



# CMS FacilitiesOps and IN2P3

[ CMS visit to IN2P3 – Lyon, 23 Oct 09]

Daniele Bonacorsi, Peter Kreuzer  
[ CMS Facilities Ops ]

Claudio Grandi, Chris Brew  
[ T1 coordination in CMS Facilities Ops ]

Andrea Sciabà, Josep Flix  
[ Site Readiness in CMS Facilities Ops ]

Nicolò Magini  
[ Data Transfer Operations and DDT in CMS Facilities Ops ]



# CMS FacilitiesOps

## CMS FacilitiesOps weekly meetings

- ◆ To discuss status of T1 and T2 sites, and related items, over last 7 days
  - CMS contacts at T1's asked to provide brief weekly reports
  - SAM and SiteReadiness status is reviewed, explanations are asked, discussion
- ◆ Weekly, Monday afternoon, 5pm GVA time

## CMS attends WLCG Ops daily calls, 3pm GVA time

- ◆ Official WLCG official minutes:
  - <https://twiki.cern.ch/twiki/bin/view/LCG/WLCGOperationsMeetings>
- ◆ Collection of CMS daily reports:
  - [https://twiki.cern.ch/twiki/bin/view/CMS/FacOps\\_WLCGdailyreports](https://twiki.cern.ch/twiki/bin/view/CMS/FacOps_WLCGdailyreports)



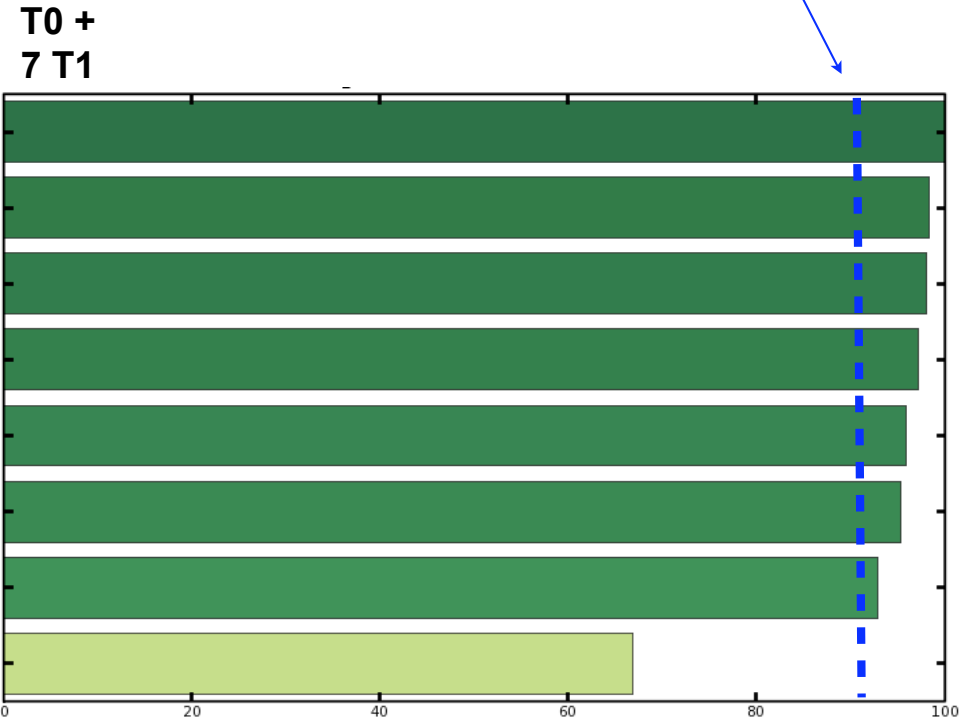
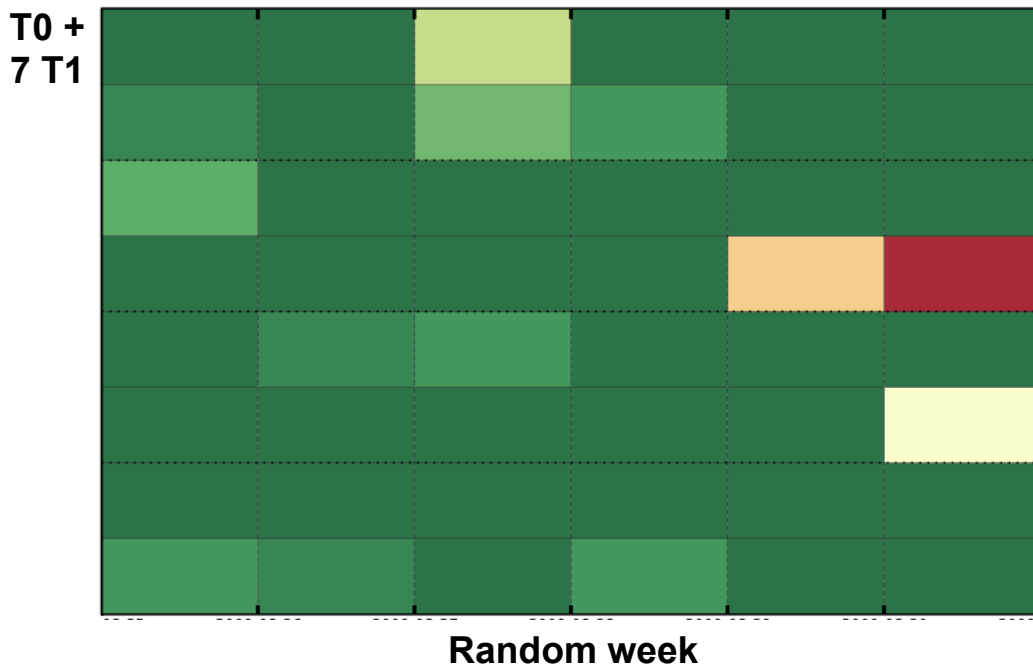
# SAM Availability for CMS T1's

## CMS-specific SAM tests

- ◆ Complementary to WLCG SAM, to mimic real CMS workflows
  - Widely documented elsewhere

## Overall SAM Availability ranking for CMS T1's: goal is **90%**

- ◆ For all **orangish/redish** boxes we discuss at FacOps weekly meetings





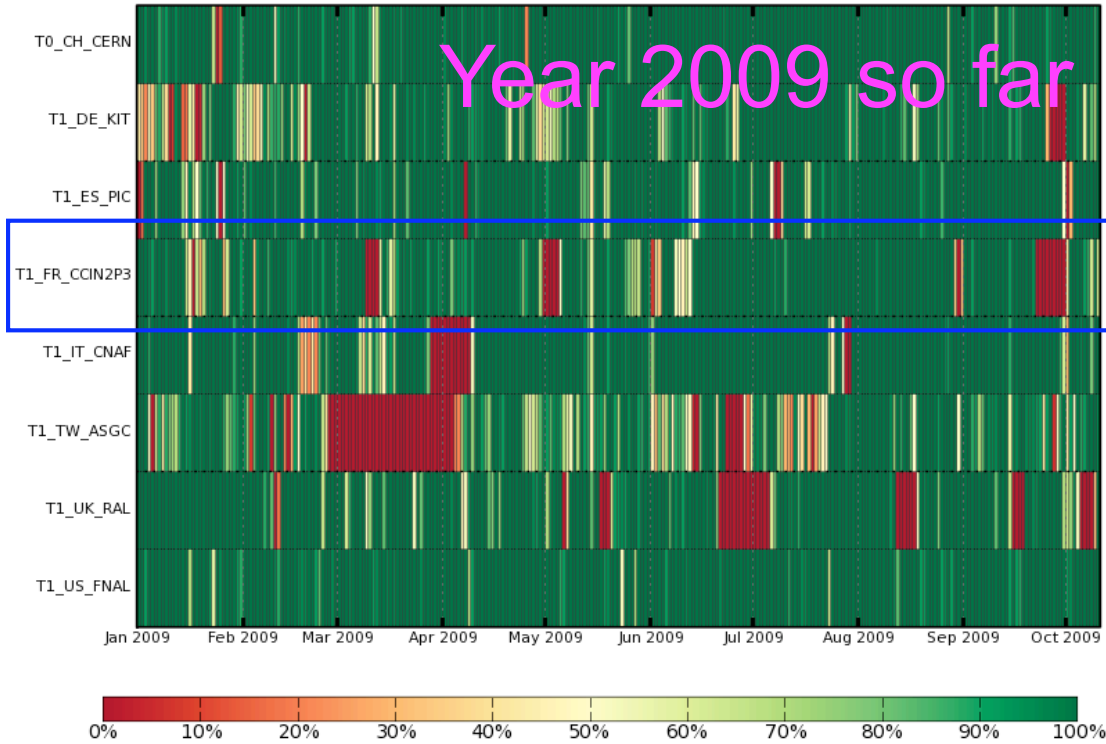
# SAM Availability for IN2P3

Looking to IN2P3 in CMS-specific SAM tests in 2009

## Site Availability

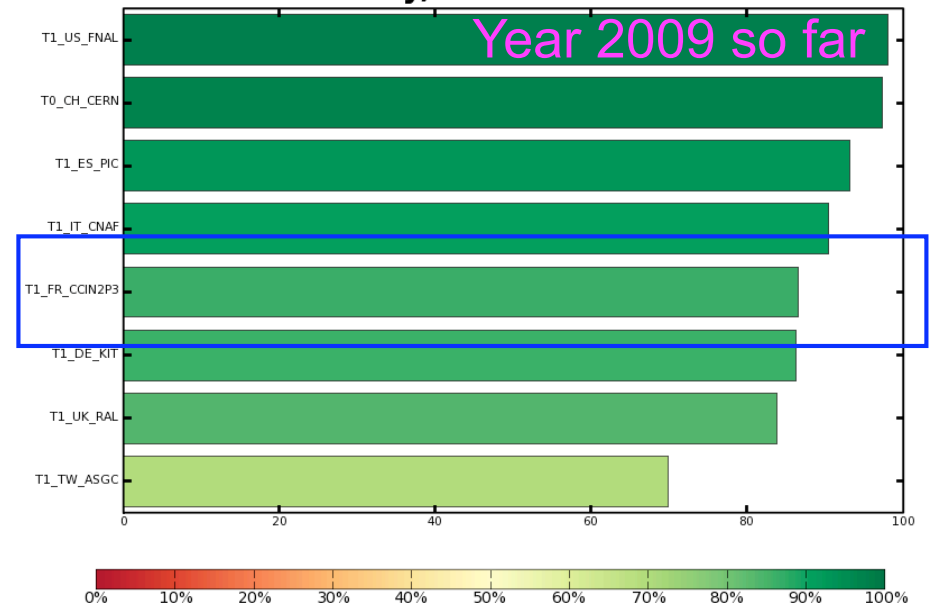
283 Days from Week 00 of 2009 to Week 41 of 2009

Year 2009 so far



## Site Availability, 2009-01-01 - 2009-10-11

Year 2009 so far





# CMS SiteReadiness

Global estimator in FacOps for the readiness of sites for daily operations

T1_[region]_[sitename]																							
Site Readiness Status:	R	W	R	W	NR	NR	NR	R	R	NR	NR	NR	NR	NR	R	R	NR	NR	NR	NR	R		
Daily Metric:	O	O	O	O	O	O	O	O	E	O	E	E	E	O	O	O	E	E	O	E	O	O	
Maintenance:	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	
Job Robot:	98%	96%	99%	98%	97%	96%	98%	96%	89%	98%	97%	96%	85%	100%	100%	100%	80%	0%	100%	100%	100%	99%	
SAM Availability:	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	88%	60%	28%	100%	100%	100%	96%	100%	100%	80%	100%	100%	
T1::downlinkT0:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
T1::downlinks/uplinksT1s:	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	8(d)- 8(u)	
T1::uplinksT2s:	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	
	19	20	21	22	23	24	25	26	27	28	29	30	01	02	03	04	05	06	07	08	09	10	
	Sep							Oct															

