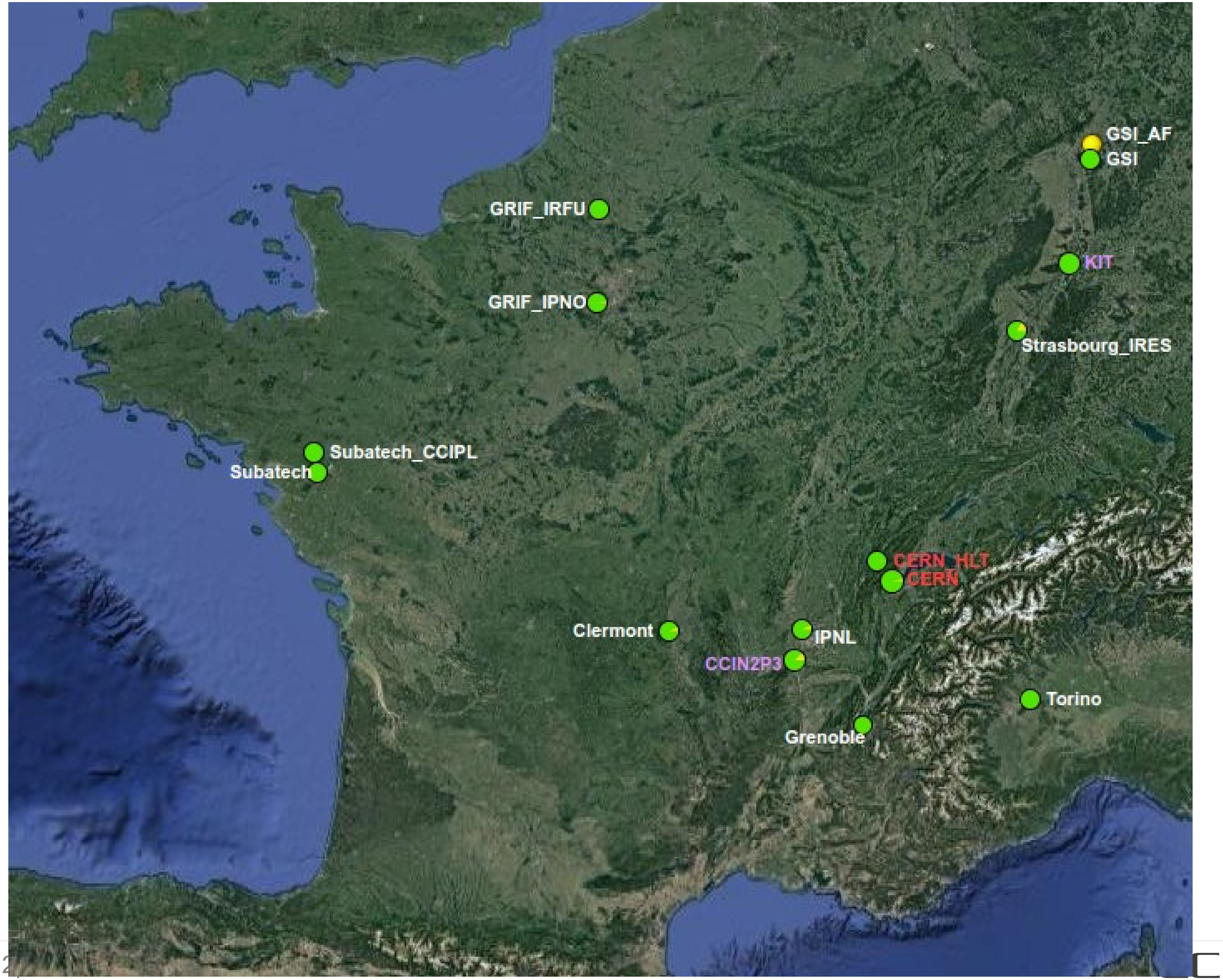
Centre de Calcul de l’Institut National de Physique Nucléaire et de Physique des Particules

French sites : status & plans

ALICE T1/T2 workshop @ Bucharest
2019-05-16
Renaud Vernet


GSI_AF
GSI
GRIF_IRFU
KIT
GRIF_IPNO
Strasbourg_IRES
Subatech_CCIPL
Subatech
CERN_HLT
CERN
Clermont
IPNL
CCIN2P3
Torino
Grenoble

