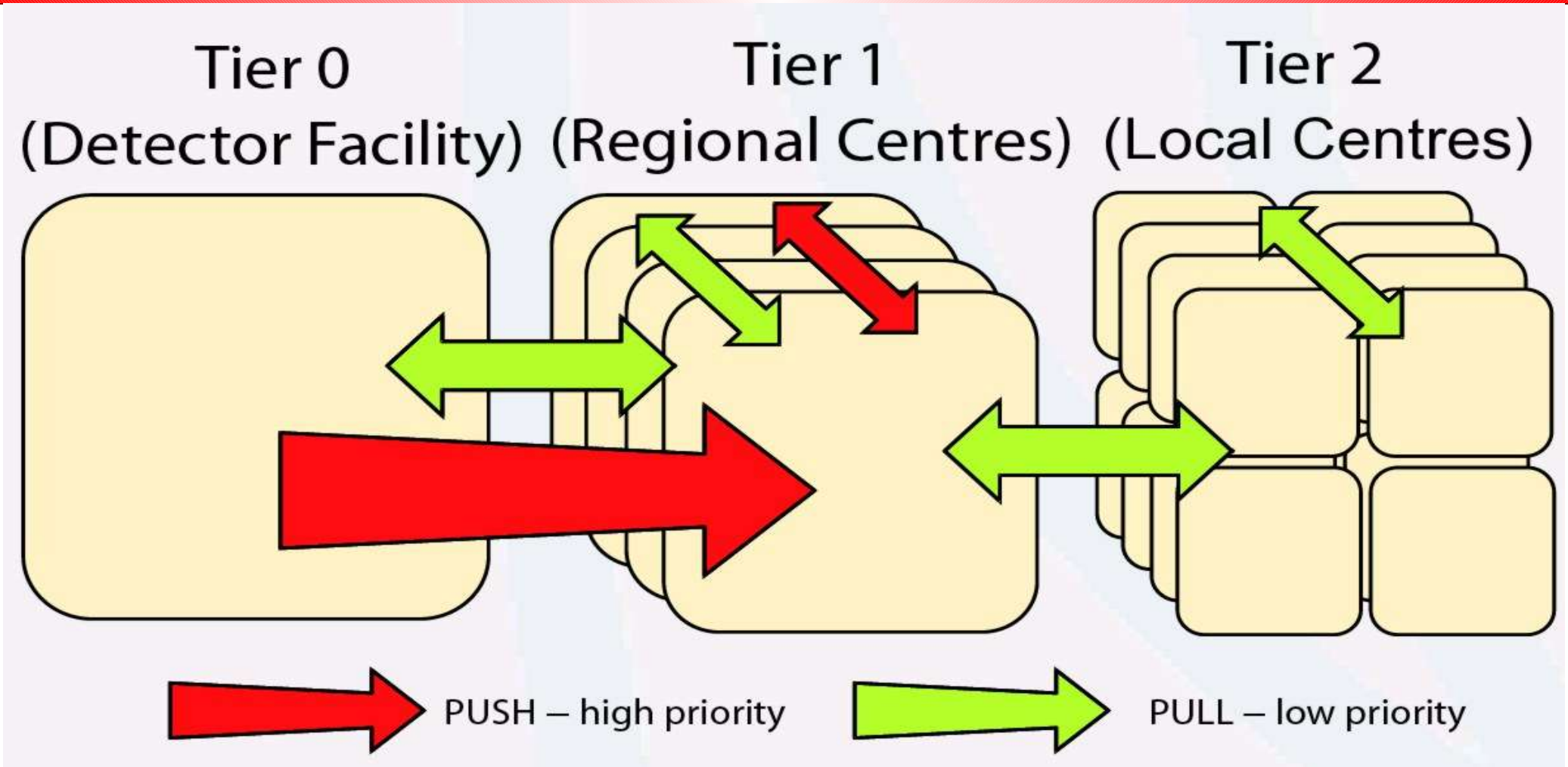
Service Challenge 3 Results

Anja Vest, Uni Karlsruhe

FZK T1-T2 workshop, October 2005

- SC3 goals
- SC3 results
- SC3 PhEDEx status
- Non SC3 activities
- Summary

Tiered data flow → SC3 goal



Main output of SC3:

Data transfer and data serving infrastructure known to work for realistic use

Qualitative SC3 goals (throughput phase)

- Overview of throughput exercise:
  - Throughput from CERN Tier-0 disk to disk and tape at Tier-1's
  - Fan out transfers to selected Tier-2's, same data but less of it
  - Target: transfer and storage systems work and are tuned
    - Using real CMS files and production systems (or to-be production)
    - Sustained operation at required throughput without significant operational interference / maintenance
- Concretely
  - Part 1: Data from disk buffer at CERN first to Tier-1/2 disks
    - Tier-2's will be subscribed subset of the data going to Tier-1's
    - Data to Tier-2's are routed via Tier-1's
  - Part 2: Same, but data goes to tape at Tier-1s
  - Transfers managed by PhEDEx
  - Files registered to local file catalogue
  - Sufficient monitoring