<https://twiki.cern.ch/twiki/bin/view/CMS/SiteCommRules>

"Site Readiness Status" as defined in [Site Commissioning Twiki](#):

- R** = READY
- W** = WARNING
- NR** = NOT-READY
- SD** = SCHEDULED-DOWNTIME

"Daily Metric" as boolean AND of all individual metrics considered for the site:

- O** = OK (All individual metrics above Site Commissioning Thresholds; "n/a" ignored)
- E** = ERROR (Some individual metrics below Site Commissioning Thresholds)
- SD** = SCHEDULED-DOWNTIME

- INDIVIDUAL METRICS -

"Scheduled Downtimes": site maintenances

- Up** = Site is not declaring Scheduled-downtime
- SD**=full-site; **SE-SD**: All CMS SE(s) in SD; **CE-SD**: All CMS CE(s) in SD
- ~** = Some SE or CE services (not all) Downtime

"SAM Availability":

- Green** = SAM availability is  $\geq 90\%$
- Red** = SAM availability is  $< 90\%$

"T1::downlinks/uplinksT1s":

- Green** = Site has  $\geq 4$  DDT commissioned uplinks and downlinks, respectively, with other T1 sites
- Red** = Otherwise

"Job Robot":

- Green** = Job success rate is  $\geq 90\%$
- Red** = Job success rate is  $< 90\%$
- = Jobs submitted but not finished
- n/a** = Job success rate is n/a

"T1::downlinkT0":

- Green** = Downlink from T0\_CH\_CERN is DDT-commissioned
- Red** = Otherwise

"T1::uplinksT2s":

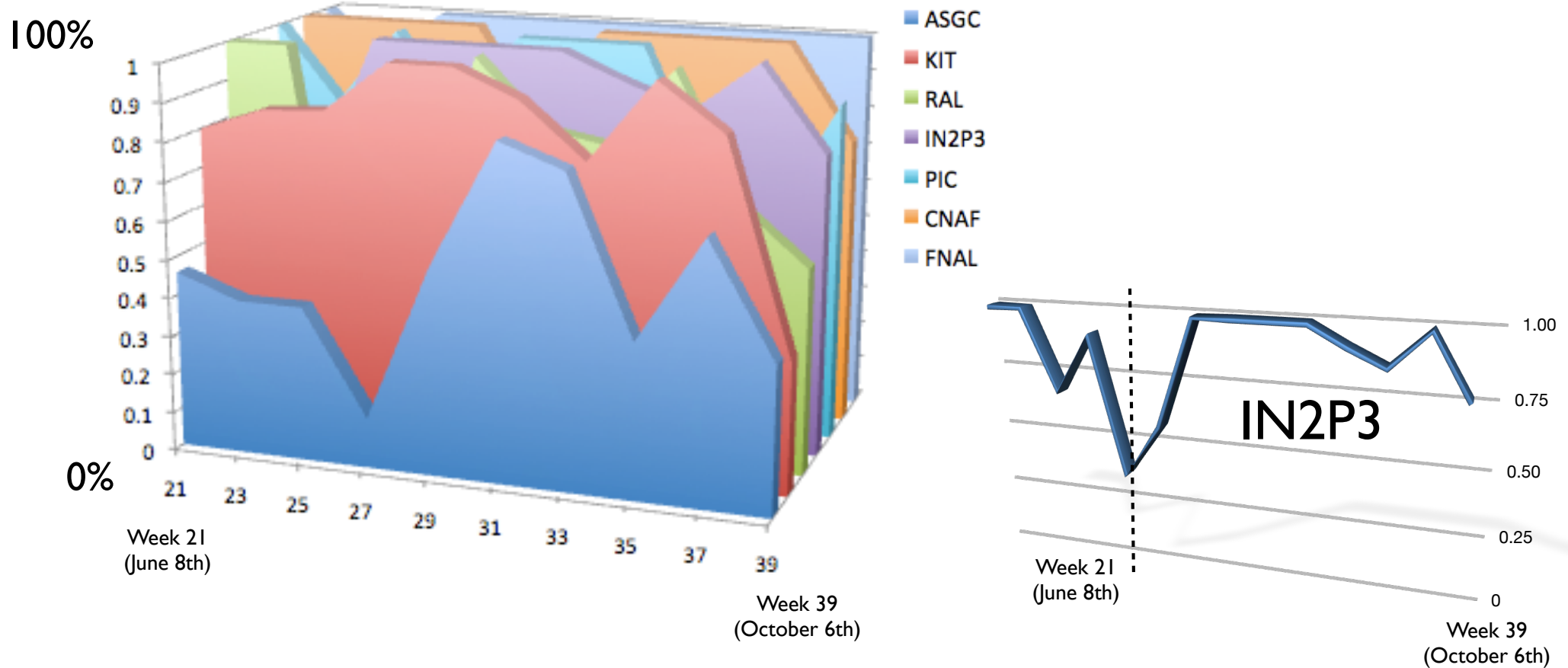
- Green** = Site has  $\geq 20$  DDT commissioned uplinks to T2 sites
- Red** = Otherwise

From CMS Site Readiness metrics:

- ✦ Site availability: fraction of time all functional tests succeed
- ✦ JobRobot efficiency: fraction of successful "fake" analysis jobs
- ✦ Links: # of commissioned data transfer links



# CMS SiteReadiness ranking for CMS T1's



SiteReadiness goal for T1's: **90%**

Achieved averages in Jun-Oct 2009:

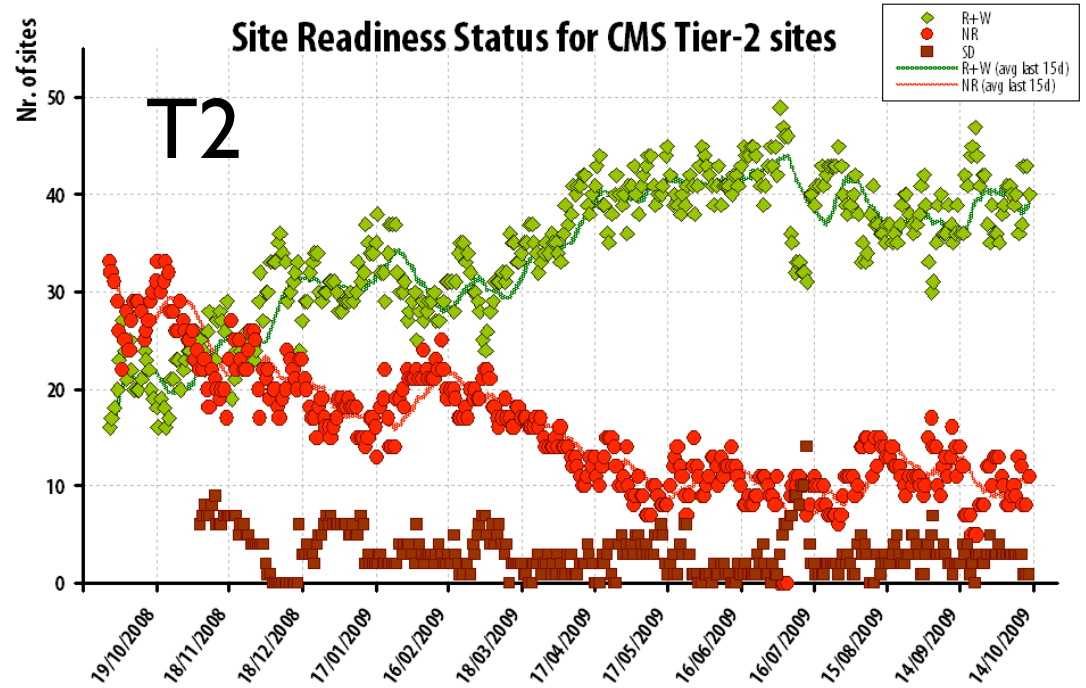
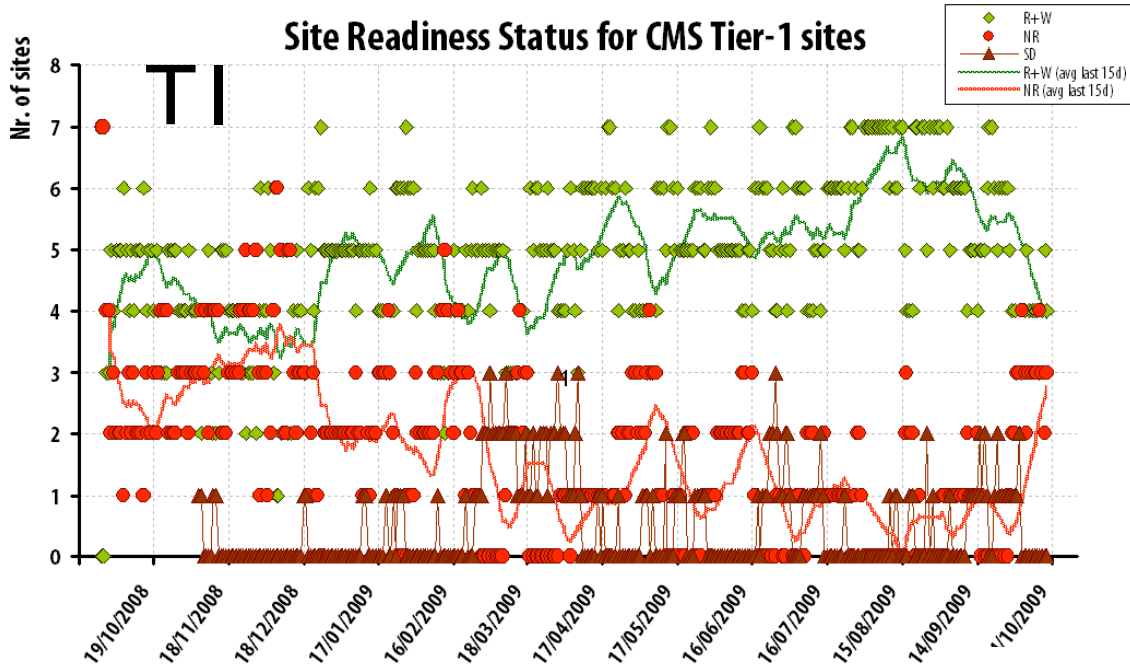
- ◆ {FNAL, CNAF} at {**99%**, **95%**}
- ◆ {PIC, IN2P3, KIT, RAL} at {**87%**, **86%**, **85%**, **73%**}
- ◆ ASGC at **50%**

**WLCG SAM (ops)** not the full picture  
**CMS-specific SAM** not the full picture  
**SiteReadiness (even!)** not the full picture  
 ◆ Need high CPU eff, disk stability, MSS solidity and performance, ...



