ALICE and gLite

Patricia Méndez Lorenzo

WA lecture, CERN 17-04-09

OVERVIEW

- Recall of the ALICE Computing Model
- VOBOXES
- ALICE Storage Model and solutions
- ALICE transfer system
- CREAM-CE

THE ALICE PHYSICS DATA CHALLENGES: GOALS

Running in continuous PDC mode since 2006

- Validation of the LCG/gLite workload management services
  - **Stability** of the services is fundamental for the entire duration of the exercise
- Validation of the data transfer and storage services
  - The **stability** and **support** of the services have to be assured beyond the throughput tests
- Validation of the ALICE distributed reconstruction and calibration model
- Integration of all Grid resources within one single – interfaces to different Grids (LCG, OSG, NDGF)
- End-user data analysis
- Full Dress Rehearsal and CCRC`08 during ALICE FDR

THE ALICE IMPLEMENTATION: ALIEN

- Own Task queue and related services
  - Pull Model service: a server holds a master queue of jobs and it is up to the CE that provides the CPU cycles. It asks for the jobs
- Use of the WLCG-WMS for agent submission
- Several Grid infrastructures available since the PDC06
  - Use of AliEn as a general front-end
  - LCG, OSG, NDGF
  - Lots of resources but different middleware
- Use high-level tools and APIs to access Grid resources
  - Developers put a lot of abstraction effort into hiding the complexity and shielding the user from implementation changes

ALICE IN THE WORLD

Maps, sites and services status monitored via MonaLisa: http://pcalimonitor.cern.ch

COMPUTING MODEL - PP

Tape T0

Alien FC

To tape

To Grid FC

CERN T0

RAW

Disk buffer T0

Pass 0 reco

Calibration

Calibration parameters

To T1s (FTS)

CAF Analysis

T2s

MC Data (xrootd)

T1s

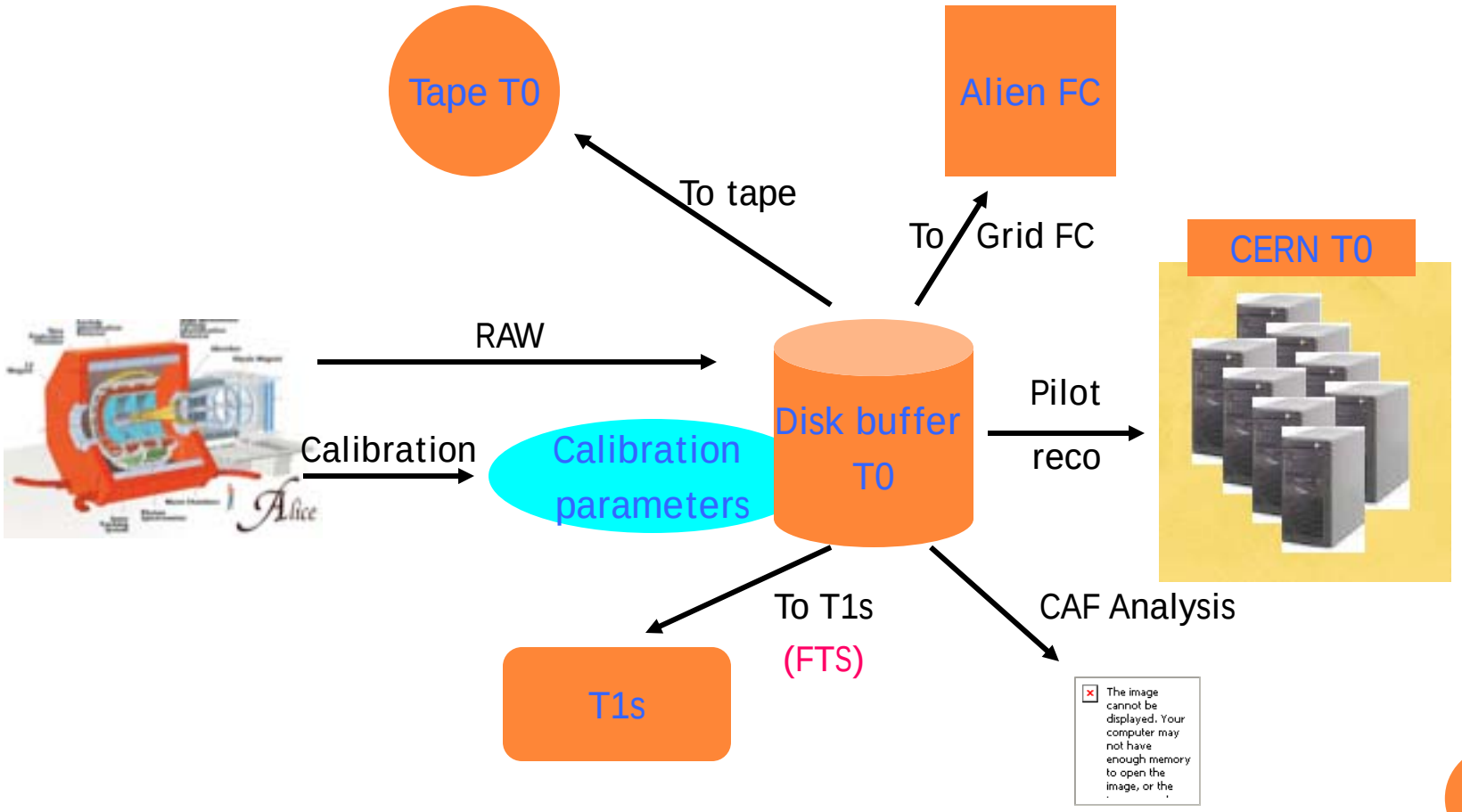
Pass 1&2 reco

Ordered analysis

End-user analysis

The image cannot be displayed. Your computer may not have enough memory to open the image, or the

COMPUTING MODEL - HI DATA TAKING



COMPUTING MODEL - LHC SHUTDOWN

Tape T0

From tape

Alien FC

CERN T0

Disk buffer T0

Pass 0 reco

To T1s

CAF Analysis

T2s

MC Data (xrootd)