Quantitative SC3 goals (throughput phase)

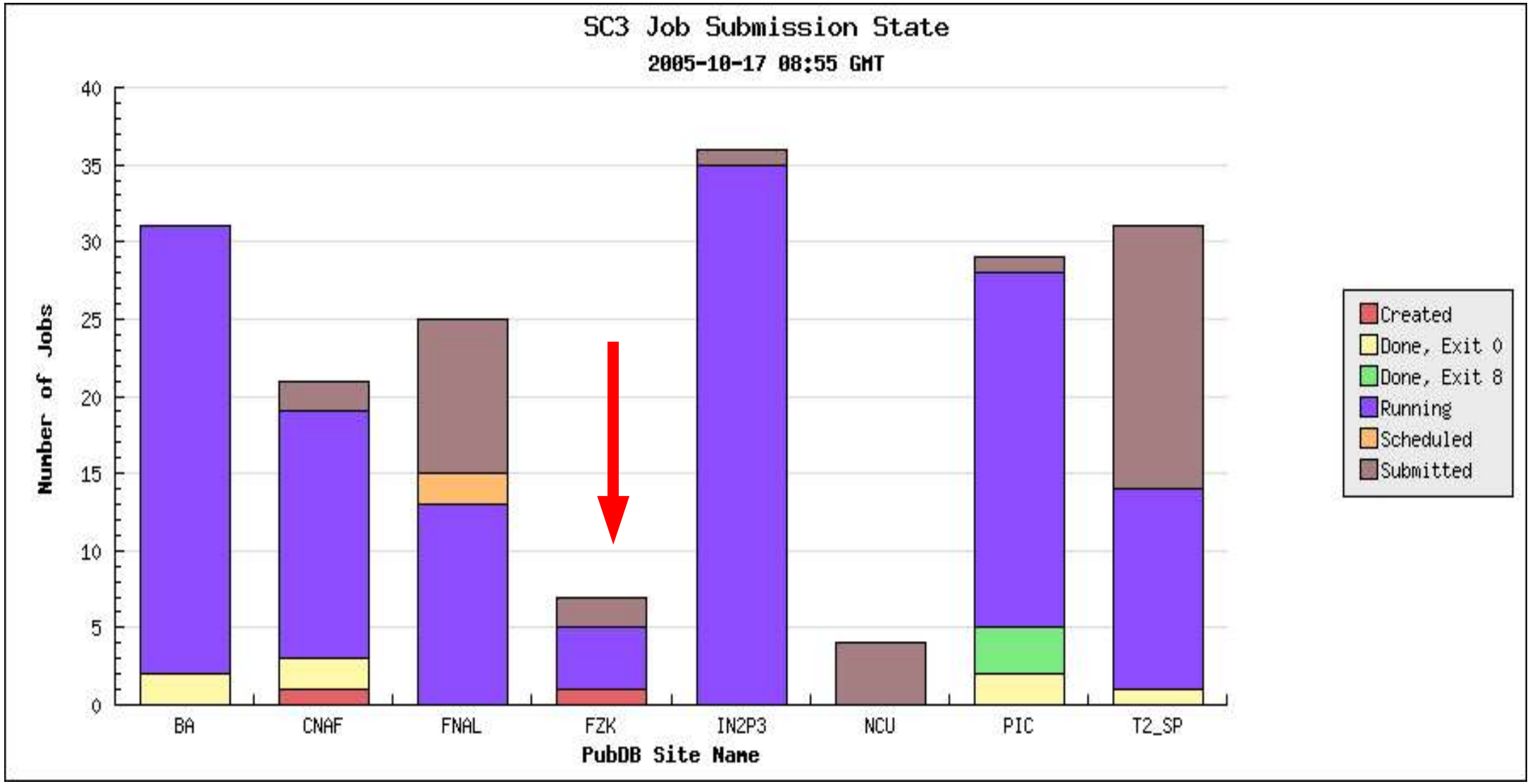
- Rates defined:
  - T0 disk to T1 disk — 150 MB/s sustained
  - T0 disk to T1 tape — 60 MB/s sustained
  - T1 disk/tape to T2 disk — ? MB/s sustained
  - T2 disk to T1 disk — < 1 MB/s sustained
  - Suggested informally 30 MB/s T1 to T2 if bandwidth is available
- Service quality:
  - Transfer failures should have no significant impact on rate
  - Transfer failures — < 0.1% of files more than 5
  - Catalogue failures after transfer — < 0.1% of files
  - File migration to tapes — (keep up with transfers)