# Readiness of sites: CMS requirements on Tiers [4/4]

*Example:* historical data on T1 and T2 sites









# SiteReadiness breakdown for IN2P3

Period / State	READY [days]	WARN [days]	NOT READY [days]	Downtime
June 09	20	0	10	0
July 09	29	2	0	0
Aug 09	27	3	0	1
Sep 09	18	2	3	7

NOTE: SiteReadiness has lately suffered from SSB instabilities when tracing scheduled downtimes. The September IN2P3 downtime was corrected on SiteReadiness tables as announced here:

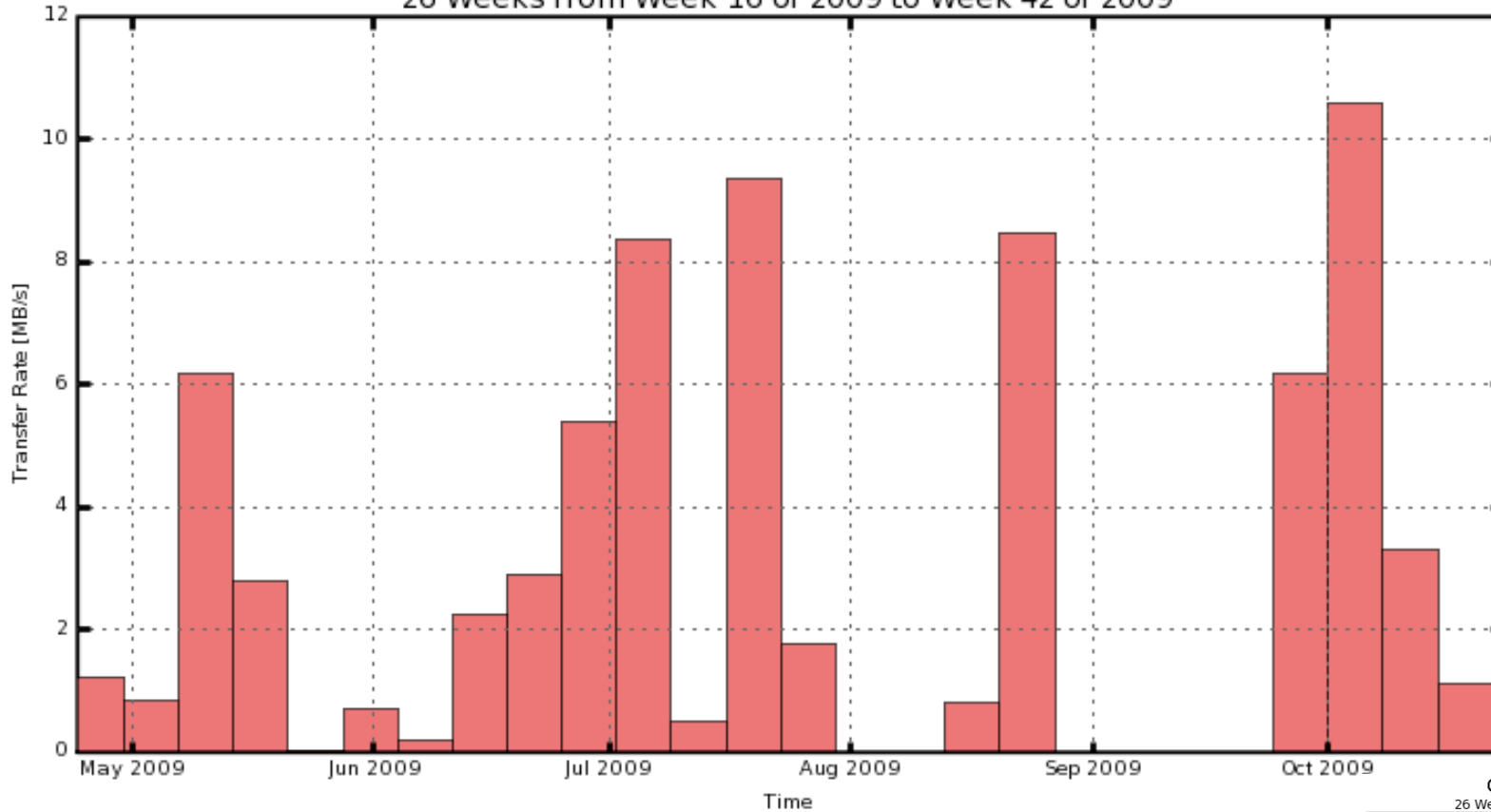
✦ <https://hypernews.cern.ch/HyperNews/CMS/get/sc4/1969.html>



# Transfer rates: T0 -> IN2P3

## CMS PhEDEx - Transfer Rate

26 Weeks from Week 16 of 2009 to Week 42 of 2009

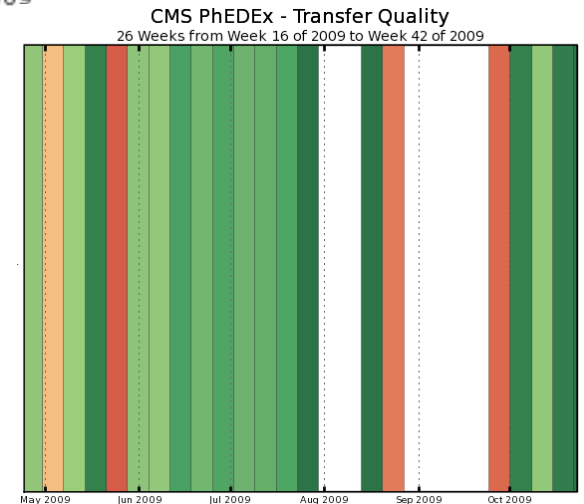


T0\_CH\_CERN\_Export to T1\_FR\_CCIN2P3\_Buffer

Maximum: 10.59 MB/s, Minimum: 0.00 MB/s, Average: 2.80 MB/s, Current: 1.12 MB/s

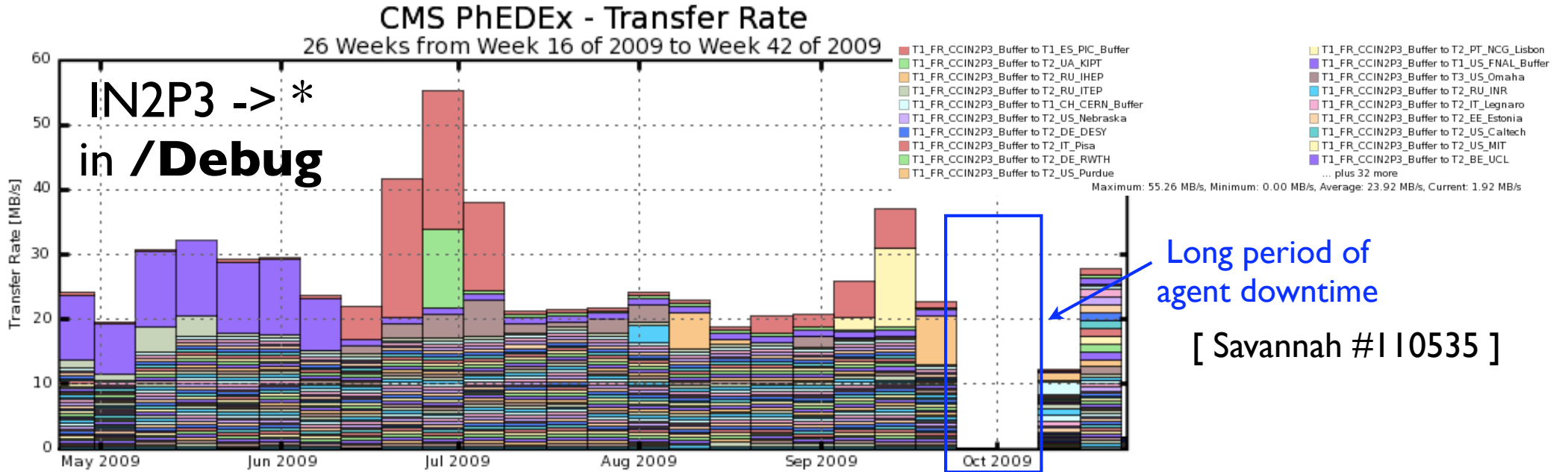
## Little activity in the PhEDEx /Prod instance

- ◆ few datasets from T0 assigned to IN2P3 as custodial site...