Pledges 2019

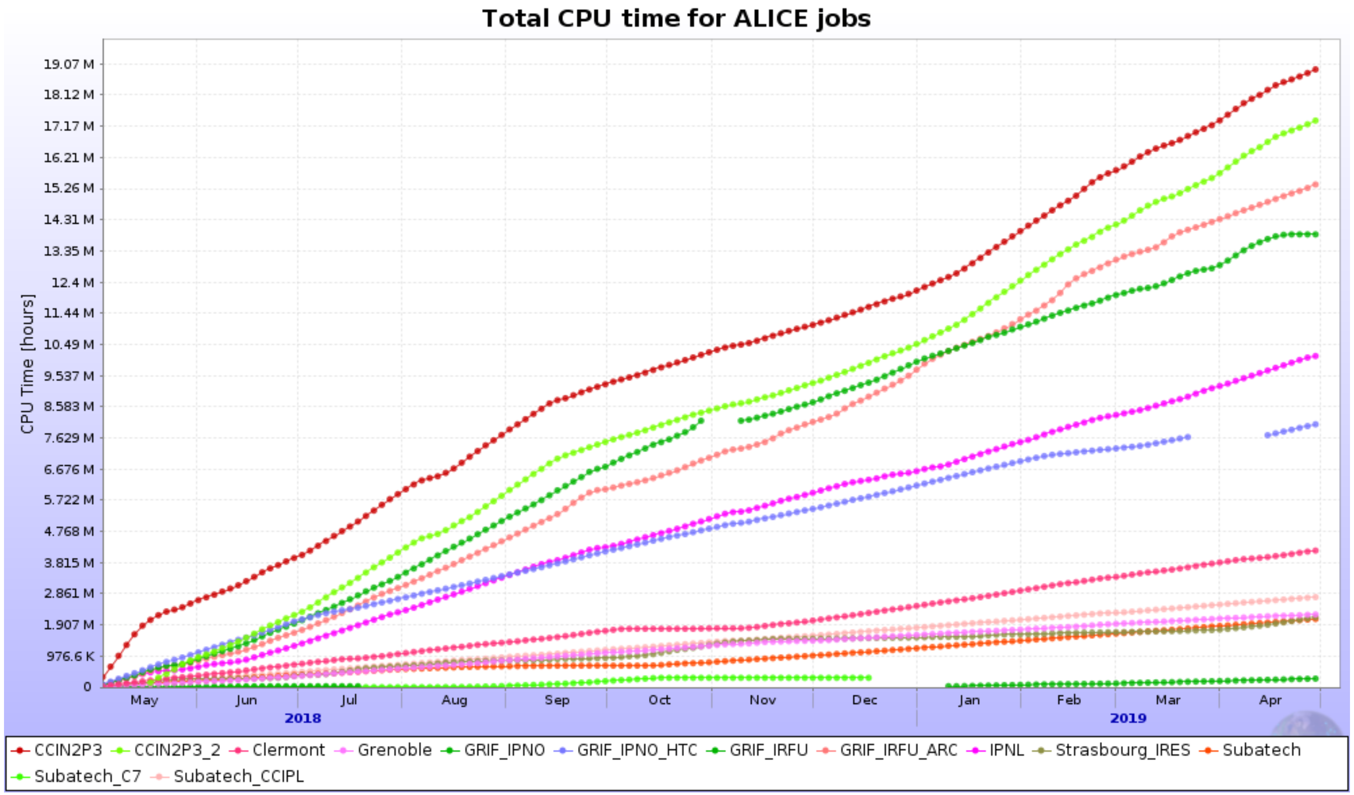
	T1			T2 (*)	
	capacity	vs T1 requ.		capacity	vs T2 requ.
CPU	41 k	11 %		45 kHS	12 %
Disk	5.1 PB	11 %		4.2 PB	12 %
Tape	6.2 PB	11 %			