Service challenge 3 results

- Job submission status (slides)
- Data transfers in throughput phase:
  - Micro participation with PhEDEx
  - GridKa (T1) ↔ DESY (T2) established
  - CERN (T0) → GridKa (T1)
  - sustained rates: ~ 60 MB/s
  - peak rates: 100 MB/s
  - Monitoring (slides):
    - Replica state
    - Data transfers
    - Transfer quality

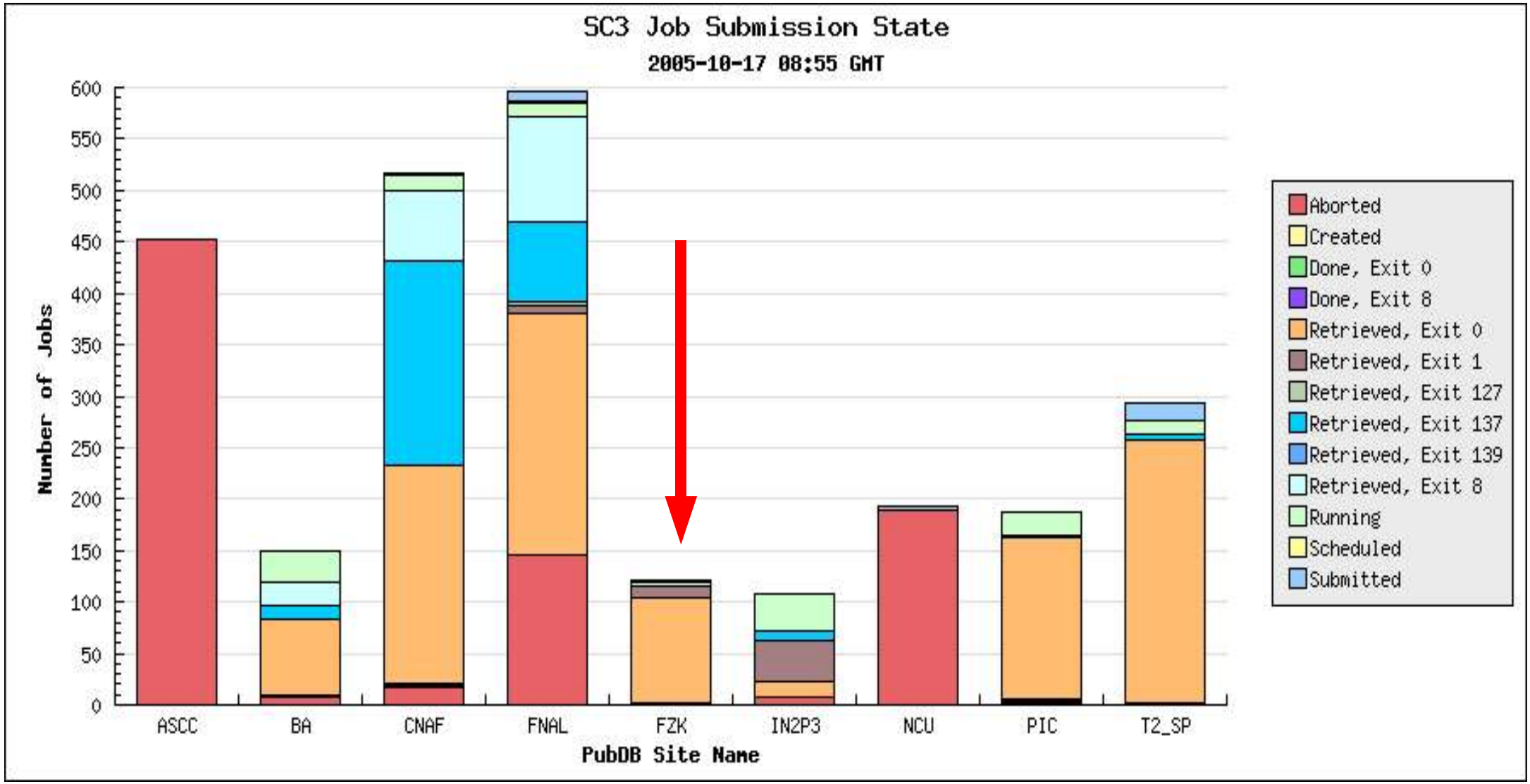
SC3 job submission status I

Actual status:

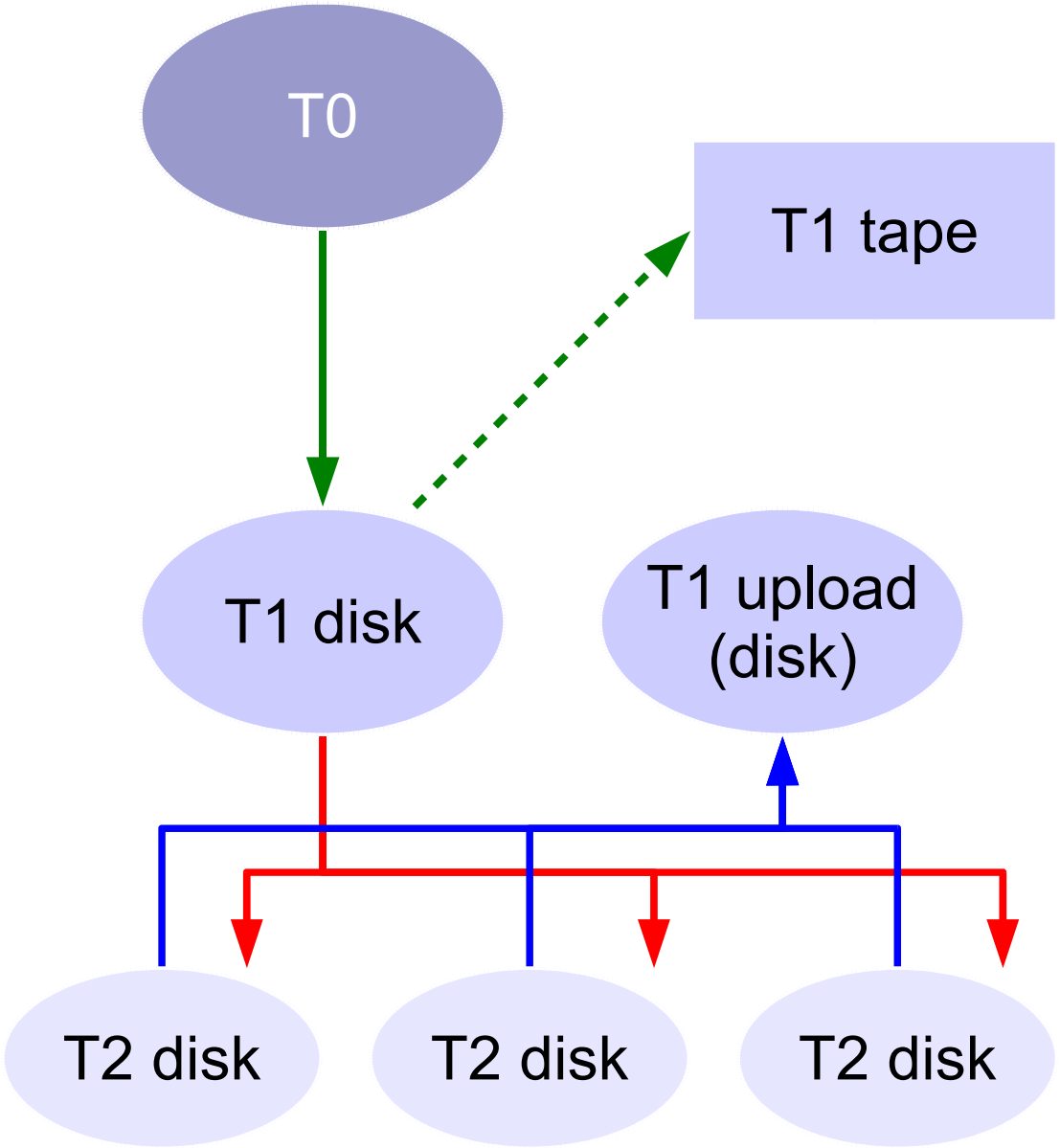


SC3 job submission status II

Average success over time:



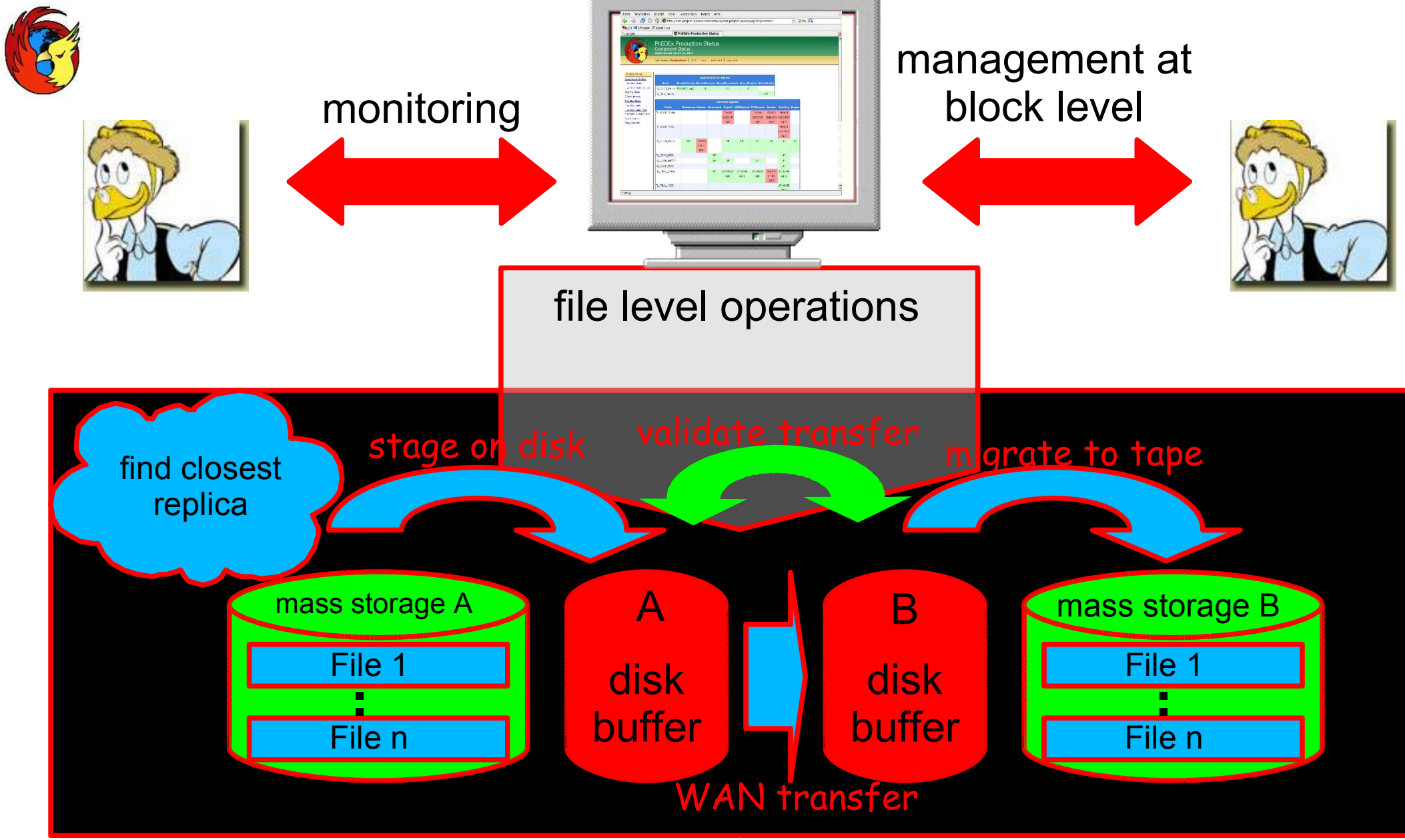
Data transfers T1 → T2 → T1 