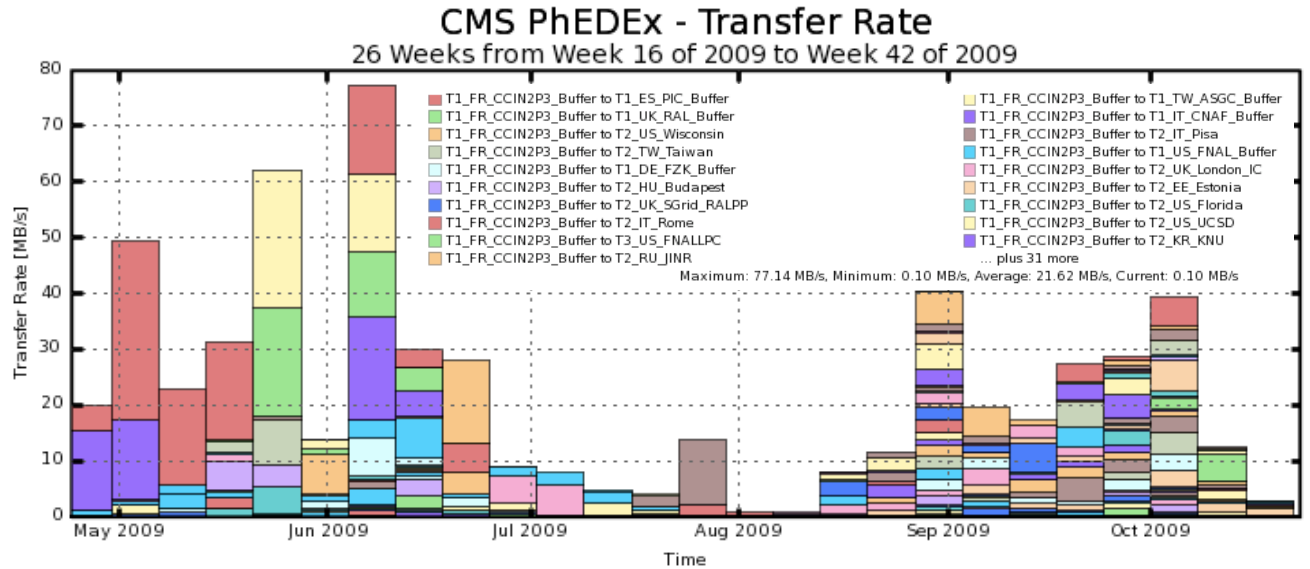




# Transfer rates: IN2P3 -> \*



**IN2P3 -> \***  
**in /Prod**

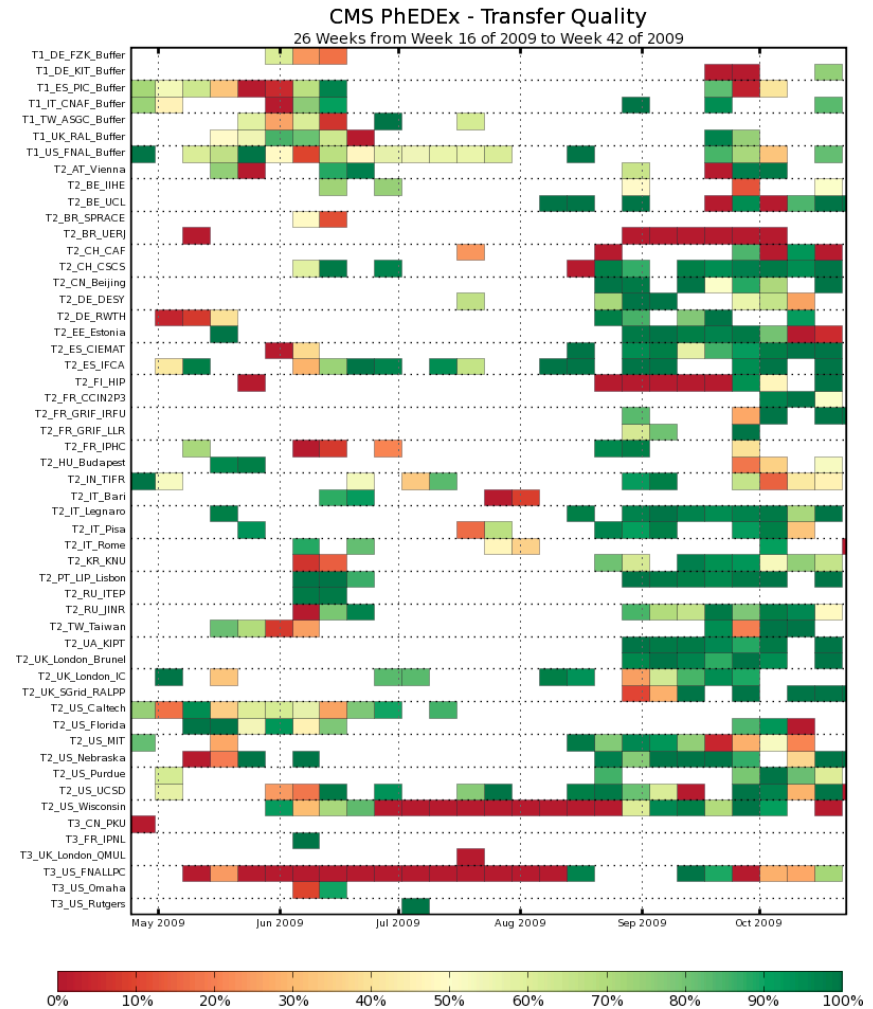
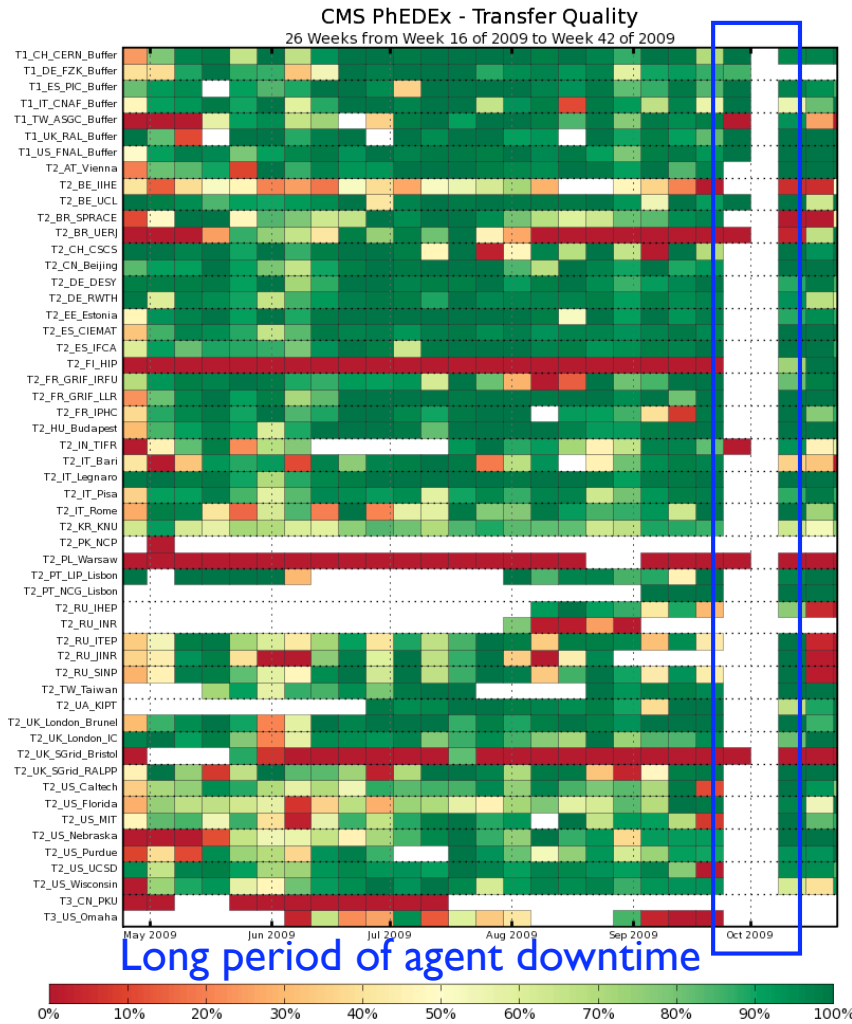




# Transfer quality: IN2P3 -> \*

## IN2P3 -> \* in /Debug

## IN2P3 -> \* in /Prod



Generally OK in /Debug

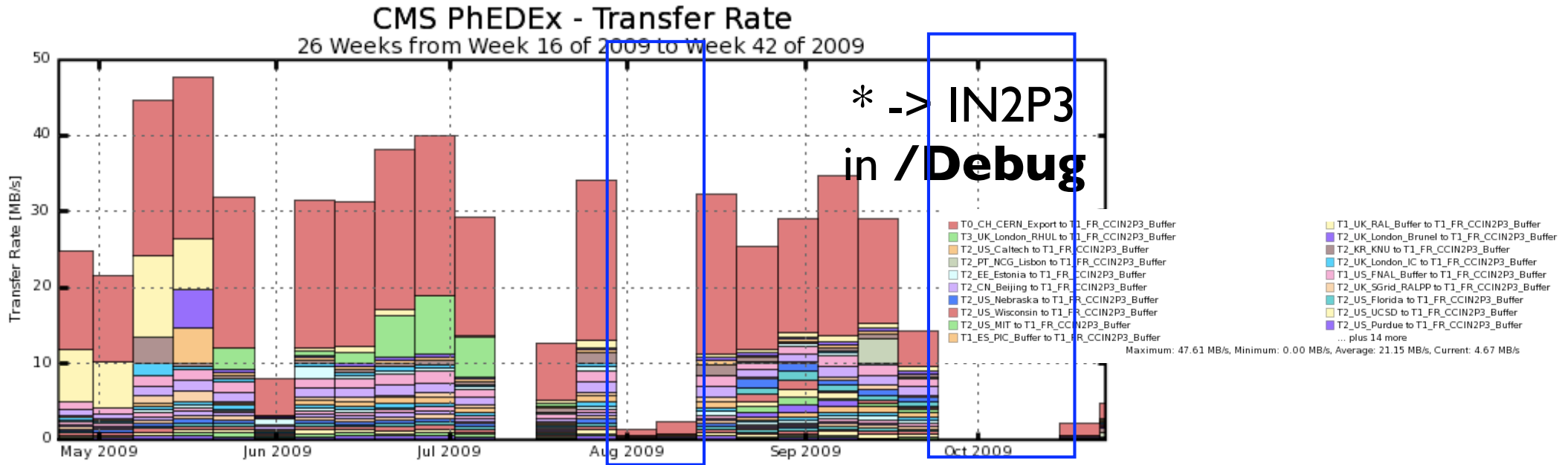
- ✦ apart from the agent downtime in late Sept

Not too bad in /Prod

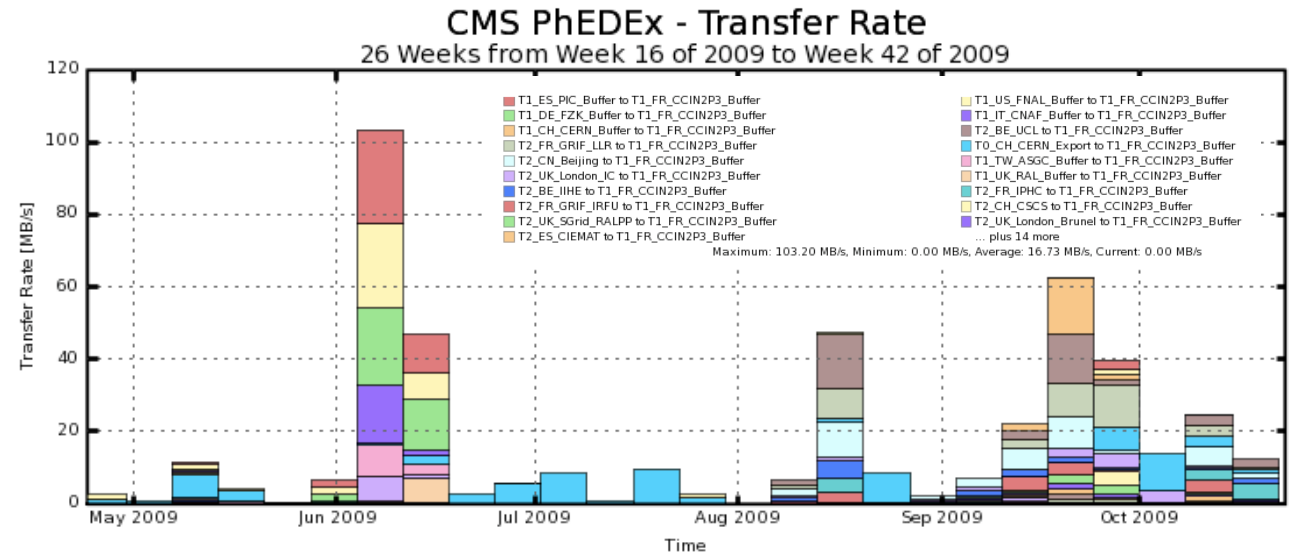
- ✦ Most frequent problem is transfer expirations due to FTS channel congestion - these are invisible in the plots...



# Transfer rates: \* -> IN2P3



\* -> IN2P3  
in /Prod

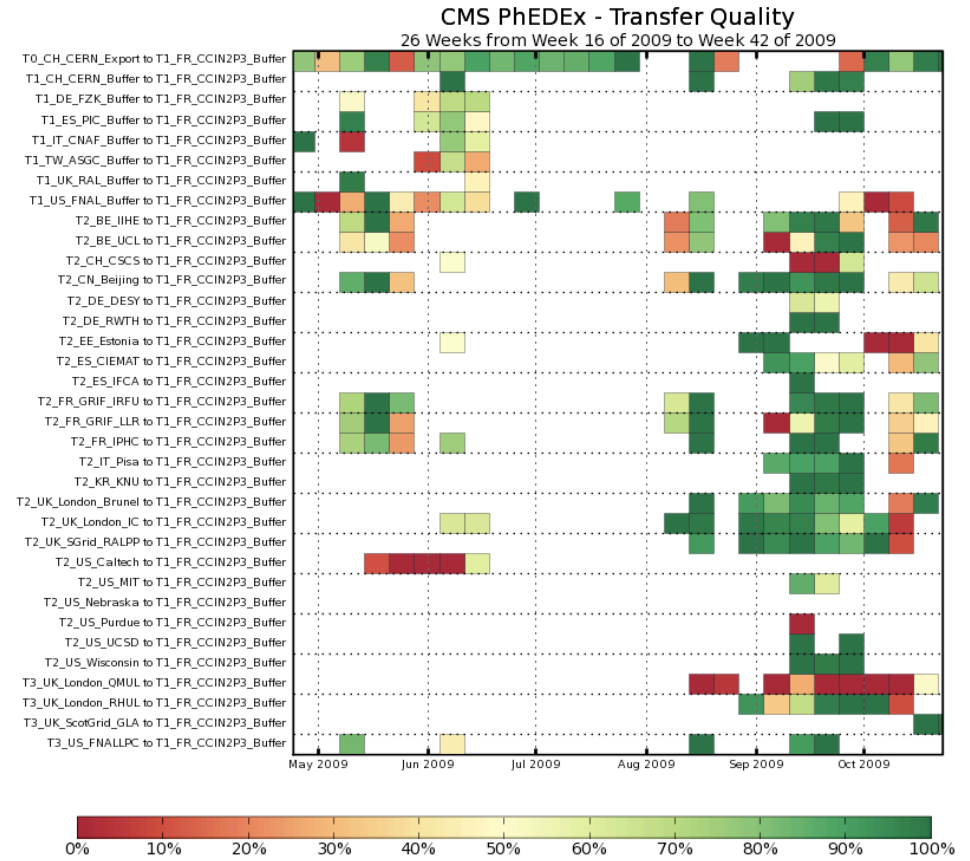
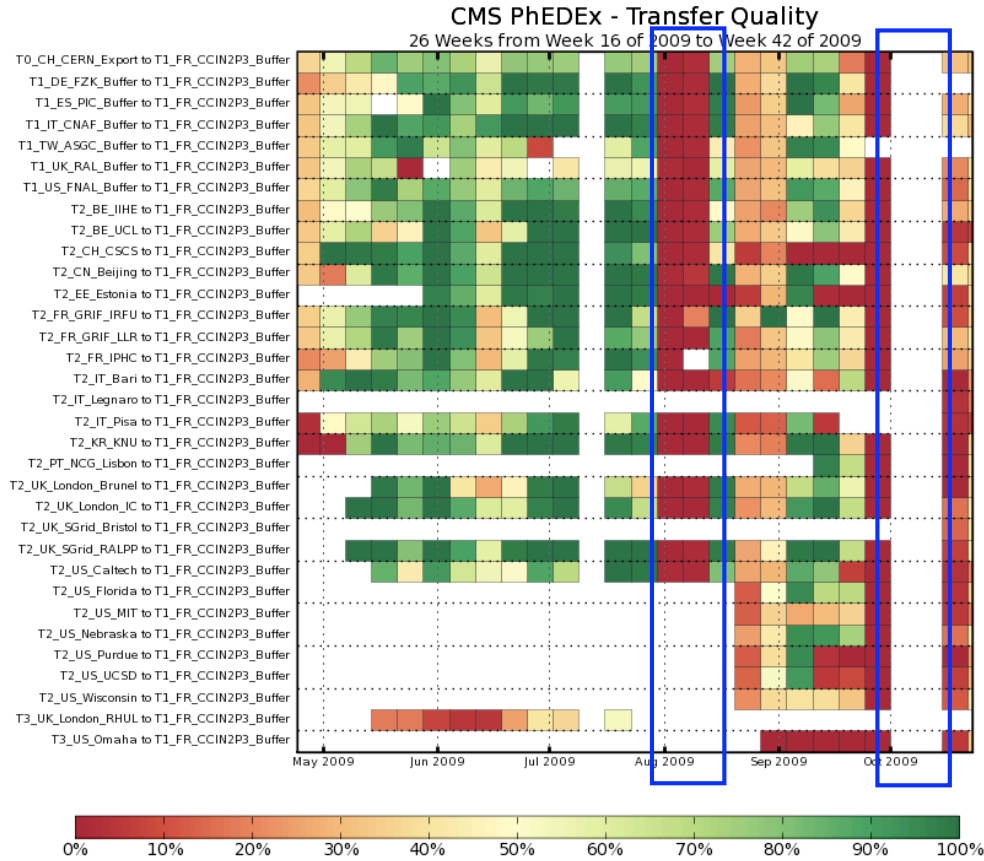