(*) IPNL T3 not accounted for

Significant budgetary support from FA maintained

French contribution to ALICE computing

- 11.7 % of total ALICE CPU time
  - was 9 % last year



CCIN2P3

IRFU

IPNL

IPNO

Clermont

Subatech
Strasbourg
Grenoble

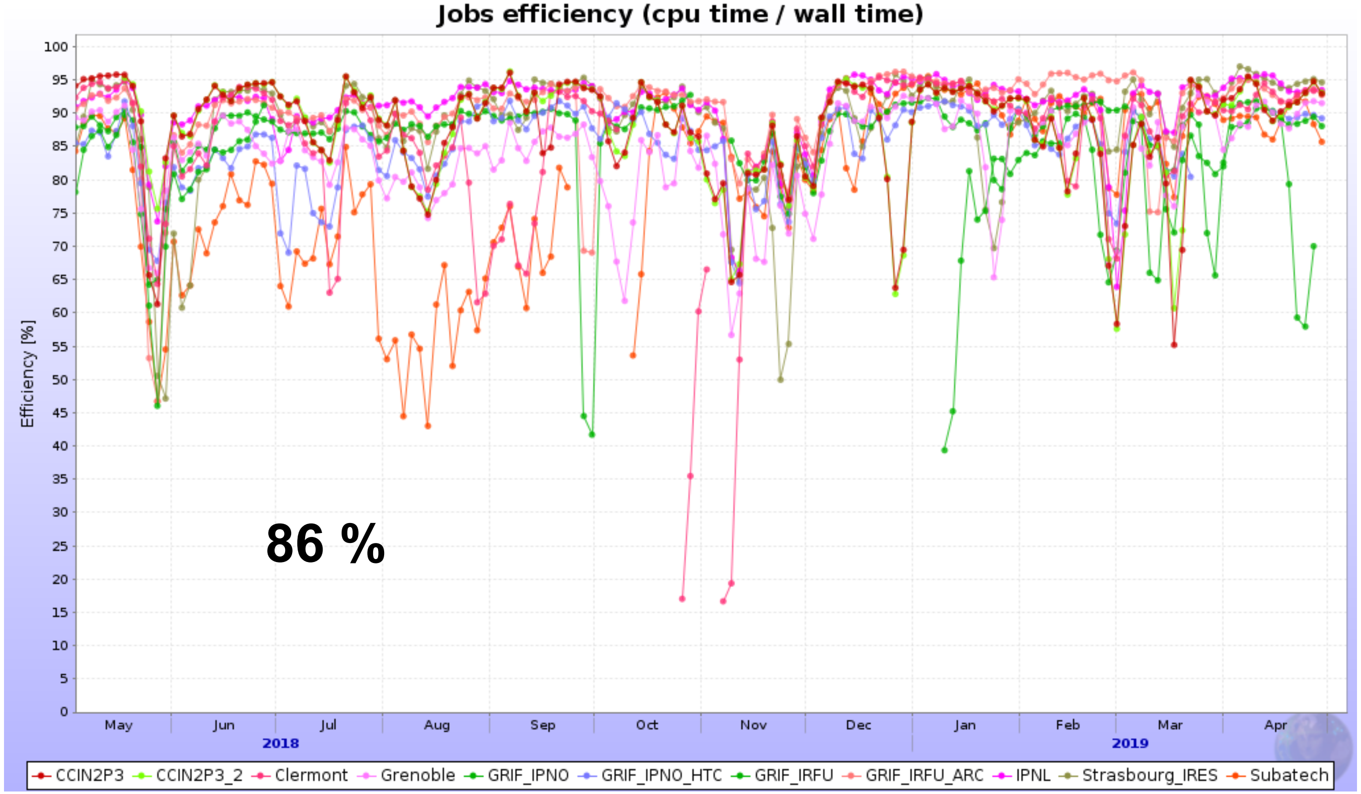
Summary CPU pledge utilization

IPNL	Nantes	GRIF	Clermont	Grenoble	Strasbourg	CCIN2P3
N/A	+100 %	+170 %	+42 %	+ 7 %	-15 %	+55 %

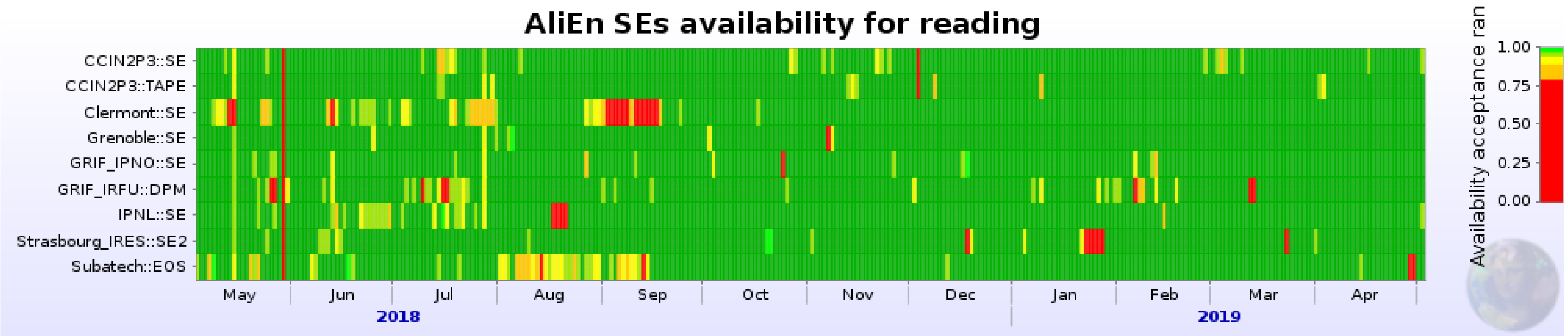
Normalized walltime / pledged CPU from May 2018 to Apr 2019
Source : EGI accounting portal