End-user analysis

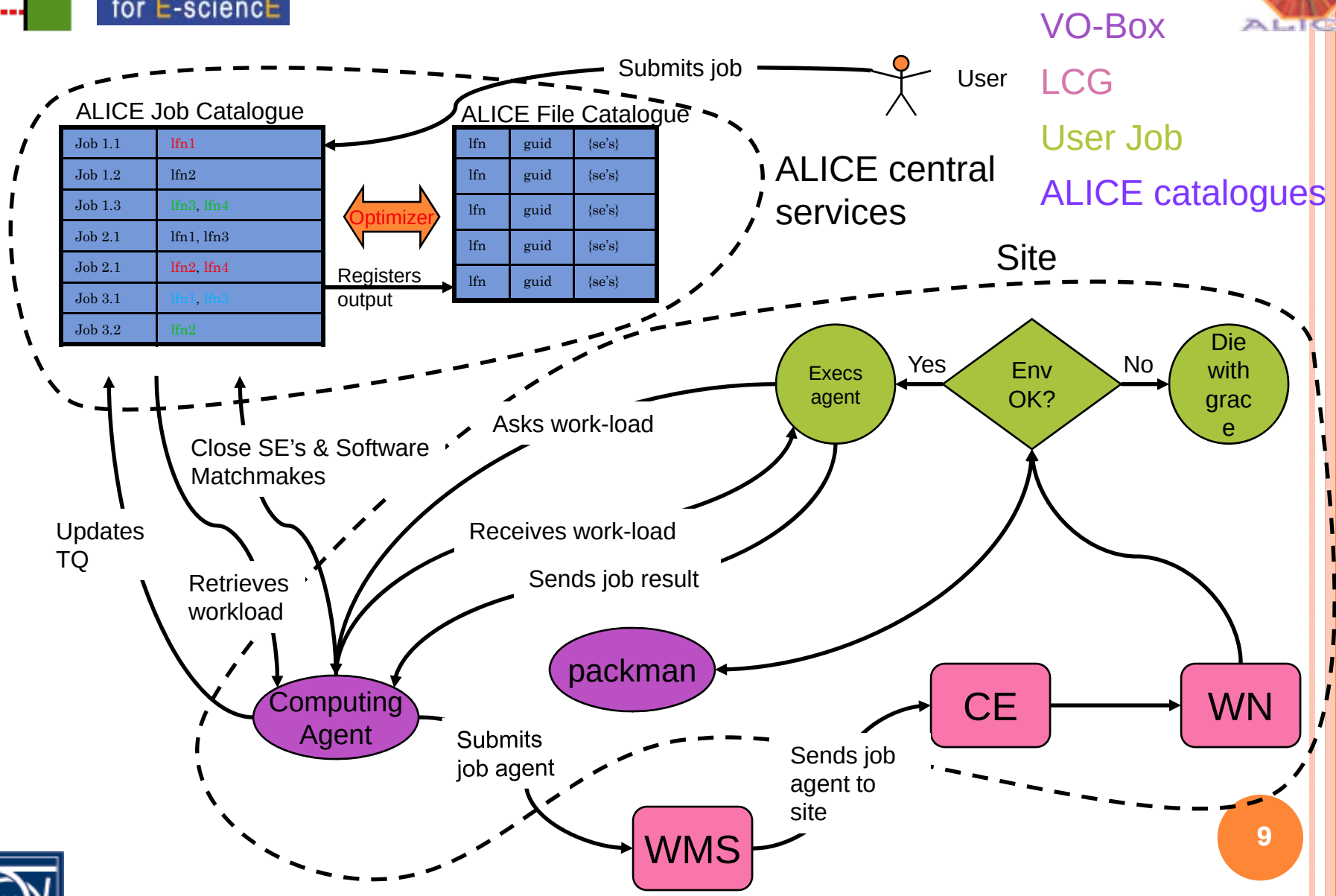
T1s

Pass 1&2 reco

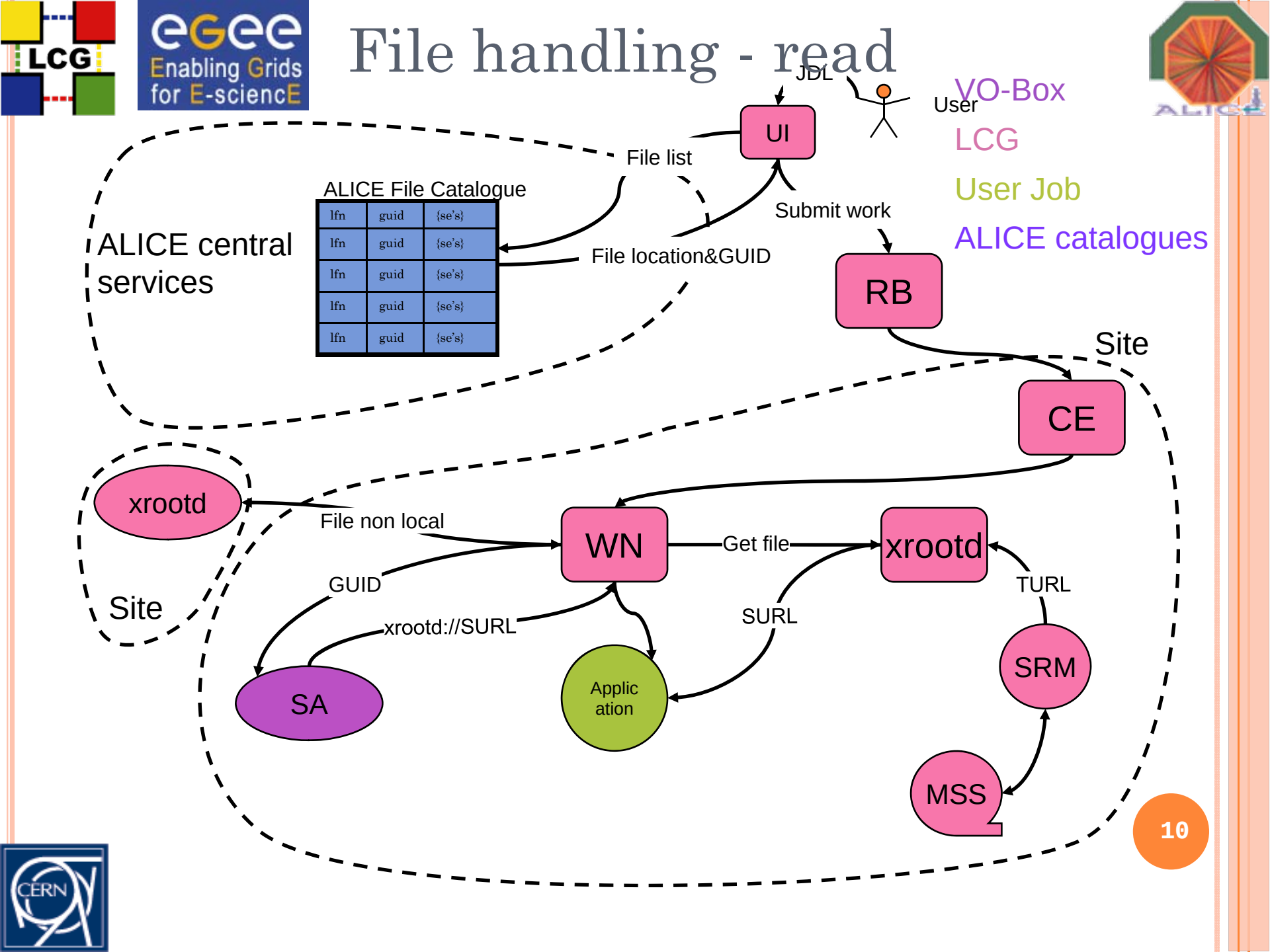
Ordered analysis

The image cannot be displayed. Your computer may not have enough memory to open the image, or the

Job Submission Structure



File handling - read



PRINCIPLE OF OPERATIONS: VOBOXES

- VO-boxes deployed at all T0-T1-T2 sites providing resources for ALICE
  - ***Mandatory requirement to enter the production***
  - Required in addition to all standard LCG Services
  - Entry door to the LCG Environment
  - Runs standard LCG components and **ALICE specific services**
- Uniform deployment
  - Same behavior for T1 and T2 in terms of production
  - Differences between T1 and T2 a matter of QoS only
- Installation and maintenance entirely ALICE responsibility
  - Based on a regional principle
  - Set of ALICE experts matched to groups of sites
- Site related problems handled by site administrators
- LCG Service problems reported via GGUS
  - Not too much, ALICE has experts in almost all sites

REQUIREMENTS FOR THE ALICE VOBOX

- WLCG Configuration
  - The VOBOX is the integration point to the WLCG
  - FULL IMPLEMENTATION OF THE WLCG-UI
  - FTS Client Services
- ALICE Configuration
  - The specific requirements have been included in a public document: **VOBOX Security and Operations Questionnaires v-0.5**
  - Distributed to all site managers before setting up
- **Support for the whole Production**
  - Regional experts handle the VOBOXES
  - Who is who perfectly established in most of the sites
  - Central support placed at CERN

ALICE Services in the VO-BOX

- Site Computing Element
  - Interfaces different WMS (LCG, OSG, etc)
  - It performs the matchmaking with the ALICE TQ
  - This service in principle could be run centrally but scaling and size problems of the CERN ALICE installation would appear
- Agent Monitoring Service
  - Control of all VO-BOX services
  - As long as the VO-BOC exists, this service will run there

ALICE Services in the VO-BOX

- PackMan