# Transfer quality: \* -> IN2P3

\* -> IN2P3 in /**Debug**

\* -> IN2P3 in /**Prod**



The import in the /Debug instance are more frequently in overall bad health

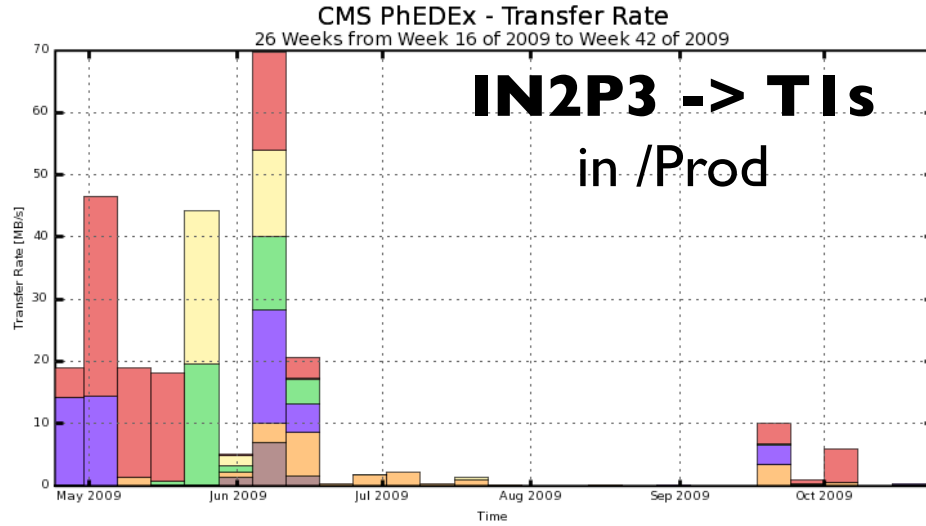
- ◆ agents down for long periods of time
- ◆ Relatively bad transfer quality in imports since summer

A large source of errors is `"*Already have 1 record(s) with pnfsPath=[...]"`

- ◆ probably a cleanup of the LoadTest target area would improve things...

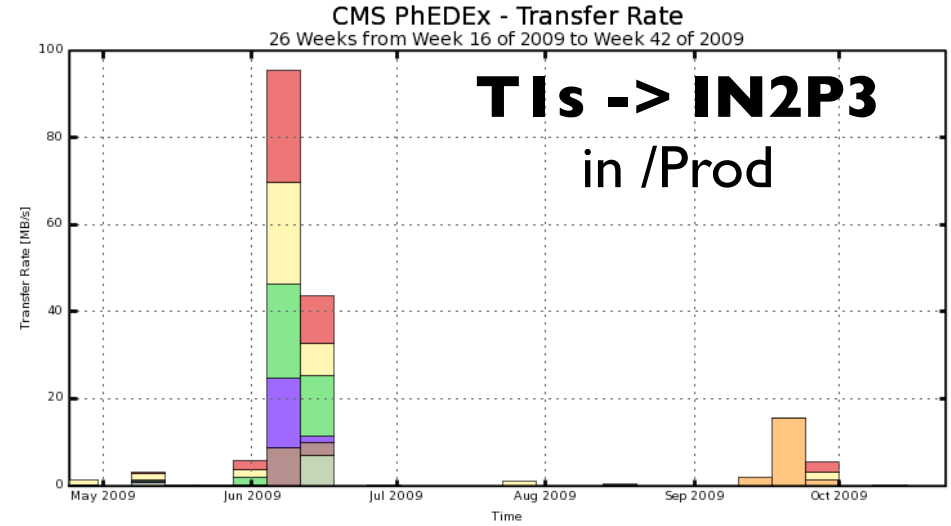


# Transfers: IN2P3 <-> T1's [1/2]



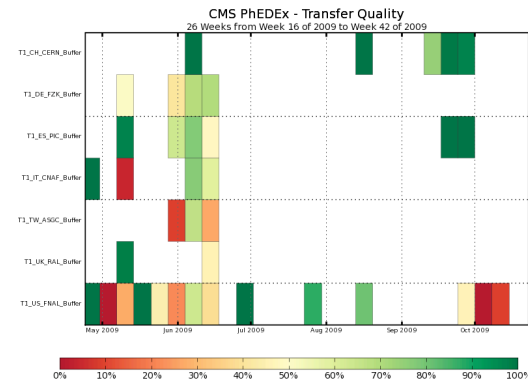
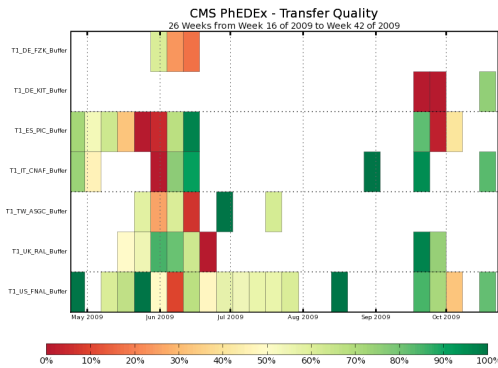
■ T1\_ES\_PIC\_Buffer   
 ■ T1\_TW\_ASGC\_Buffer   
 ■ T1\_US\_FZK\_Buffer   
 ■ T1\_IT\_CNAF\_Buffer   
 ■ T1\_US\_FINAL\_Buffer  
■ T1\_DE\_FZK\_Buffer   
 ■ T1\_DE\_KIT\_Buffer

Maximum: 69.67 MB/s, Minimum: 0.00 MB/s, Average: 10.18 MB/s, Current: 0.39 MB/s



■ T1\_ES\_PIC\_Buffer   
 ■ T1\_US\_FINAL\_Buffer   
 ■ T1\_DE\_FZK\_Buffer   
 ■ T1\_IT\_CNAF\_Buffer   
 ■ T1\_CH\_CERN\_Buffer  
■ T1\_TW\_ASGC\_Buffer   
 ■ T1\_US\_FZK\_Buffer

Maximum: 95.33 MB/s, Minimum: 0.00 MB/s, Average: 6.95 MB/s, Current: 0.00 MB/s



Almost no activity outside STEP09

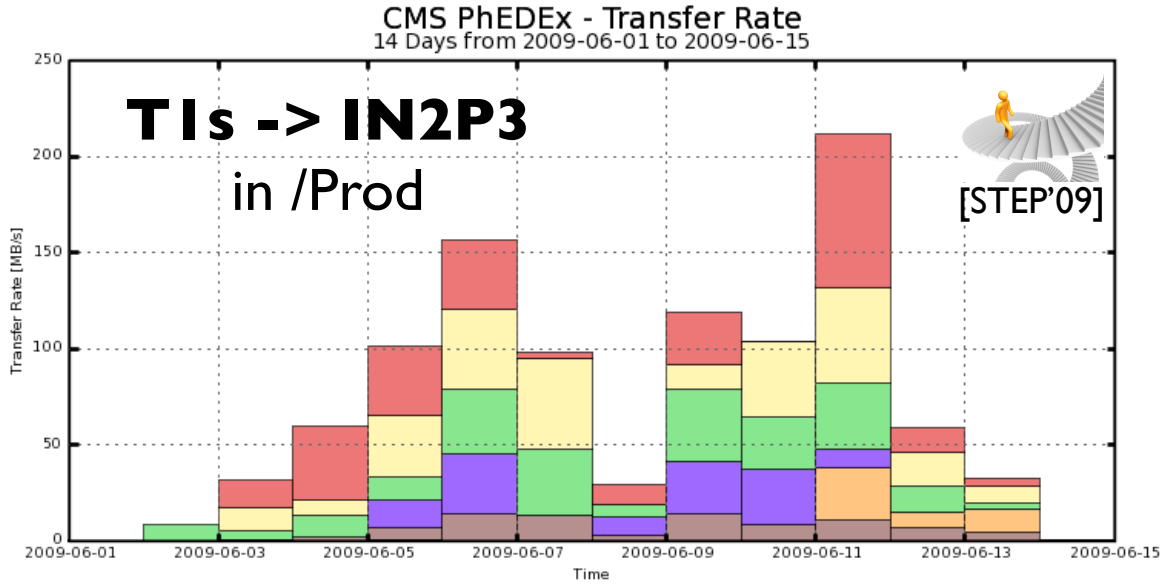
During STEP09, very good rates (*more in the back-up slides*)

- ◆ Targets (assuming no rerouting in PhEDEx) were 185 MB/s in, 105 MB/s out - exceeded in one day



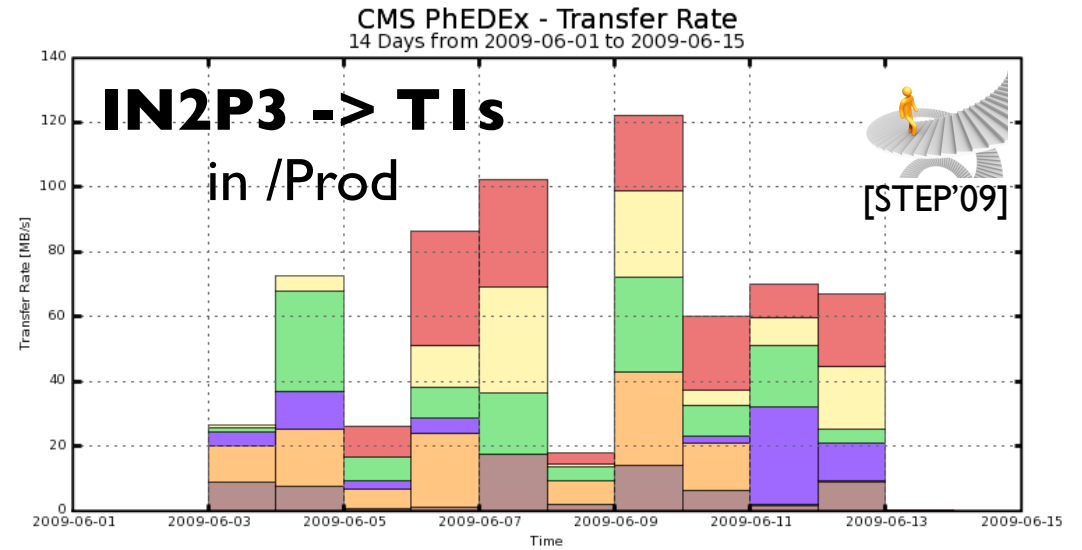


# Transfers: IN2P3 <-> T1's [2/2]



- T1\_DE\_FZK\_Buffer to T1\_FR\_CCIN2P3\_Buffer
- T1\_US\_FNAL\_Buffer to T1\_FR\_CCIN2P3\_Buffer
- T1\_UK\_RAL\_Buffer to T1\_FR\_CCIN2P3\_Buffer
- T1\_CH\_CERN\_Buffer to T1\_FR\_CCIN2P3\_Buffer
- T1\_ES\_PIC\_Buffer to T1\_FR\_CCIN2P3\_Buffer
- T1\_IT\_CNAF\_Buffer to T1\_FR\_CCIN2P3\_Buffer
- T1\_TW\_ASGC\_Buffer to T1\_FR\_CCIN2P3\_Buffer