in SC3



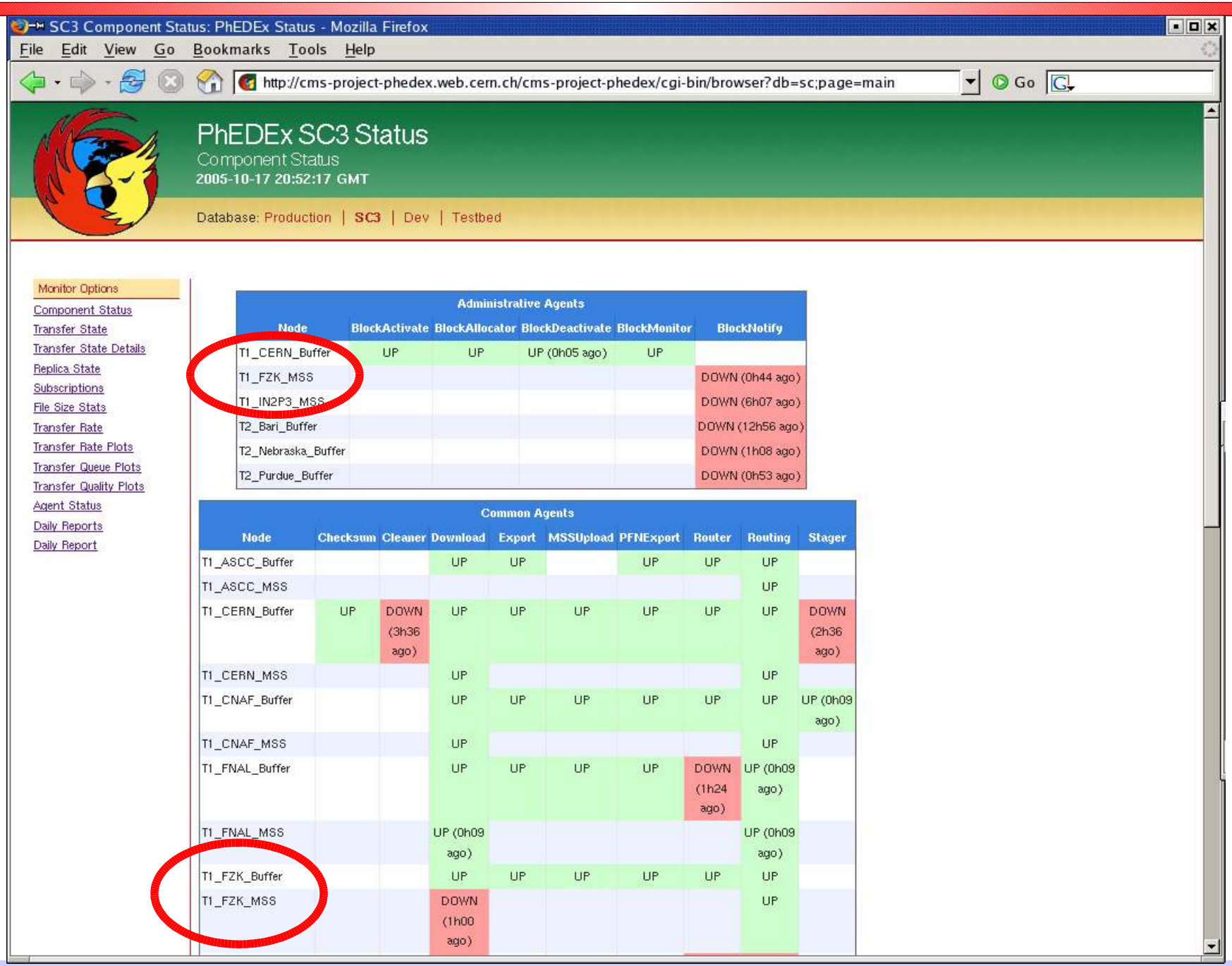
GridKa / DESY:

- Data export from GridKa to DESY
- Site services:
  - Storage: dCache/SRM (+ tape at DESY)
  - Transfers: PhEDEx/SRM (srmcp)
  - File catalogue: POOL, MySQL

HEP data replication, PhEDEx workflow



SC3: PhEDEx component status



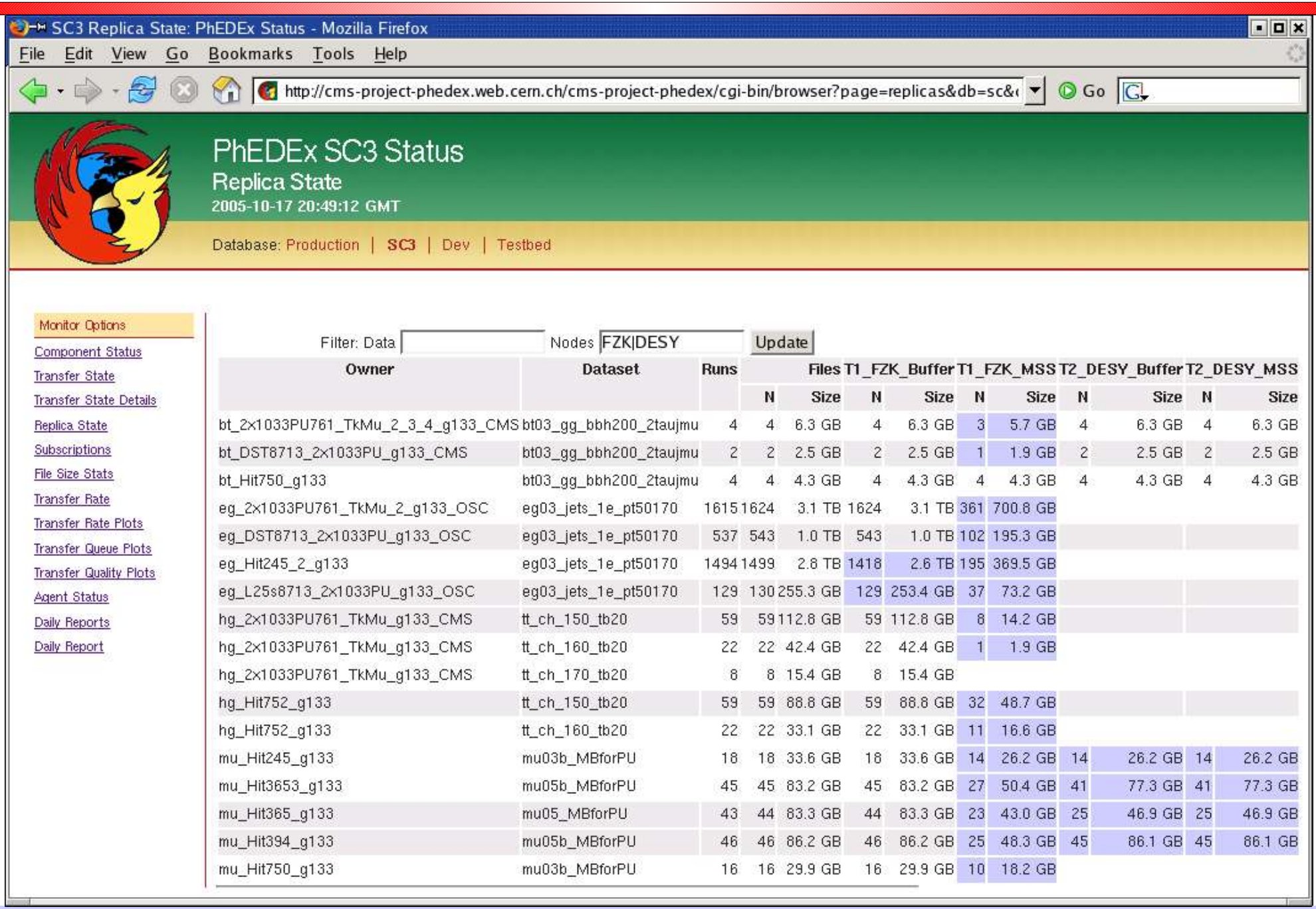
Administrative Agents

Node	BlockActivate	BlockAllocator	BlockDeactivate	BlockMonitor	BlockNotify
T1_CERN_Buffer	UP	UP	UP (0h05 ago)	UP	
T1_FZK_MSS					DOWN (0h44 ago)
T1_IN2P3_MSS					DOWN (6h07 ago)
T2_Bari_Buffer					DOWN (12h56 ago)
T2_Nebraska_Buffer					DOWN (1h08 ago)
T2_Purdue_Buffer					DOWN (0h53 ago)

Common Agents

Node	Checksum	Cleaner	Download	Export	MSSUpload	PFNExport	Router	Routing	Stager
T1_ASCC_Buffer			UP	UP		UP	UP	UP	
T1_ASCC_MSS								UP	
T1_CERN_Buffer	UP	DOWN (3h36 ago)	UP	UP	UP	UP	UP	UP	DOWN (2h36 ago)
T1_CERN_MSS			UP					UP	
T1_CNAF_Buffer			UP	UP	UP	UP	UP	UP	UP (0h09 ago)
T1_CNAF_MSS			UP					UP	
T1_FNAL_Buffer			UP	UP	UP	UP	DOWN (1h24 ago)	UP (0h09 ago)	
T1_FNAL_MSS			UP (0h09 ago)					UP (0h09 ago)	
T1_FZK_Buffer			UP	UP	UP	UP	UP	UP	
T1_FZK_MSS			DOWN (1h00 ago)					UP	