  - It distributes, installs and configures the software needed by ALICE jobs
  - Includes versioning and test tools
  - If a job needs a certain software to run, this service automatically install it before pulling it from the TQ
  - It is a complex and reliable service needed by the ALICE architecture
  - It writes the software under VO_ALICE_SW_DIR
    - It must therefore run in the VOBOX to access this area

ALICE Services in the VO-BOX

- MonaLisa
  - It monitors the job status, the storage and the traffic
  - It includes specific ALICE monitoring
  - ALICE would like to have it as BS service following other experiment initiatives
  - It is installed in the VOBOX because:
    - It runs specific LCG test suites which must be executed from the VOBOX (monitoring of the VO-BOX itself)
    - If a previous local aggregation is performed it will minimize the monitoring traffic

ALICE Services in the VO-BOX

- Storage Adaptor
  - Handles the communication with LFC to translate GUID to TURL/SURL
  - Builds the TURL
  - Can act as a volume manager
  - Starts up xrootd services
  - Handles communication with FTD
  - Monitor site storage configuration
  - It must run in the VOBOX because
    - To avoid communication with the central service (catalog)
    - Need to communicate with the local LFC
    - Need to be local to start xrootd and to monitor the storage

ALICE Services in the VO-BOX

- xrootd
  - Posix I/O
  - Insulates application from local storage systems
  - Efficient handling of storage
  - Handles user-level file authorization
  - It runs in the VOBOX because:
    - It needs to communicate with the local SE
  - An official requirement of ALICE to include it in the middleware ongoing

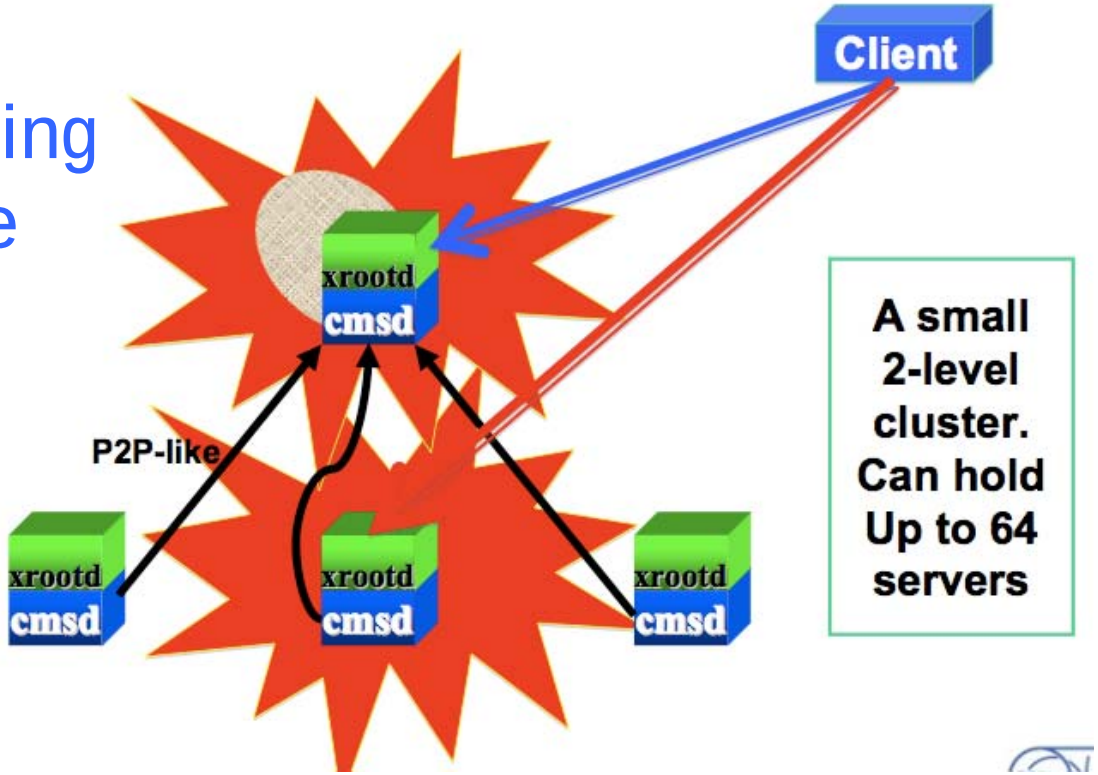
STORAGE SOLUTIONS AND CURRENT STATUS

- ALICE requires the **xrootd protocol** (I/O and transfers) interfaced to all Storage Systems
- dCache – stable
  - CCIN2P3, GridKA, NDGF, NL-T1
- Castor2 - stable, most heavily used
  - CERN, CNAF
  - RAL
- dCache and DPM at T2s - stable, used for MC production and user analysis