Maximum: 211.91 MB/s, Minimum: 0.00 MB/s, Average: 72.21 MB/s, Current: 0.50 MB/s



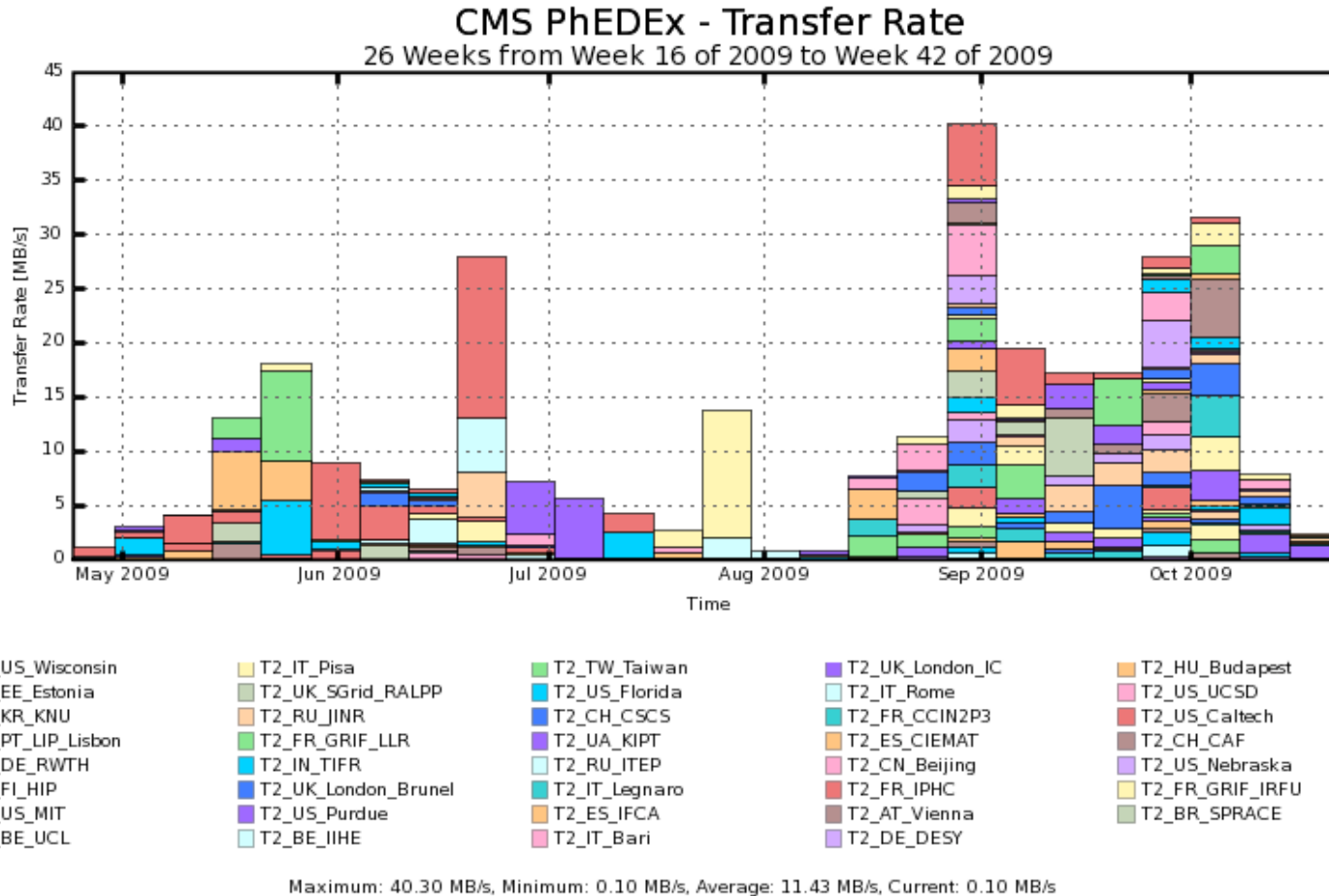
- T1\_FR\_CCIN2P3\_Buffer to T1\_IT\_CNAF\_Buffer
- T1\_FR\_CCIN2P3\_Buffer to T1\_ES\_PIC\_Buffer
- T1\_FR\_CCIN2P3\_Buffer to T1\_TW\_ASGC\_Buffer
- T1\_FR\_CCIN2P3\_Buffer to T1\_UK\_RAL\_Buffer
- T1\_FR\_CCIN2P3\_Buffer to T1\_US\_FNAL\_Buffer
- T1\_FR\_CCIN2P3\_Buffer to T1\_DE\_FZK\_Buffer

Maximum: 122.06 MB/s, Minimum: 0.00 MB/s, Average: 50.07 MB/s, Current: 0.39 MB/s





# Transfers: IN2P3 -> T2's



Constant activity since September

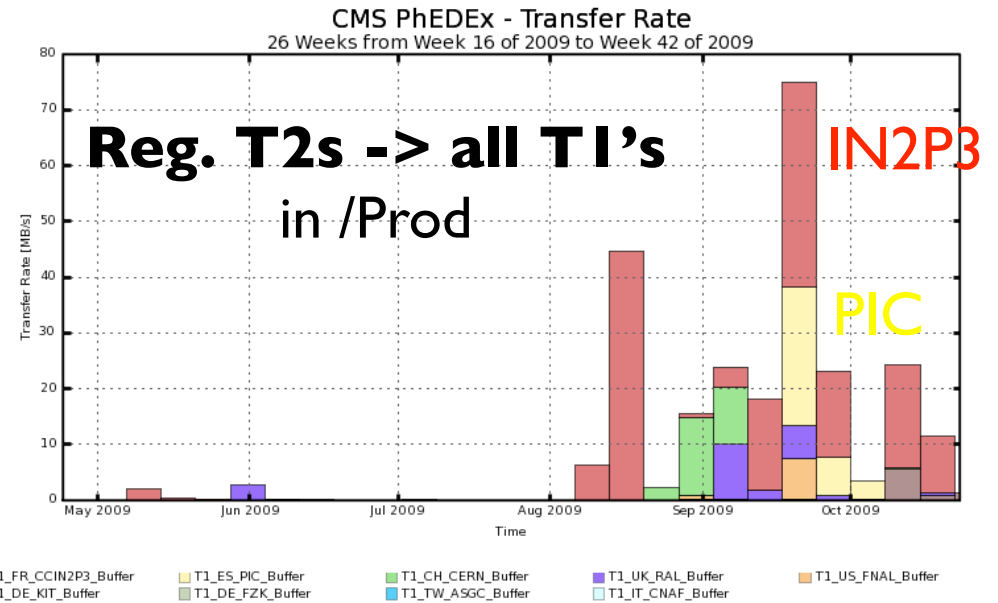
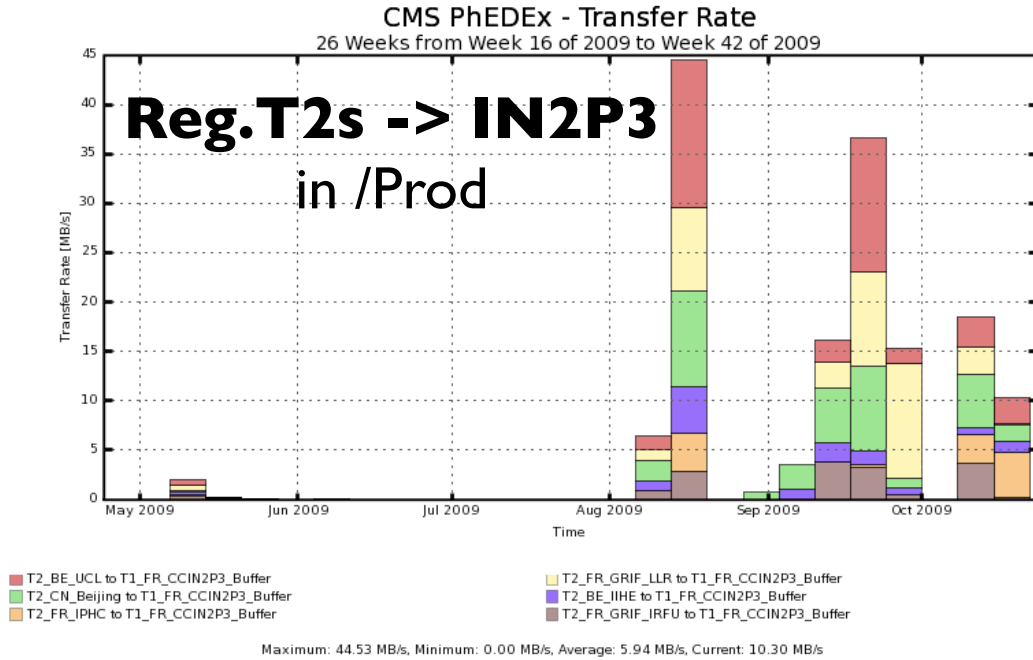
Target in CCRC08 was ~80 MB/s averaged over a long period

- ✦ still below, despite lots of data custodial at T1\_FR\_CCIN2P3 (~600 TB).

DataOps scheduled a round of 'DDT-style' tests IN2P3->T2\_\* last week to measure export rates



# Transfers: T2's -> IN2P3 and other T1's



Assuming CCRC'08 targets:

- the MC production rate from T2s in the France/Belgium/China region averaged over a long period should be 7.4 MB/s

We are way higher than that after the summer



# Link commissioning status

[http://lhweb.pic.es/cms/CommLinksReports/CommissionedLinks\\_Sites.html](http://lhweb.pic.es/cms/CommLinksReports/CommissionedLinks_Sites.html)

## T1\_FR\_IN2P3:

- ◆ Export links commissioned, except for some in T2\_RU/T2\_TR region (rate limitations)
  - [http://cmsweb.cern.ch/phedex/prod/Components::Links?from\\_filter=T1\\_FR&andor=and&to\\_filter=.\\*&Update=Update#](http://cmsweb.cern.ch/phedex/prod/Components::Links?from_filter=T1_FR&andor=and&to_filter=.*&Update=Update#)
- ◆ Import links OK, also some non-regional links (not all of them, though)
  - [http://cmsweb.cern.ch/phedex/prod/Components::Links?from\\_filter=&andor=and&to\\_filter=T1\\_FR&Update=Update#](http://cmsweb.cern.ch/phedex/prod/Components::Links?from_filter=&andor=and&to_filter=T1_FR&Update=Update#)

## T2's in France/Belgium/China region:

- ◆ All fully equipped with downlinks and with many backup uplinks
- ◆ T2\_FR\_CCIN2P3 exports still inactive during namespace migration
- ◆ Remarkably, T2\_FR\_GRIF\_LLJ also has lots of T2<->T2 links



# Ops efficiency and Communication

A good coverage of CMS Ops includes:

- ◆ Fulfill your site contact responsibilities
  - Good summary in DataOps slides (next talk)
- ◆ Attend regularly the Ops weekly meetings
  - Provide the brief weekly report every Monday to FacOps
  - Come prepared and discuss current issues on SAM, JR, ... in full depth
  - Give feedback to DataOps on production activities
- ◆ Give complete and precise answers to questions by FacOps and DataOps
  - Meetings, HN, private communications, ...
- ◆ Ask questions yourself !