Job efficiency

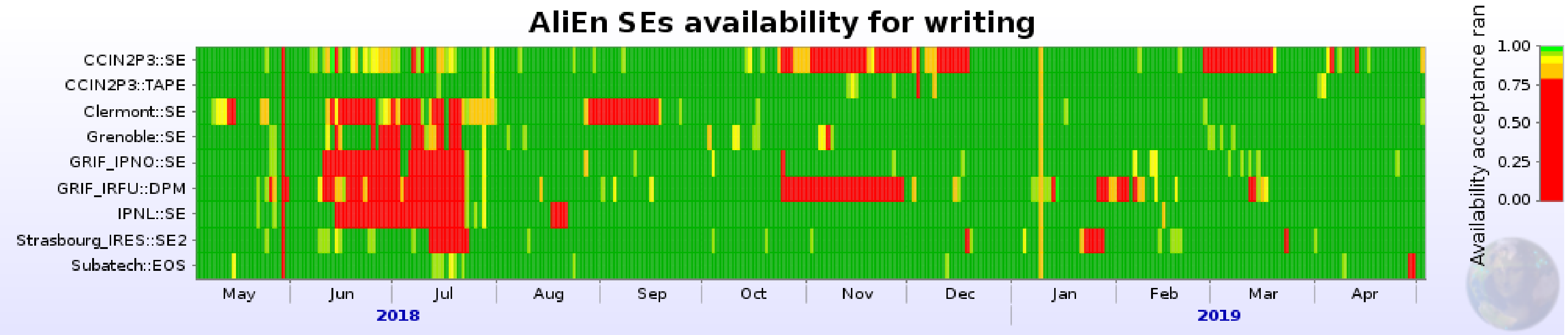
- No significant recurring problem



Storage



> 97 %

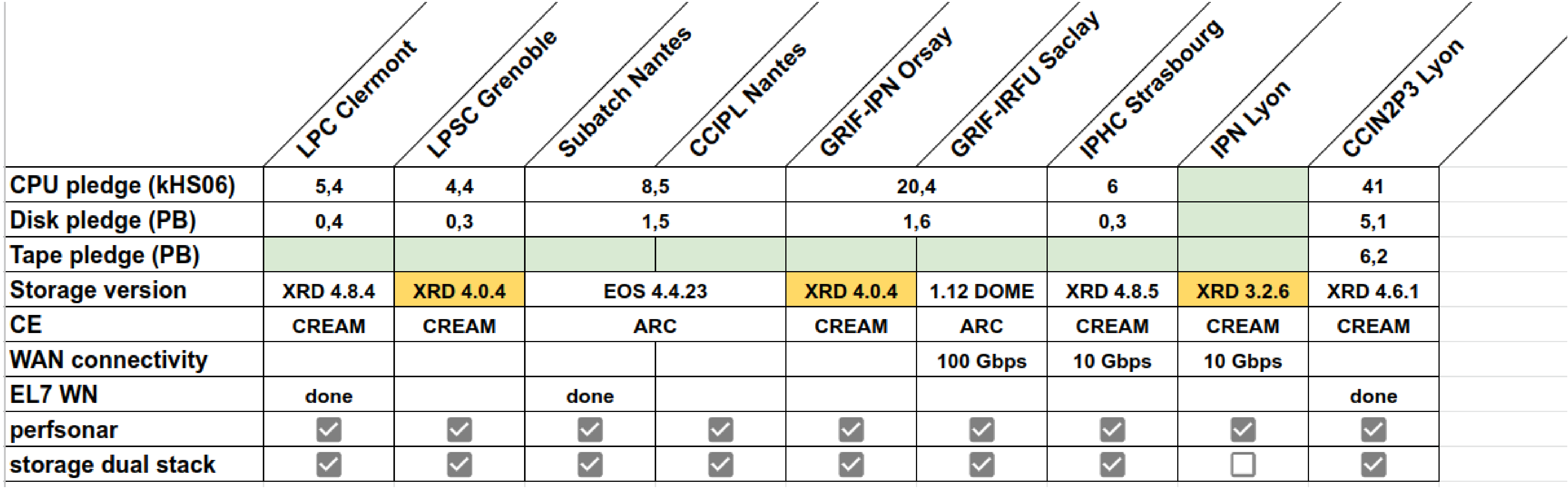


~ 93 %

(almost) all sites provide dual stack storage

News from sites

Summary table (WIP)