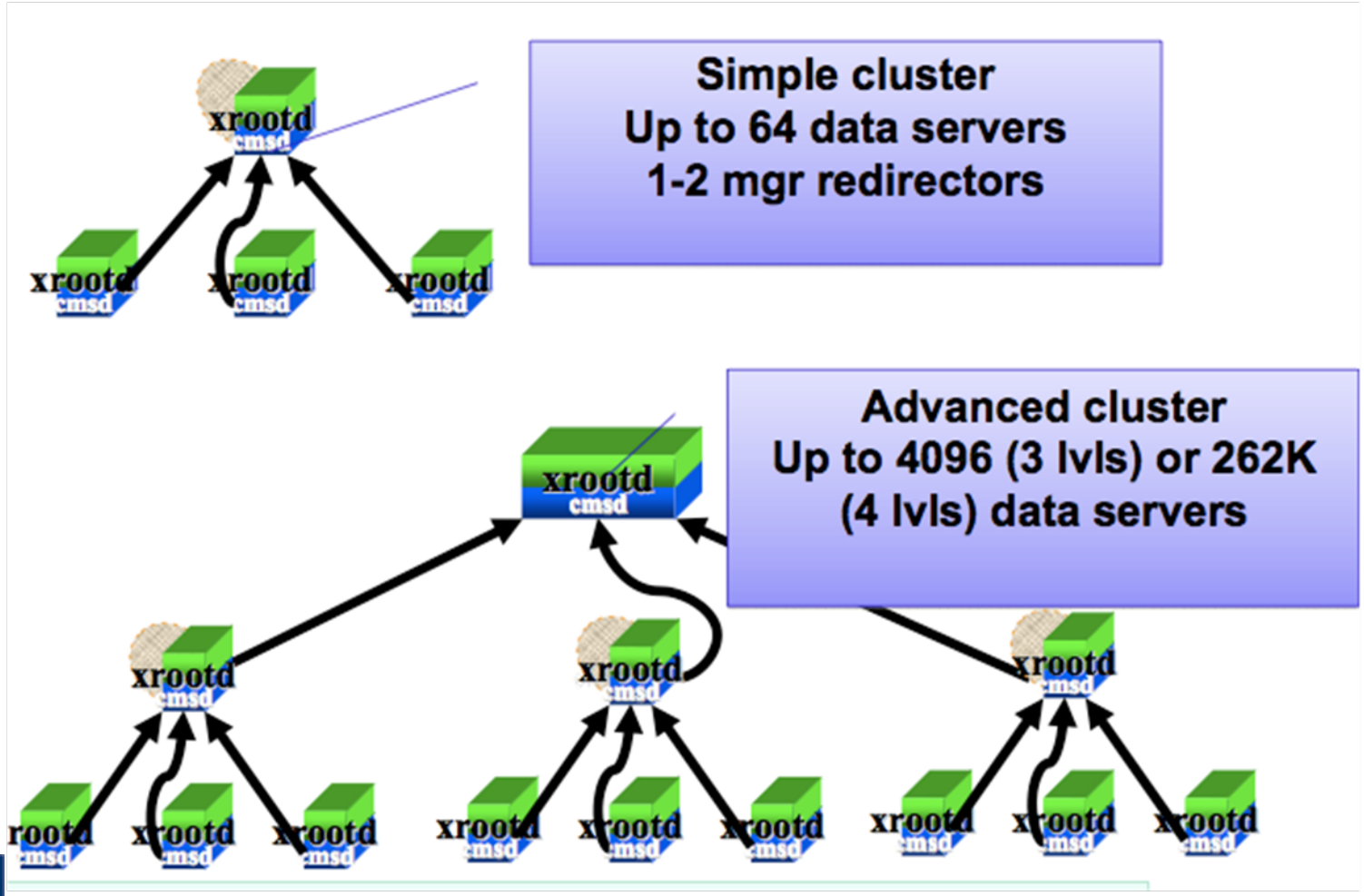
SOME INFORMATION ABOUT XROOT (I)

- Xrootd is the protocol chosen by ALICE to access data (I/O) and also as transfer protocol between T1 and T2 sites

Basic working principle



SOME INFORMATION ABOUT XROOT (II)



Monitor Tools

- We are controlling the status of the transfers with different tools
  - MonaLisa controls all FTD status
  - Dashboard follows the FTS errors
  - The status of the transfers are fully monitored also in the VO-BOXES through the FTD logs
- All problems have been reported inmediatly using GGUS
- Good support of the SC Experts

ALICE MONITORING SYSTEM

VO Box machine status

What is this about?

Machine status (last hour average values)

