastructure
  - Temperature monitoring with alarms (mail and automatic phone call or pager)
  - UPS monitoring with alarms
- Hardware and OS
  - Monitoring of important Unix parameters : CPU,load,RAM
  - Disk space
- Services
  - Daemons
  - Specific probes (see next slide)

Incident detection

- Nagios specific probes for the Grid Services
  - CE : APEL accounting synchronisation
  - CE : Inactive jobs (CPU time not increasing over 30')
  - CREAM-CE : Tomcat daemon
  - DPM-xrootd-disk : Daemons (GPFS, gridftp, xrootd)
  - DPM-xrootd-head : Daemons (DPM,DPNS,Mysql,xrootd
  - Xrootd : Daemons listening
  - Worker nodes : local disk switching in RO mode, PBS Mom

Example : Nagios for CREAM-CE

Service Status Details For Host 'nanlcg04'

Host	Service	Status	Last Check	Duration	Attempt	Status Information
nanlcg04	Glite-LB-logd	OK	05-22-2009 08:44:36	24d 15h 26m 51s	1/3	Glite-LB-Logd OK: Daemon is running. Host is listening.
	LDAP	OK	05-22-2009 08:39:59	0d 20h 9m 8s	1/3	LDAP OK - 0,105 secondes de temps de réponse
	NTP	OK	05-22-2009 08:35:21	10d 18h 1m 47s	1/4	NTP OK: Offset -0.000441 secs, jitter 3.010 msec, peer is stratum 6
	PING	OK	05-22-2009 08:42:44	24d 1h 3m 46s	1/3	PING OK - Paquets perdus = 0%, RTA = 0.18 ms
	SSH	OK	05-22-2009 08:43:06	24d 21h 55m 44s	1/3	SSH OK - OpenSSH_3.9p1 (protocole 1.99)
	TOMCAT	OK	05-22-2009 08:43:29	24d 15h 27m 59s	1/3	TOMCAT_SERVER OK: Daemon is running. Host is listening.
	Unix_CPU	OK	05-22-2009 08:43:51	24d 20h 23m 32s	1/3	OK: CPU is 3% used
	Unix_DISK	OK	05-22-2009 08:44:14	24d 20h 54m 1s	1/3	DISK OK - free space: /boot 198 MB (79%):
	Unix_LOAD	OK	05-22-2009 08:44:36	24d 20h 53m 38s	1/3	OK - load average: 0.04, 0.28, 0.67
	Unix_PROCS	OK	05-22-2009 08:39:59	23d 16h 49m 3s	1/3	PROCS OK: 115 processes
	Unix_TRIPWIRE	OK	05-22-2009 08:40:21	6d 15h 30m 49s	1/3	Tripwire OK - violation found: 0
	Unix_ZOMBIE	OK	05-22-2009 08:42:44	24d 20h 57m 31s	1/3	PROCS OK: 0 processes with STATE = Z
	bdii	OK	05-22-2009 08:43:06	24d 20h 37m 26s	1/3	TCP OK - 0,000 second response time on port 2170
	grid-ftp	OK	05-22-2009 08:43:29	24d 20h 37m 4s	1/3	TCP OK - 0,000 second response time on port 2811
	inactive_jobs	WARNING	05-22-2009 08:33:51	4d 13h 39m 47s	3/3	Bad Jobs on nanpbs: 13153 13166 13205
	mysql	OK	05-22-2009 08:44:14	24d 20h 39m 19s	1/3	TCP OK - 0,000 second response time on port 3306

Automatic actions

- Early detection of incidents makes it possible to take action, automatically in some cases
  - Temperature crossing a certain threshold => halt part of the worker nodes. Ultimate protection can be enforced by cutting all power if a (high) threshold is reached.
  - Same if the power is down and the UPS batteries become low
  - Need to distinguish between vital and less vital services
  - Restart daemons that have crashed or are stuck (DPM)
  - Switch to a spare server (heartbeat)

Monitoring the Site

- Apart from incident detection which raise an alarm, the Site Manager can also keep an eye on the Services :
  - MonaLisa Monitoring
  - Nagios graphs
  - logs and tools available
  - Network usage and traffic (CC netstat, Extra)

Change management

- Changes break things, but cannot be avoided
  - Glite updates, OS updates,
- One should always be able to backtrack
  - Virtual machines : revert to snapshot
  - Quattor : redeploy previous profiles
  - Backup or disk image
- We have to keep track of changes
  - Logbook : elog or something similar

Communication with the VO

- The Site Manager needs to know :
  - What is the expected behaviour of the services he runs
  - If the site does not behave well from the VO point of view
  - What are the requirements of the VO and their evolution along time
- The Site Manager have to notice the VO :
  - In case of downtimes or services disruption
  - If he thinks something is wrong

Open Questions and Issues

- Tracing Jobs
  - Understanding AliEn job numbers and translate into the job numbers in the local batch system, looking at output
- LCG-CE to CREAM transition
  - Keep one cluster (how to do it cleanly) or split the cluster
- Job Efficiency : understanding errors
- Blocked jobs : to be understood
- Xrootd : how to provide redundency ?
- Accounting in ML : move to SPEC-HEP06
- What about support of final users ?

Perspectives for SUBATECH

- Test SL5 on worker nodes
- Upgrade network connection to RENATER from 1Gbits/s to 10Gbits/s
- Build a test Analysis Facility based on PROOF
- Decommission the DPM-xrootd storage

Conclusion

- The quality of service for the grid-site is strongly linked to the quality of infrastructure (room, network, monitoring) and the whole site policy for this matter
- Supporting only one LHC VO has probably permitted more involvement to debug problems and test new solutions for Alice

References

- SUBATECH : http://www-subatech.in2p3.fr/
- IN2P3 : http://www.in2p3.fr/
- LCG-FR : http://lcg.in2p3.fr/
- Nagios : http://www.nagios.org/
- Elog : https://midas.psi.ch/elog/
- Quattor : http://quattor.org/