	LPC Clermont	LPSC Grenoble	Subatch Nantes	CCIPL Nantes	GRIF-IPN Orsay	GRIF-IRFU Saclay	IPHC Strasbourg	IPN Lyon	CCIN2P3 Lyon
CPU pledge (kHS06)	5,4	4,4	8,5		20,4		6		41
Disk pledge (PB)	0,4	0,3	1,5		1,6		0,3		5,1
Tape pledge (PB)									6,2
Storage version	XRD 4.8.4	XRD 4.0.4	EOS 4.4.23		XRD 4.0.4	1.12 DOME	XRD 4.8.5	XRD 3.2.6	XRD 4.6.1
CE	CREAM	CREAM	ARC		CREAM	ARC	CREAM	CREAM	CREAM
WAN connectivity						100 Gbps	10 Gbps	10 Gbps	
EL7 WN	done		done						done
perfsonar	☑	☑	☑	☑	☑	☑	☑	☑	☑
storage dual stack	☑	☑	☑	☑	☑	☑	☑	☐	☑

GRIF-IRFU *(Paris Saclay)*

Admins
F. Schaer
S. Ferry

- Compute
  - CREAM scheduled for decommission
  - ARC CE
    - on SSD → dramatic improvement
    - ARC6 ready for deployment (still beta)
- Storage
  - Still issues with small ALICE files
- Network
  - 100 Gb/s ready to be used
  - But still problems at NREN level
- Budget
  - Little budget available for off-site representation
  - Provision of future pledged resources ?
    - Unknown

GRIF-IPNO *(Paris - Orsay)*

Admin
C. Diarra

- Compute
  - No growth this year
  - No success (yet) in smart integration of ALICE on HPC farm (~900 cores)
- Storage
  - +200 TB
- Fusion of several Parisian labs
  - IPNO + LAL + other labs
  - Common pool of resources (~2020)
  - Will fight try to keep native xrootd for ALICE

IPNL *(Lyon)*

Admin
D. Pugnère

- Resources
  - No growth, CPU contribution will drop (no number yet)
  - Stay WLCG T3 nonetheless
- Connectivity
  - no planned evolution : traffic low compared to available bandwidth
- Efforts sustainable
- Dual stack for storage
  - Planned for 1st semester of 2019

Subatech + CCIPL *(Nantes)*

Admin
J-M. Barbet

- Subatech will probably close in 2023

- Plan
  - + 4kHS06 (2019) at CCIPL
  - + 1 PB disk (2020) at Subatech
- Compute
  - 50 % WN in Centos7 with ARC+HTCondor
  - 50 % CCIPL (HPC Center)
  - 1 Vobox Centos7
  - 1 CREAM decommissioned
- EOS
  - in dual stack
  - Managers reinstalled in Centos7

Subatech + CCIPL *(Nantes)*

- ARC-CE
  - Jobs in status « hold » taken as « running » by ARC (→ be careful)
    - AliEn CE mistaken
    - Solution : create cron deletes « hold » jobs
  - Agressive memory management when submitting to condor (default)
    - → unset this option on ARC config
  - Jobs local DB *jobs.dat* corrupted
    - → needed to remove the file

- Some storage availability issues
  - Revealed large packet drops in internal network infrastructure
  - 2 switches replaced

Clermont

Admin
J-C. Chevaleyre

- Upgraded all xrootd servers to 4.8.4
- +28 % storage this year
- Migrated WN’s to Centos7

Grenoble

Admin
C. Gondrand

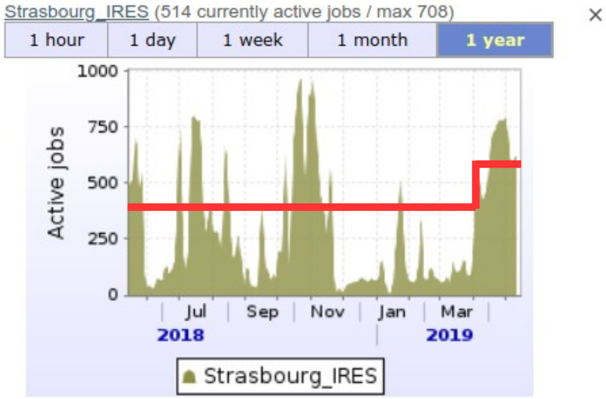
- Admin « team » understaffed
  - 0.5 FTE for grid activities
- Storage
  - Full dual stack
  - Servers in SL6, xrootd 4.0.4
  - Redirector in Centos7, xrootd 4.4.8
- Future of Grenoble site
  - Several system admins in lab will retire within 5 years
  - End of local financial support in 2021
  - End of ALICE site under consideration (nothing decided yet)
  - Possibly continue for a few years
  - diskless site ?

ALICE point of view ?

Strasbourg

Admin
Y. Patois

- Stable staffing, compatible with commitment
  - As long as techos & config do not change too much
  - Move to EOS would need a bit more effort
  - As long as native xrootd is supported, all is fine !
- Compute
  - Will soon move from CREAM+PBS/Maui to ARC+Condor
- Conectivity
  - Should move from 10 Gbps to 2x25 Gbps « soon »
- Use of CPU pledges
  - Noticed only recently
  - Probably a config problem
  - won't happen again, pledge delivery monitoring put in place


Strasbourg_IRES (514 currently active jobs / max 708)

CCIN2P3 *(Lyon)*

Contact
R. V.