		CPU										Mem [% MB]		Swap [% MB]		Eth0 [KB/s]			Eth1 [KB/s]		Eth2 [KB/s]		System
Site name	Last see online	Load5	User	System	IOWait	Int	SoftInt	Nice	Steal	Idle	Cnt	MHz	Usage	Total	Usage	Total	In	Out	In	In	Out	Procs	Socks
1. Aalborg	2007-02-01 09h	0.92	4.799	0.833	7.023	0.024	0.316	0	0	87	2	2793	77	2026	12.67	3121	8.181	8.038	55.49	2.278	-	-	169	144
2. Athens	2007-01-29 21h	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3. Bari	2007-01-22 11h	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4. Birmingham	2007-02-01 09h	0.201	7.804	1.719	0.166	0.014	0.29	0	-	90.01	2	800	55	1001	7.468	2047	13.54	34.98	-	-	-	-	94	39
5. BITP	2007-02-01 09h	0.411	1.026	0.431	12.28	0.018	0.215	0	-	86.03	4	3192	23	3999	0	8181	41.03	43.81	11.14	2.092	-	-	94	91
6. Bologna	2007-02-01 09h	0.063	1.208	0.479	0.077	0.003	0.114	0	-	98.12	4	3067	19	4005	0	2000	-	-	17.35	7.977	-	-	94	43
7. Cagliari	2007-02-01 09h	0.058	2.867	1.089	0.092	0.015	0.273	0	-	95.66	2	3199	34	2007	0	2000	22.28	11.75	-	-	-	-	90	55
8. Catania	2007-02-01 09h	0.202	3.447	0.965	0.725	0.017	0.17	0	-	94.68	4	2799	38	2006	0	4094	32.47	18.77	-	-	-	-	85	52
9. CCIN2P3	2007-02-01 09h	16.76	24.77	54.19	1.643	0.035	0.254	0	-	19.11	4	3000	51	2007	12.41	2047	1.683	0.093	32.87	21.05	-	-	250	229
10. CERN	2007-02-01 09h	1.019	16.86	4.55	1.106	0.037	0.893	0	-	76.56	4	2388	52	5768	2.959	8000	91.44	62	-	-	-	-	539	1179
11. CERN_gLite	2007-02-01 09h	0.099	3.849	1.337	0.388	0.005	0.164	0	-	94.26	2	3000	17	3995	0	4094	9.95	1.712	-	-	-	-	96	35
12. CERN-L	2007-02-01 09h	0.731	12.26	3.18	1.942	0.04	0.27	0	-	82.3	2	2793	22	3991	0	4094	34.33	14.08	-	-	-	-	119	55
13. CERNMAC	2007-02-01 09h	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14. Clermont	2007-02-01 09h	0.081	5.753	1.057	0.012	0.167	0.164	0	-	92.85	1	2007	18	3013	0.007	8189	18.43	9.165	-	-	-	-	109	56
15. CNAF	2007-02-01 09h	0.023	1.519	0.687	0.071	0.011	0.164	0	-	97.55	2	3067	9	4005	0	4000	14.37	5.678	-	-	-	-	89	35
16. Cyfronet	2007-02-01 09h	0.037	1.708	0.296	0.288	0.011	0.073	0	-	97.62	2	1300	34	1982	0.186	1000	-	-	16.36	11.06	-	-	70	30
17. FZK	2007-02-01 09h	3.023	14.71	5.382	27.56	0.02	0.278	0	-	52.06	4	3000	67	2007	8.054	3827	23.81	15.24	51.01	84.12	-	-	151	190
18. GRIF_DAPNIA	2007-02-01 09h	0.118	6.042	1.096	0.133	0.053	0.105	0	-	92.57	1	2793	25	2001	0	2047	19.4	12.04	-	-	-	-	103	59
19. GSI	2007-02-01 09h	0.182	8.336	1.751	0.07	0.023	0.253	0	0	89.57	1	2667	58	820.9	0	512	42.78	23.17	-	-	-	-	83	65
20. Houston	2007-02-01 09h	0.05	2.898	0.578	0.049	0.011	0.347	0	-	96.12	1	1396	29	4014	13.04	4095	0.015	0.034	12.72	4.809	-	-	111	84
21. IHEP	2007-01-31 10h	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
22. IPNO	2007-02-01 09h	0.072	1.824	0.727	0.06	0.003	0.116	0	-	97.27	2	2394	15	3994	0	1992	16.51	8.048	-	-	-	-	99	43
23. ISS	2007-02-01 09h	1.943	4.217	3.768	15.06	0.16	0	0	-	76.8	4	2392	67	1009	49.62	996.2	2575	597.5	297.8	2513	-	-	159	160
24. ITEP	2007-02-01 09h	0.073	2.831	0.892	0.464	0.008	0.161	0	-	95.65	2	2999	60	1000	3.843	1992	18.25	10.2	-	-	-	-	115	50
25. JINR	2007-02-01 09h	0.065	2.603	0.864	0.228	0.052	0.261	0	-	95.99	2	2793	47	2005	0	2047	20.68	243.4	-	-	-	-	167	45
26. Jyvaskyla	2006-12-12 14h	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
27. KFKI	2007-02-01 09h	0.071	2.479	0.908	0.089	0.013	0	0	-	96.51	2	3392	15	4052	0	4094	16.3	8.323	-	-	-	-	113	49
28. KISTI	2007-01-25 18h	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29. KNU	2007-02-01 09h	1.585	34.8	18.96	-	-	-	0	-	46.24	2	1000	54	1008	4.4	1913	32.55	63.95	13.4	3.698	-	-	289	123
30. Kolkata	2007-02-01 09h	0.215	2.486	2.308	4.128	0.021	0.273	0	-	90.78	4	2399	25	3829	0	2000	0.541	0.425	-	-	91.18	29.85	168	81
31. Kosice	2007-02-01 09h	0.091	6.247	1.283	0.176	0.124	0.086	0	-	92.08	1	3207	27	2009	0.133	2047	0.104	0.081	20.64	12.74	-	-	90	52
32. LBL	2007-01-30 11h	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Status of the VOBOX, ALICE and WLCG services are monitored through ML

VOBOX AliEn and LCG services status

What is this about?