Savannah somehow gives a feeling of the rate of issues notifications

- ◆ No Savannah gets opened if a problem is monitored, seen, fixed by CMS contacts onsite before any operator / shifter / user sees it
  - <http://snipurl.com/savannah-in2p3>
    - IN2P3 102, CNAF 84, ASGC 79, FNAL 74, RAL 48, PIC 30, [ KIT 8 - before: FZK, no history]

We strongly rely on CMS contacts at T1 sites for efficient operations



# Back-up



# CMSSW deployment

## CMSSW installed via Grid job on EGEE and OSG sites

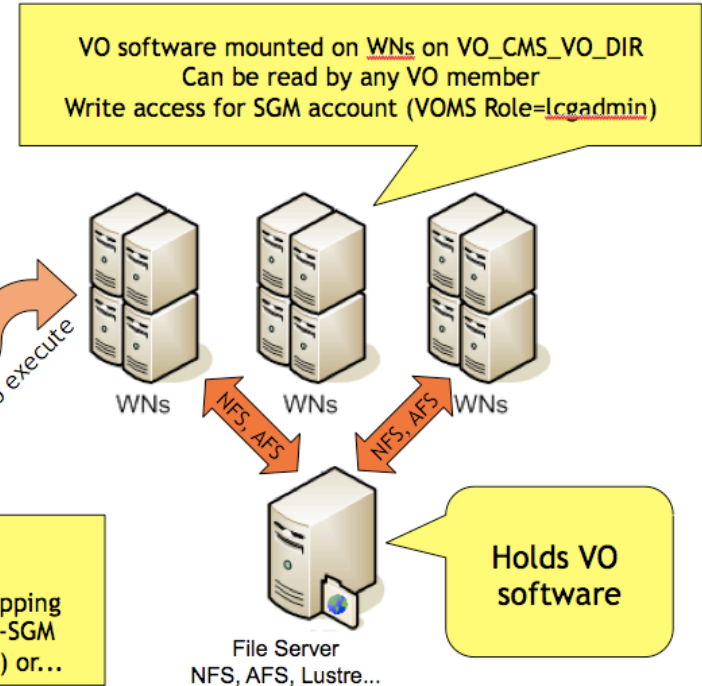
- Basic strategy: use RPM (with apt-get) in CMS SW area

### CMSSW\_314 deployment

- release announcement on Saturday Oct 3rd at 13h22
  - first installation jobs submitted at 13h41
- status: CMSSW\_314 release deployed and ready for Oct-X start-up on da
  - EGEE: submitted to 51 Computing Elements (CE), 44 were DONE after few hrs (see plot); started to follow up on tails over the weekend already
  - OSG: release not tagged into the tag collector, so installed manually; smoothly and quickly completed in most OSG T2/T3



*Credits:*  
 CMSSW deployment teams in Facilities Ops:  
 Bockjoo Kim,  
 Joris Maes,  
 Lukas Vanelderden,  
 Petra Van Mulders,  
 Ilaria Villella,  
 Christoph Wissing



### On EGEE and OSG:

- CMSSW releases get routinely installed smoothly in most sites within few hrs from the release

OSG	Year 2009
Installed Releases (2_2_4 TO 3_3_0)	33
Total number of CEs	25
Total Installations	621
Total Removals	365

*Credits:* Bockjoo Kim



# Ticketing systems [1/2]

## GGUS

## Savannah

Item ID	Summary	Submitted On	Assigned To	Submitted By
#110213	Commissioning Muon T2-T2 links	2009-09-29 13:42	cmscompinfrasup-ddt	afanfani
#110216	WMS setup for cms Role=priorityuser and pool account limit	2009-09-29 16:25	cmscompinfrasup-glitewms	afanfani
#110274	October exercise: Fix required on Padova CE (deny /cms/Role=priorityuser) to avoid running on Padova instead of Legnaro	2009-10-02 09:43	cmscompinfrasup-t2itlegnaro	afanfani
#110291	wms202 and wms218 at CERN show problem in proxy delegation	2009-10-04 11:19	cmscompinfrasup-glitewms	afanfani
#110309	error with working dir, nomore priorityuser mapping, accessing a datasetat Purdue	2009-10-05 09:34	cmscompinfrasup-t2uspurdue	afanfani
#110386	Aborted jobs at Padova for users accessing dataset in Legnaro	2009-10-07 15:59	cmscompinfrasup-t2itlegnaro	afanfani
#110539	Grid Aborted at T2_FR_CCIN2P3	2009-10-13 09:47	cmscompinfrasup-t2frcin2p3	afanfani
#110555	failures accessing MC files (8001/8020) for instance ppMuX/Summer09-MC_31X_V3_SD_DoubleMu-v1/GEN-SIM-RECO	2009-10-13 15:13	cmscompinfrasup-t2uspurdue	afanfani
#110517	Some CMSSW version published but not present	2009-10-12 16:57	cmscompinfrasup-cmsswdeploy	aholguin
#110518	SRMV2 for T3_CO_Uniandes	2009-10-12 17:00	cmscompinfrasup-sam	aholguin
#110550	Problem with CC-IN2P3 AF	2009-10-13 13:27	cmscompinfrasup-t2frcin2p3	andriusj

## GGUS

- ◆ Long tradition of the standard Global Grid User Support system
- Reaches the WLCG site-admins and the fabric-level experts

## Savannah

- ◆ Problem tracking, troubleshooting reference, statistics, ...
- Reaches 'squads' easy to define: CMS contacts at Tiers, tools/services experts, ...
- More: baseline tool for Offline Computing shifts, integrated with other CMS projects, ...





# Ticketing systems [2/2]

Wouldn't a single ticketing system be preferable?

- ◆ Of course. BUT: is there one with all the features CMS uses for Ops?

CMS requested a Savannah-to-GGUS bridging

- ◆ Work finalized. Now ready to be used. Start soon to gain experience in Ops
  - Thanks to Guenter Grein (GGUS), Yves Perrin (LCG/SPI) and Simon Metson (CMS) for their great efforts in the technical implementation and testing

- Main features of the tool
  - those sites who want GGUS should add to their "site-squad" the "GGUS Robot" account.
  - For these sites, when a new Savannah ticket is opened, e.g. by the CMS Computing shifter, the bridging will be activated (It can be manually turned off via Savannah option "Use GGUS == NO")
  - The GGUS ticket is opened by the GGUS Robot. The GGUS cross-reference URL and a "GGUS on hold" status are automatically set in the Savannah ticket
  - Once the GGUS ticket is closed, the Savannah ticket is automatically closed

**To Do:**  
- Provide instructions to CMS sites and for CMS Operators

**Note:**  
- GGUS Alarm/Team tickets treated separately





# STEP'09 :: IN2P3



Pre-staging started on June 8-12th due to scheduled HPSS upgrade

- ◆ Site-operated pre-staging approach was chosen (1)
- ◆ HPSS v.6.2 interfaced to TReqs interface was used
  - files sorting based on the file position on tape

Sizable multi-VO activity throughout STEP'09

High loads observed on HPSS (June 8-13th) (2):

- ◆ Due to all CMS activities simultaneously, in particular CMS analysis at the T2, and also other VOs activities
  - Decided to suspend T2 analysis activity during STEP'09

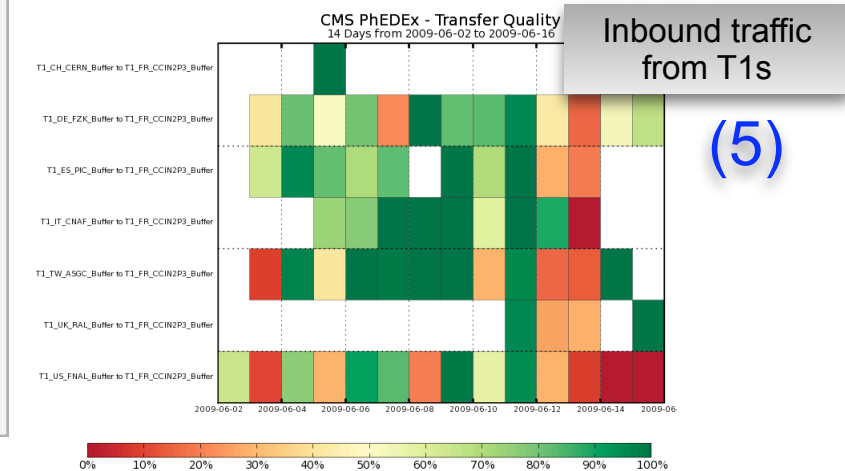
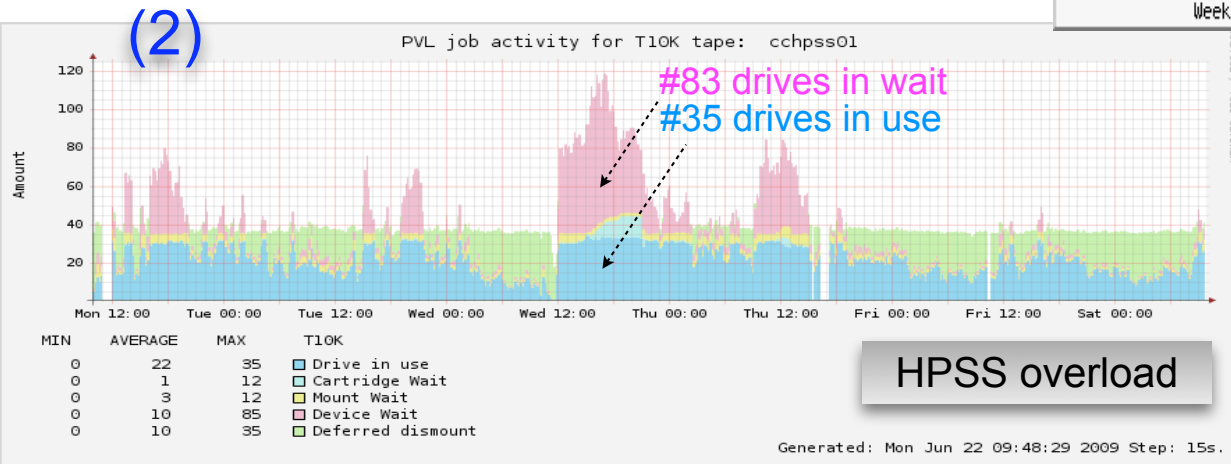
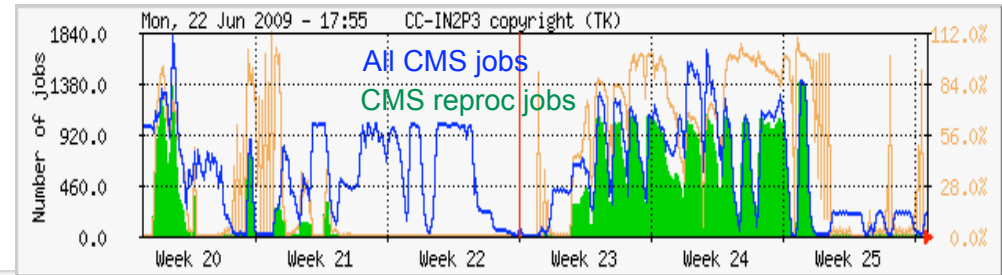
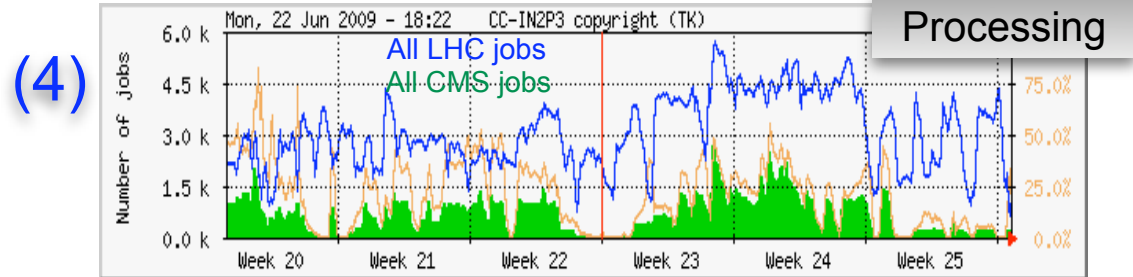
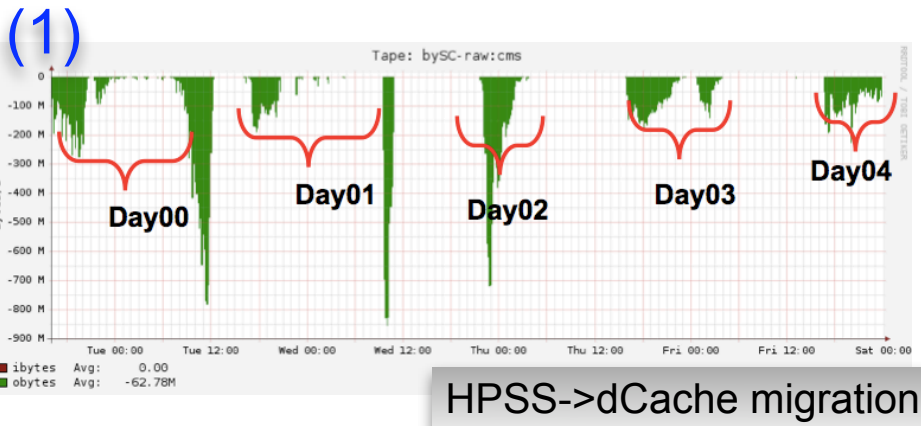
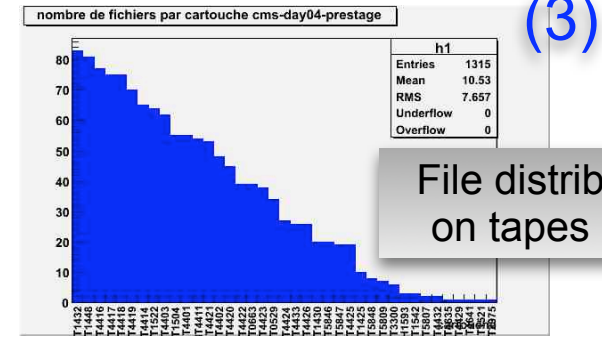
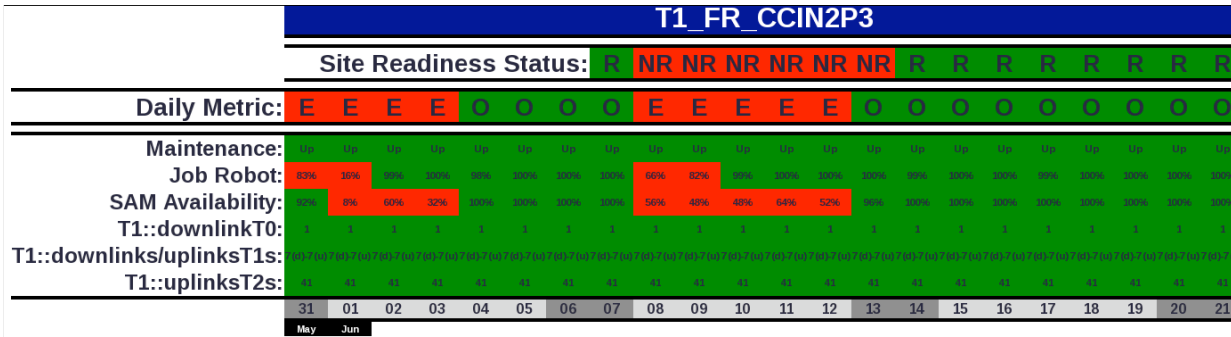
Reprocessing