- Storage
  - 1 server lost (RAID issues) → all data lost on server
- Connectivity
  - 40 Gbps to LHC-ONE network
- Compute
  - Univa Grid Engine + CREAM
  - Centos7
  - HTCondor pool likely to be put in place (for grid jobs)

General issues

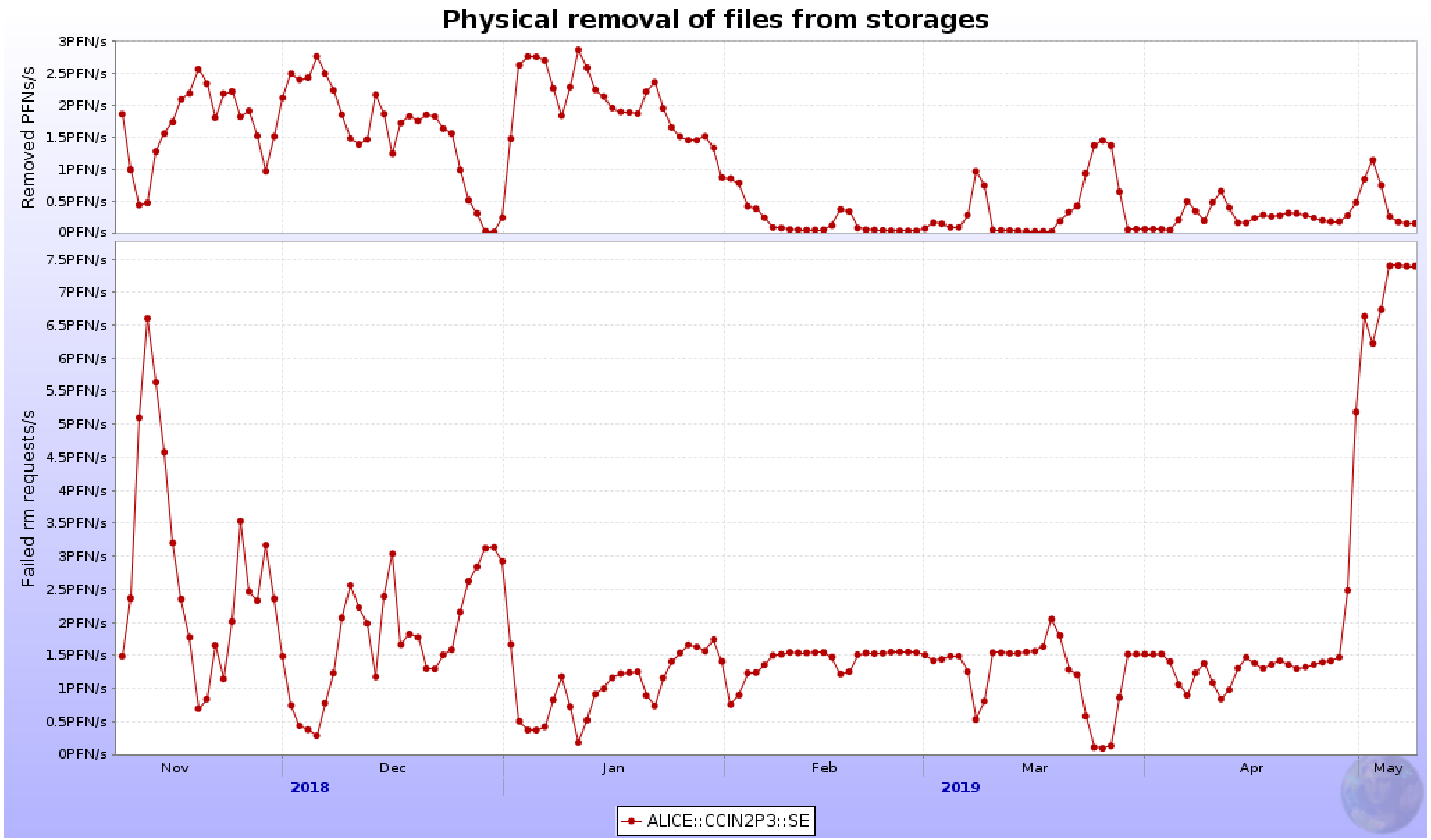
- Running services with Latchezar’s proxy
  - Compatibility issues between AliEn and new French CA
  - jAlien should fix that (more recent openssl)