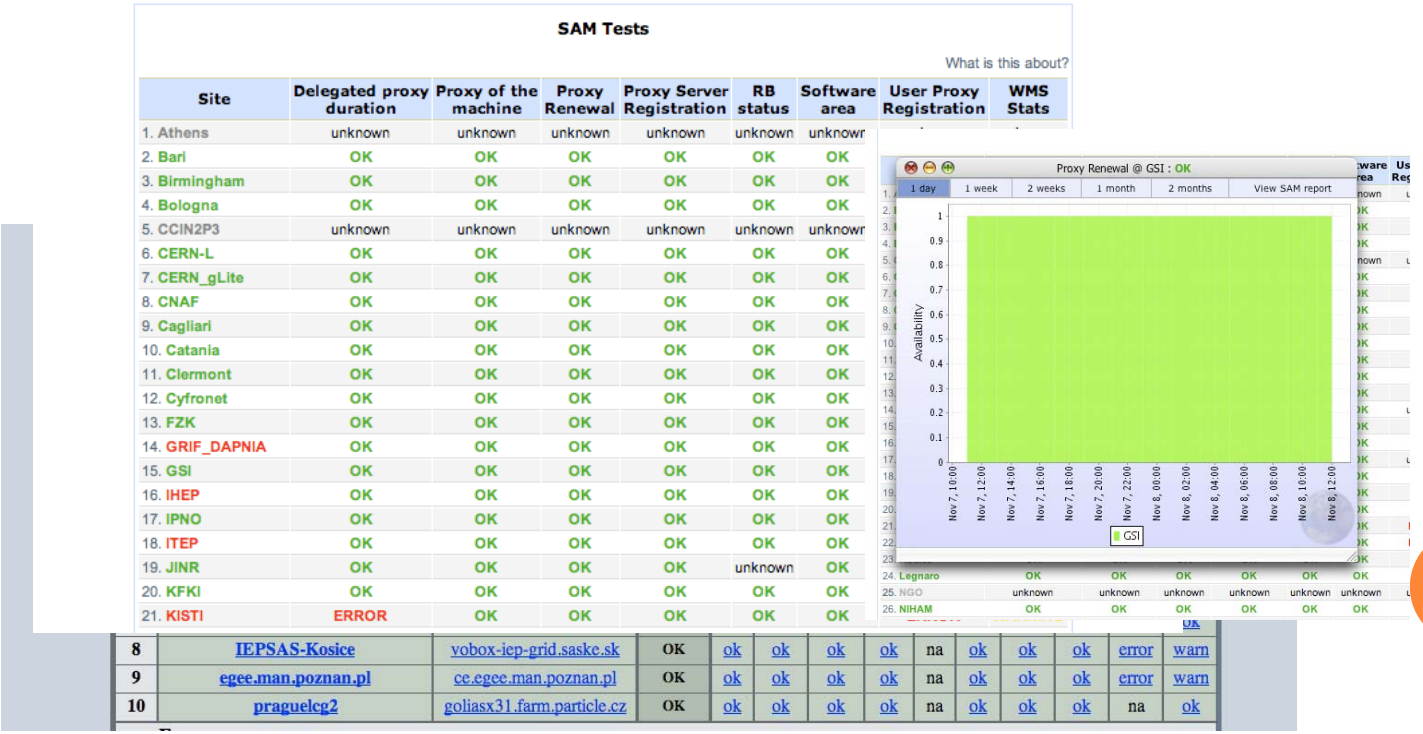
AliEn services

VOBox		AliEn services						AliEn tests			
Address	AliEn version	Monitoring script	CE	SE	PackMan	Monitor	FTD	add	get	rm	whereis
francis.grid.aau.dk	v2-12_57	Glob...					-	Feb ...			
xg010.inp.demokritos.gr			-	-	-	-	-	-	-	-	-
alicegrid6.ba.infn.it			-	-	-	-	-	-	-	-	-
epbf008.ph.bham.ac.uk	v2-12_71						-	Feb ...			
alice9.bitp.kiev.ua	v2-12_71						-	Feb ...			
bcalice6.bo.infn.it	v2-12_71						-	Feb ...	Fail...		Fail...
vobox.ca.infn.it	v2-12_71						-	Feb ...	Fail...		Fail...
vobox.ct.infn.it	v2-12_71						-	Feb ...	Fail...		Fail...
cclcgalice.in2p3.fr	v2-12_71	-	-	-	-	-	-	Time...	Fail...	Time...	Fail...
aliendb4.cern.ch	v2-12_50		-	-	-	-	-	Feb ...			
vcalice03.cern.ch	v2-12_71			DEAD...	DEAD...		-	Feb ...	Fail...		Fail...
lxb7281.cern.ch	v2-12_71						-	0 Fe...	Fail...		
alimacx01.cern.ch	Failed wit...						-	Fail...	Fail...	Fail...	Fail...
clrvoboxalice.in2p3.fr	v2-12_71						-	Feb ...			
ui01-alice.cr.cnaf.infn.it	v2-12_71		DEAD...	DEAD...	DEAD...	DEAD...	-	Fail...	Fail...		Fail...
ares01.cyf-kr.edu.pl	Failed wit...		DEAD...	DEAD...	DEAD...	DEAD...	-	Fail...	Fail...	Fail...	Fail...
alice-fzk.gridka.de	v2-12_71						-	Time...	Fail...		Fail...
node21.datagrid.cea.fr	v2-12_71						-	Feb ...			
grid1.gsi.de	v2-12_71						-	Time...	Fail...		Fail...
login2.tic2.uh.edu	Failed wit...		DEAD...				-	Fail...	Fail...	Fail...	Fail...
ui0001.m45.ihep.su		-	-	-	-	-	-	-	-	-	-
ipnvobox.in2p3.fr	v2-12_71						-	Feb ...			
alien.spacescience.ro	Timeout af...						-	Feb ...	Time...		
gliocl.itep.ru	v2-12_71						-	0 Fe...			
lcgvob01.jinr.ru	v2-12_71						-	srm ...	Fail...		Fail...

- Sites are encouraged to check the status through these pages
- Alarm system established for the moment (Gridmap system planned)

SAM SYSTEM FOR ALICE

- ALICE and SAM developers have implemented an own test suite to test the VOBOXES based in the following requirements:
  - Full freedom to create the test suite, to manipulate it and to chose the list of VOBOXES to tests at any moment
- The tests suite is launched each 2 hours to each site and the results are published into the SAM page
  - The experiment has full control of the test suite
  - Results are also visible from MonaLisa



SAM Tests

What is this about?

Site	Delegated proxy duration	Proxy of the machine	Proxy Renewal	Proxy Server Registration	RB status	Software area	User Proxy Registration	WMS Stats
1. Athens	unknown	unknown	unknown	unknown	unknown	unknown		
2. Bari	OK	OK	OK	OK	OK	OK		
3. Birmingham	OK	OK	OK	OK	OK	OK		
4. Bologna	OK	OK	OK	OK	OK	OK		
5. CCIN2P3	unknown	unknown	unknown	unknown	unknown	unknown		
6. CERN-L	OK	OK	OK	OK	OK	OK		
7. CERN_gLite	OK	OK	OK	OK	OK	OK		
8. CNAF	OK	OK	OK	OK	OK	OK		
9. Cagliari	OK	OK	OK	OK	OK	OK		
10. Catania	OK	OK	OK	OK	OK	OK		
11. Clermont	OK	OK	OK	OK	OK	OK		
12. Cyfronet	OK	OK	OK	OK	OK	OK		
13. FZK	OK	OK	OK	OK	OK	OK		
14. GRIF_DAPNIA	OK	OK	OK	OK	OK	OK		
15. GSI	OK	OK	OK	OK	OK	OK		
16. IHEP	OK	OK	OK	OK	OK	OK		
17. IPNO	OK	OK	OK	OK	OK	OK		
18. ITEP	OK	OK	OK	OK	OK	OK		
19. JINR	OK	OK	OK	OK	unknown	OK		
20. KFKI	OK	OK	OK	OK	OK	OK		
21. KISTI	ERROR	OK	OK	OK	OK	OK		
24. Legnaro	OK	OK	OK	OK	OK	OK		
25. NGO	unknown	unknown	unknown	unknown	unknown	unknown		
26. NIHAM	OK	OK	OK	OK	OK	OK		