- ◆ High reprocessing load by CMS and other VOs (4)
  - Failures mainly due to stage-out
- ◆ File distribution per tape on a typical day averages at ~10 (3)

Transfer

- ◆ Relatively smooth
  - some structure (in quality) to be cured, mainly in T1-T1 (5)

# STEP'09 :: IN2P3 in plots



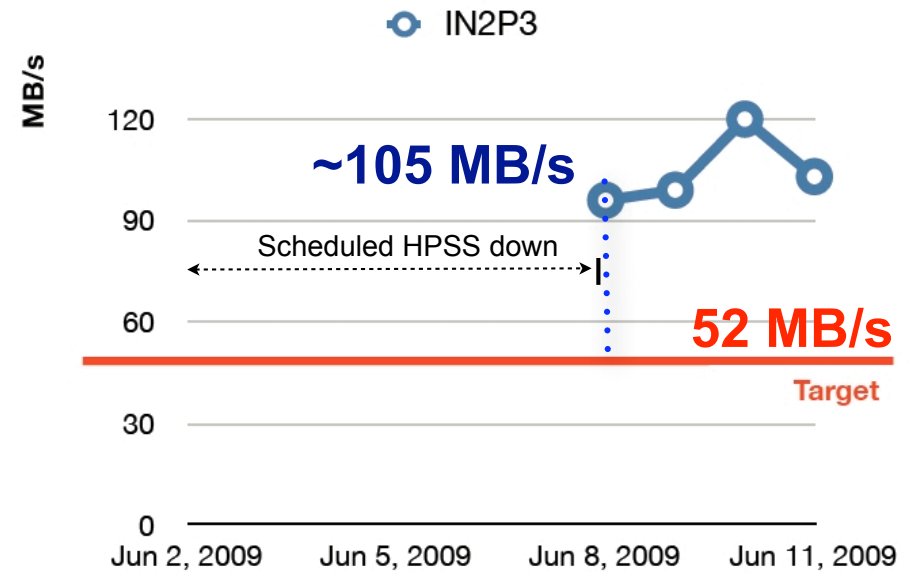


# STEP'09 :: IN2P3

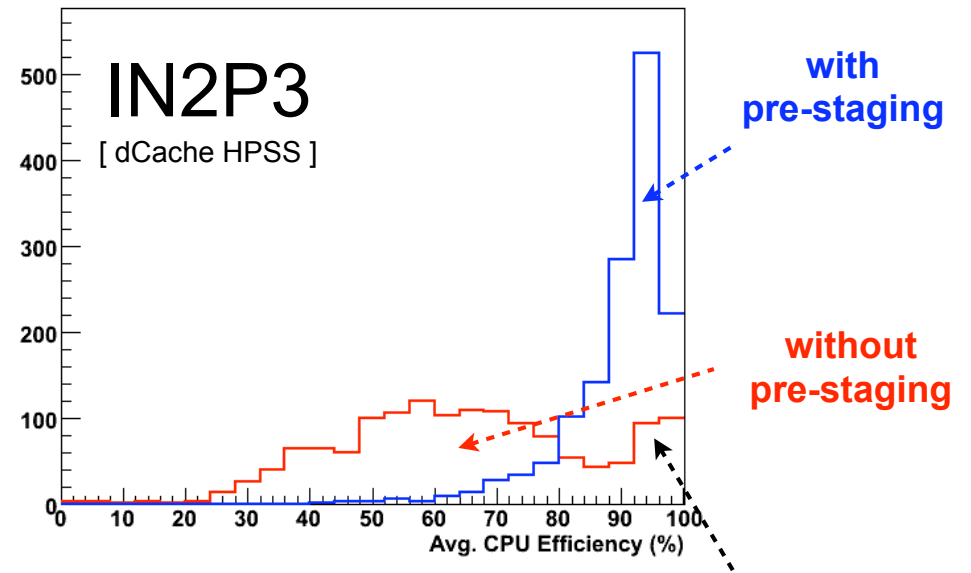
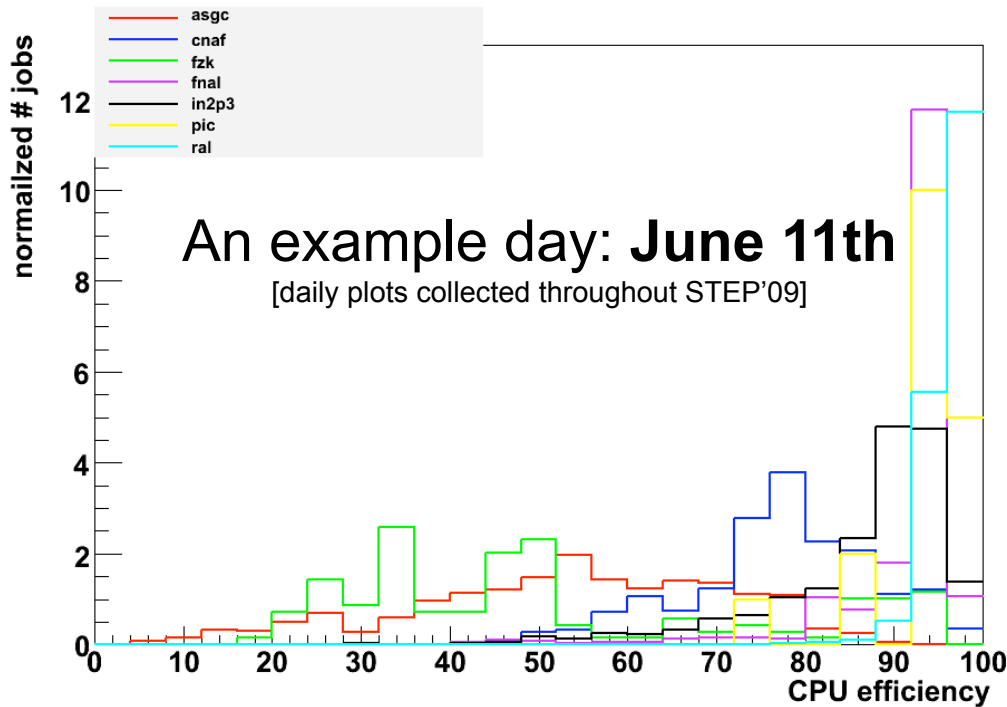
[pre-staging]



	Site-operated	Central SRM script	PhEDEx agent
ASGC			X
CNAF		X	Castor + StoRM: need additional work in PhEDEx
FNAL	X		
FZK	X	Tape issues: preferred manual	Tape issues: preferred manual
IN2P3	X	HPSS downtime on week-1: preferred manual	HPSS downtime on week-1: preferred manual
PIC			X
RAL			X



	Target [MB/s]	2-Jun	3-Jun	4-Jun	5-Jun	6-Jun	7-Jun	8-Jun	9-Jun	10-Jun	11-Jun	
ASGC	73	<i>Digesting migration</i>	140	170	190	160	145	150	140	150	220	
CNAF	56	380	300	160	240	240	270	105	80	125	240	
FNAL	242	280	200	200	120	<i>Still staging previous day</i>	<i>Recovering from backlog</i>		379	380	400	
FZK	85	<i>Tape system not available [unscheduled downtime]</i>							<i>Participated in pre-staging but performance not clear</i>			
IN2P3	52	<i>Tape system not available [scheduled downtime]</i>							96	99	120	103
PIC	50	60	61	106	83	<i>Samples not purged</i>	<i>Samples partially on</i>	99	142	123	142	
RAL	40	250	230	160	140	135	190	170	100	220	180	



Previously failed jobs might have already triggered the pre-staging

Measured every day, at each T1 site. Mixed results:

- ◆ Very good CPU efficiency for **FNAL, IN2P3, (PIC), RAL**
- ◆ Not so good CPU efficiency for **ASGC, CNAF**
- ◆ Test not significant for **FZK**

Current understanding:

- ◆ Test demonstrated the significant effect of pre-staged data for processing
- ◆ Site specifics to be investigated: **IN2P3 not one of these**