ALICE::CCIN2P3::SE

- 4 PB Storage Element
  - Operations OK with jobs
- Many files to be deleted
  - Dark data (not registered in catalog)
- Deletion rate not good
  - ~ 2Hz
  - Dark data stacks up
  - Early 2019 : 180M files total, 100 Mfiles to delete
- 2 symptoms observed by Costin
  - Xrootd takes time to return answer (why?)
  - Large number of errors during deletion (why?)
- Temporary solution
  - Files deleted manually on site
  - Need to solve deletion speed in future

https://doc.cc.in2p3.fr/intranet:lcg:coordination:problem:aliceperformancesuppression

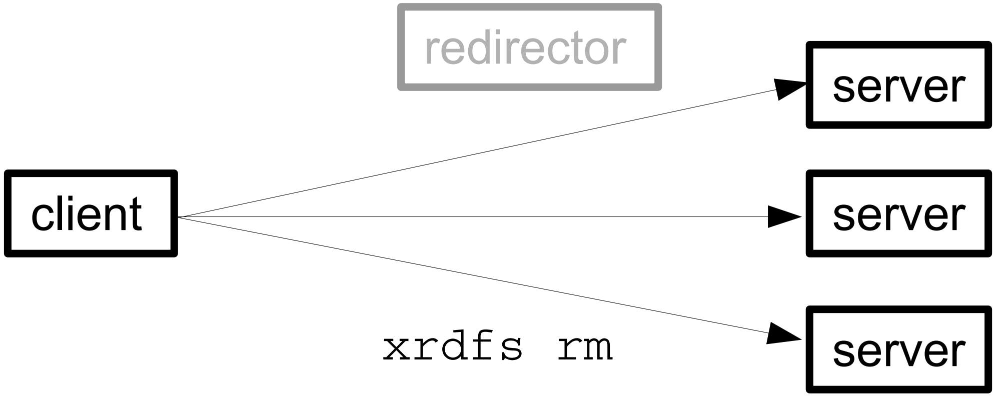
ALICE::CCIN2P3::SE



Deletion speed

Error rate

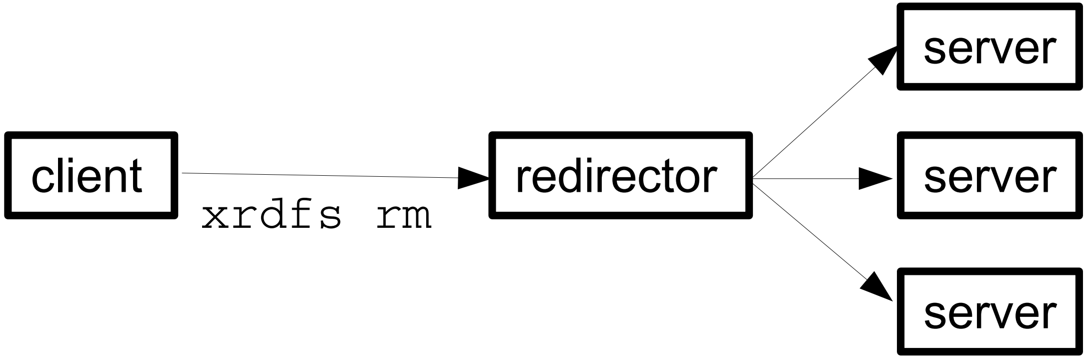
Bypassing redirector



τ ~ 10 ms

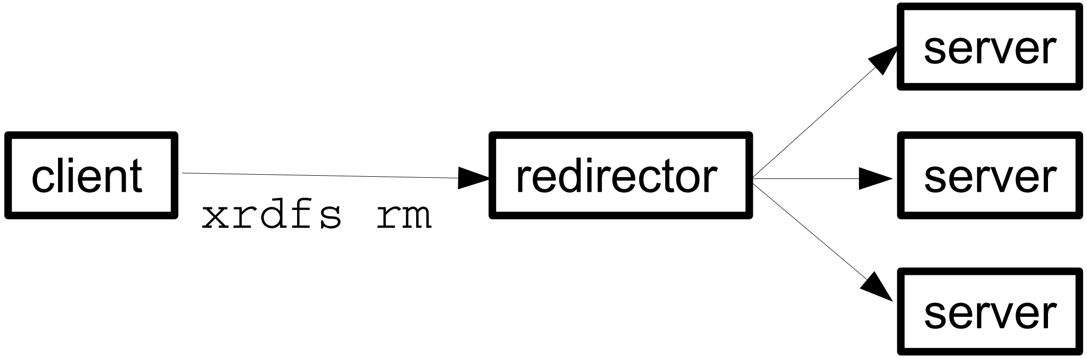
Through redirector

Files freshly written :



τ ~ 10 ms

After 'some time' :