Proxy Renewal @ GSI : OK

1 day | 1 week | 2 weeks | 1 month | 2 months | View SAM report

Availability

GSI

8	IEPSAS-Kosice	vobox-iep-grid.saske.sk	OK	ok	ok	ok	ok	na	ok	ok	ok	error	warn
9	egee.man.poznan.pl	ce.egee.man.poznan.pl	OK	ok	ok	ok	ok	na	ok	ok	ok	error	warn
10	praguelcg2	goliasx31.farm.particle.cz	OK	ok	ok	ok	ok	na	ok	ok	ok	na	ok

T0-T1 TRANSFERS: FTS

- FileTransfer Service deployed at all sites
  - Used for scheduled replication of data between computing centers
  - Used as plug-in in the AliEn File Transfer Daemon (FTD)
    - FTD running in the VO-box as one of the ALICE services
- Access to the SRM SE at all sites also required
- Monitored via Dashboard

FTS EFFICIENCY

Click on any Site, and you will have a breakdown according to the errors trasnfering files to that site

This table presents the transfers that have been done from CERN to the ALICE T1

Transfers done on: Sun 15 Oct 2006

Site (click on any site)	Successful transfers	Failed transfers	Efficiency
ALICE::LCG::SARA	0	2412	0.00 %
ALICE::LCG::RAL	57	456	11.11 %

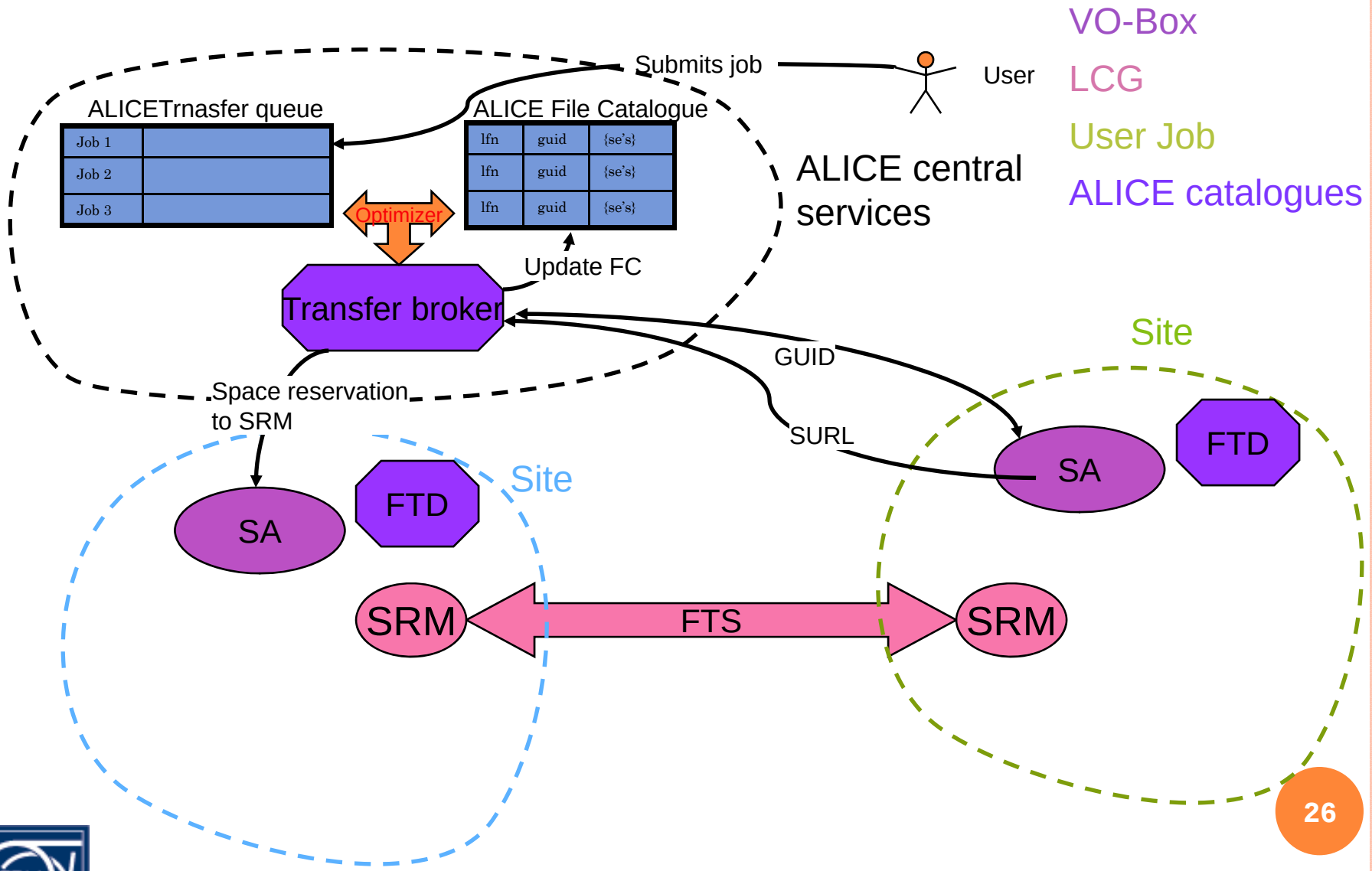
Error message	Counter
The FTS transfer _transferid_ failed (Transfer failed. ERROR the server sent an error response: 425 425 Cannot open port: java.lang.Exception: Pool request timed out : csfnfs62_1)	241
The FTS transfer _transferid_ failed (Transfer failed. ERROR the server sent an error response: 451 451 Local resource failure: malloc: Cannot allocate memory.)	169
contacting the Manager/Transfer	13
asking the SE:	13
adding the file to the LVM at /stage/sl3-lcg-exp/alicesgm/alien/lib/perl5/site_perl/5.8.7/AliEn/Service/SE.pm line 1119.	10
The FTS transfer _transferid_ failed (Operation was aborted (the gridFTP transfer timed out).)	8
The FTS transfer _transferid_ failed (Failed on SRM get: Failed To Get SURL. Error in srm__get: service timeout.)	1
syntax error at line 1, column 0, byte 0 at /stage/sl3-lcg-exp/alicesgm/alien/lib/perl5/site_perl/5.8.7/i686-linux/XML/Parser.pm line 187 500 Can't connect to aliendb1.cern.ch:8095 (connect: timeout)	1

ALICE::LCG::CNAF	1645	542	75.22 %
ALICE::LCG::FZK	985	80	92.49 %
ALICE::LCG::CCIN2P3	2601	65	97.56 %

FTS Tests: Strategy

- FTS is the transfer protocol chosen by Alice between T0 and T1 sites
- T1-T2 transfers are managed by xrootd
- Data types
  - T0-T1: Migration of raw and 1st pass reconstructed data
  - T1-T2 and T2-T1: Transfers of ESDs, AODs (T1-T2) and T2 MC production for custodial storage (T2-T1)
  - T1-T1: Replication of ESDs and AODs

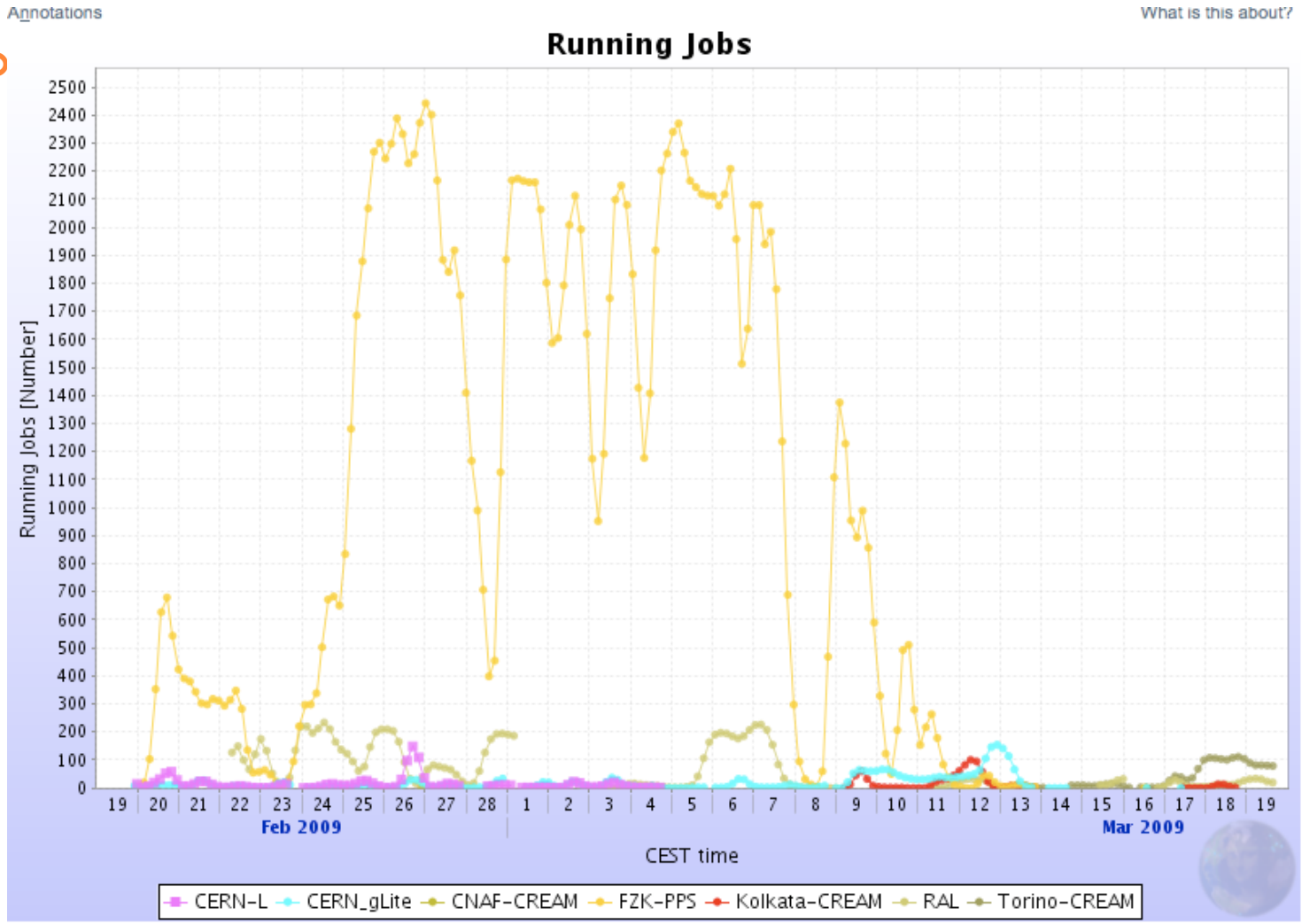
File Replication



CREAM-CE

- ALICE is interested in the deployment of the CREAM-CE service at all sites which provide support to the experiment
  - **GOAL: Deprecation of the WMS use in benefit of the direct CREAM-CE submission**
  - WMS submission mode to CREAM-CE not required
    - The experiment is not limited by the issues observed while using the WMS submission mode
  - In addition the proxy renewal feature was neither required
    - 48h voms extensions ensured by the security team@CERN
    - Enough to run production/analysis jobs without any addition extension
- ALICE has began to test the CREAM-CE since the beginning of Summer 2008 into the real production environment
- ALICE testing priority list:
  - CREAM-CE
  - SLC5 (DONE)
  - glexec/SCAS (Beginning of the summer 2009)

- The 1st test phase of the CREAM-CE
  - Performed in summer 2008 at FZK (T1 site, Germany)
  - Tests operated through a second VOBOX parallel to the already existing service at the T1 (operating in WMS submission mode)
  - Access to the local CREAM-CE was ensured through the PPS infrastructure
    - Initially 30 CPUs
    - Moved to the ALICE production queue in few weeks (production setup)
  - Intensive functionality and stability tests from July to September 2008
    - Production stopped to create an ALICE CREAM module into AliEn and to allow the site to upgrade their system
  - Results:
    - More than 55000 jobs successfully executed through the CREAM-CE in the mentioned period
    - No interventions in the VOBOX required during the testing phase



- During the 2nd test phase more than **67000** jobs have been successfully executed through all CREAM-CE system