τ = 5 s

ALICE::CCIN2P3::SE

- Many email exchanges to understand the reason
  - Cern ↔ ccin2p3 ↔ xrootd

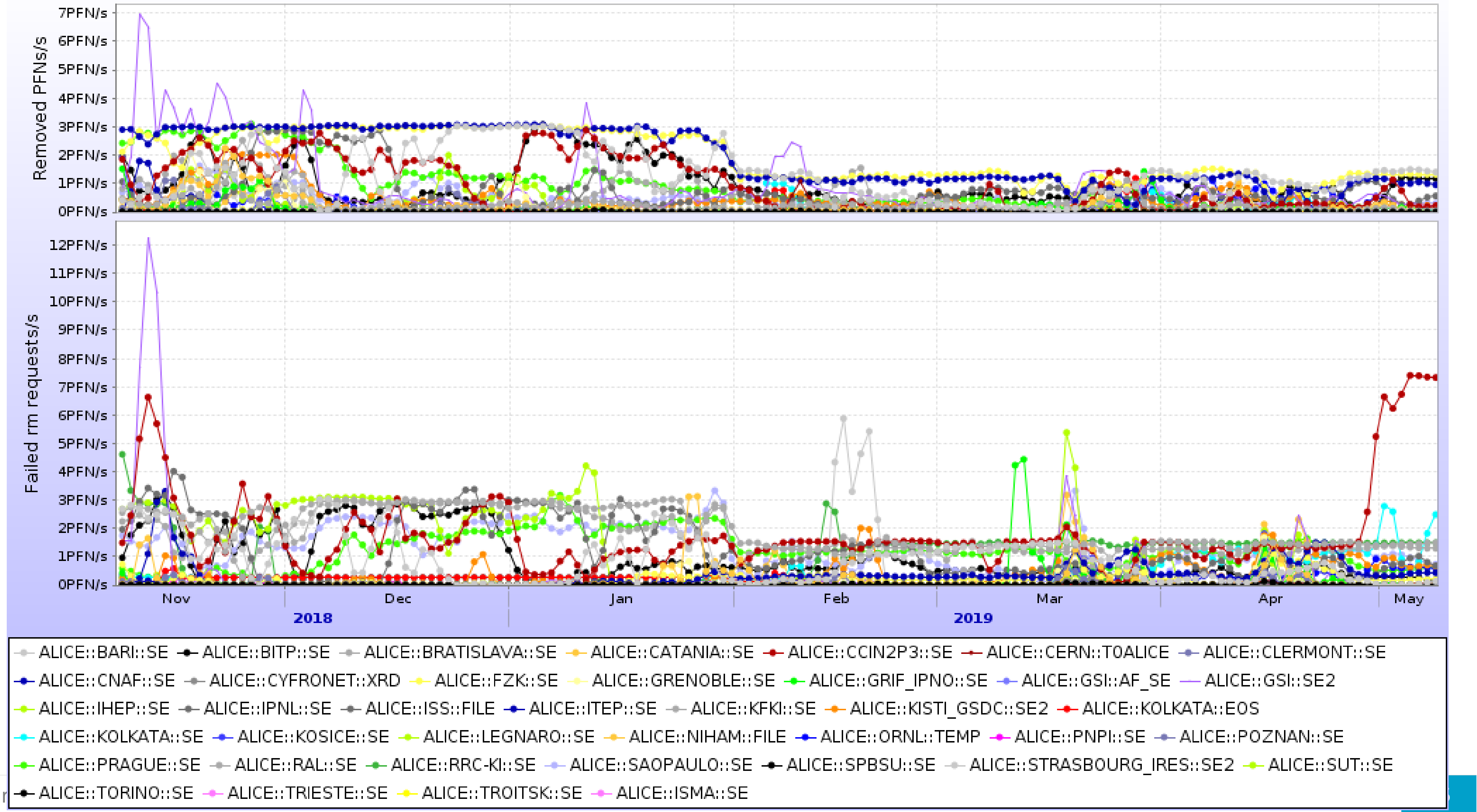
- (my personal) current conclusions
  - Cache effects
  - If file not in cache, **`cms.delay`** drives response time (default is 5 s)
  - Is that normal ? we don’t know

- Xrootd support not conclusive yet

- Need more support from experts (who ?)

Is CCIN2P3 the only site in trouble ?

Physical removal of files from storages



Conclusions

- Service delivery OK
  - Deficit in CPU @ Strasbourg largely compensated by other French sites
  - Storage availabiltity above requirement
- Funding OK at national level
  - Local funding not so clear
  - Subatech quits
  - Grenoble uncertain
- Human effort so far constant
  - Will probably decrease in a few years
  - Not much time to test new technos
- Globally smooth operations
  - But small files on DPM @ IRFU
  - Troubleshooting on xrootd ongoing @ T1