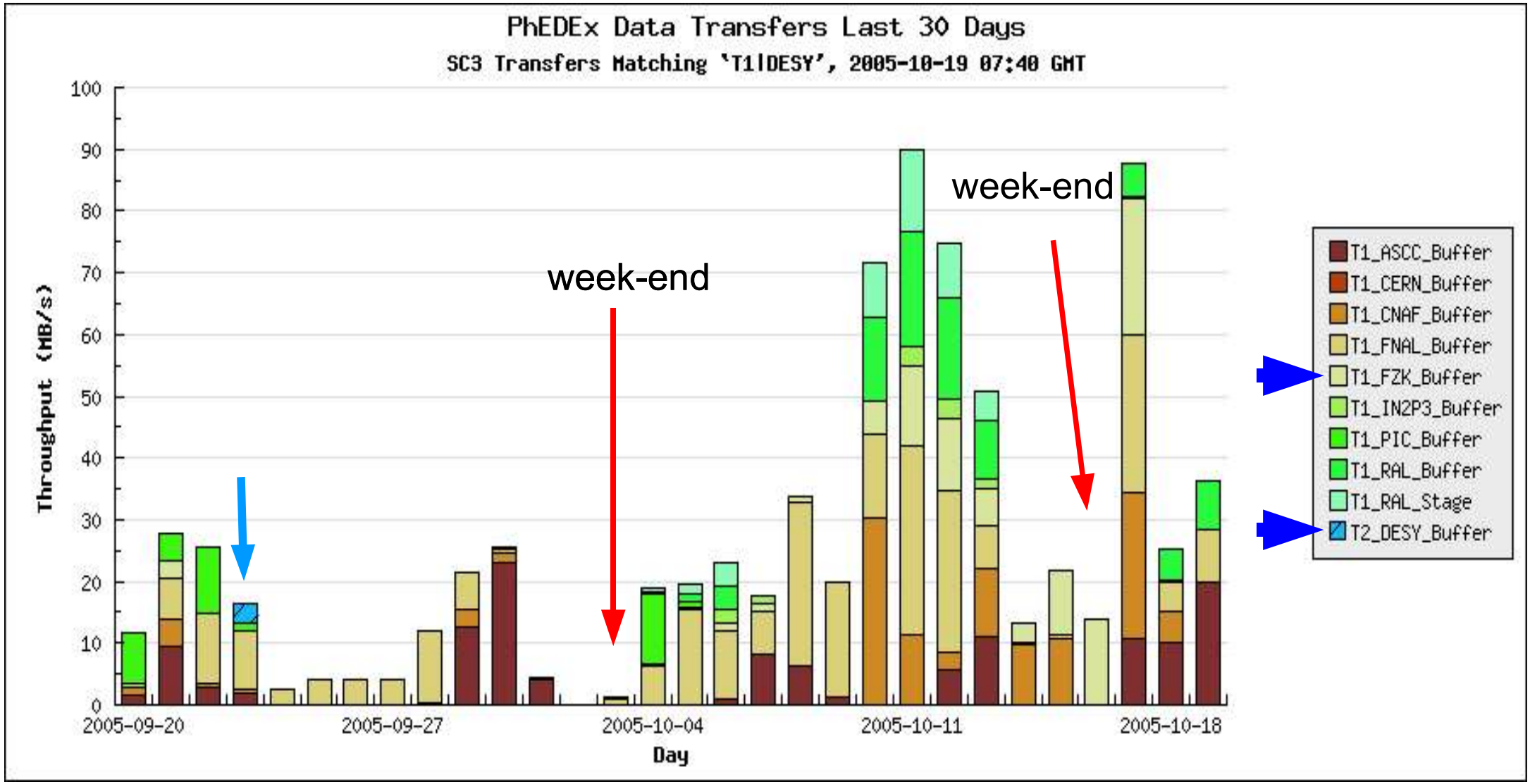
SC3: PhEDEx replica status



Owner	Dataset	Runs	Files		T1_FZK_Buffer		T1_FZK_MSS		T2_DESY_Buffer		T2_DESY_MSS	
			N	Size	N	Size	N	Size	N	Size	N	Size
bt_2x1033PU761_TkMu_2_3_4_g133_CMS	bt03_gg_bbh200_2taujmu	4	4	6.3 GB	4	6.3 GB	3	5.7 GB	4	6.3 GB	4	6.3 GB
bt_DST8713_2x1033PU_g133_CMS	bt03_gg_bbh200_2taujmu	2	2	2.5 GB	2	2.5 GB	1	1.9 GB	2	2.5 GB	2	2.5 GB
bt_Hit750_g133	bt03_gg_bbh200_2taujmu	4	4	4.3 GB	4	4.3 GB	4	4.3 GB	4	4.3 GB	4	4.3 GB
eg_2x1033PU761_TkMu_2_g133_OSC	eg03_jets_1e_pt50170	1615	1624	3.1 TB	1624	3.1 TB	361	700.8 GB				
eg_DST8713_2x1033PU_g133_OSC	eg03_jets_1e_pt50170	537	543	1.0 TB	543	1.0 TB	102	195.3 GB				
eg_Hit245_2_g133	eg03_jets_1e_pt50170	1494	1499	2.8 TB	1418	2.6 TB	195	369.5 GB				
eg_L25s8713_2x1033PU_g133_OSC	eg03_jets_1e_pt50170	129	130	255.3 GB	129	253.4 GB	37	73.2 GB				
hg_2x1033PU761_TkMu_g133_CMS	tt_ch_150_tb20	59	59	112.8 GB	59	112.8 GB	8	14.2 GB				
hg_2x1033PU761_TkMu_g133_CMS	tt_ch_160_tb20	22	22	42.4 GB	22	42.4 GB	1	1.9 GB				
hg_2x1033PU761_TkMu_g133_CMS	tt_ch_170_tb20	8	8	15.4 GB	8	15.4 GB						
hg_Hit752_g133	tt_ch_150_tb20	59	59	88.8 GB	59	88.8 GB	32	48.7 GB				
hg_Hit752_g133	tt_ch_160_tb20	22	22	33.1 GB	22	33.1 GB	11	16.6 GB				
mu_Hit245_g133	mu03b_MBforPU	18	18	33.6 GB	18	33.6 GB	14	26.2 GB	14	26.2 GB	14	26.2 GB
mu_Hit3653_g133	mu05b_MBforPU	45	45	83.2 GB	45	83.2 GB	27	50.4 GB	41	77.3 GB	41	77.3 GB
mu_Hit365_g133	mu05_MBforPU	43	44	83.3 GB	44	83.3 GB	23	43.0 GB	25	46.9 GB	25	46.9 GB
mu_Hit394_g133	mu05b_MBforPU	46	46	86.2 GB	46	86.2 GB	25	48.3 GB	45	86.1 GB	45	86.1 GB
mu_Hit750_g133	mu03b_MBforPU	16	16	29.9 GB	16	29.9 GB	10	18.2 GB				

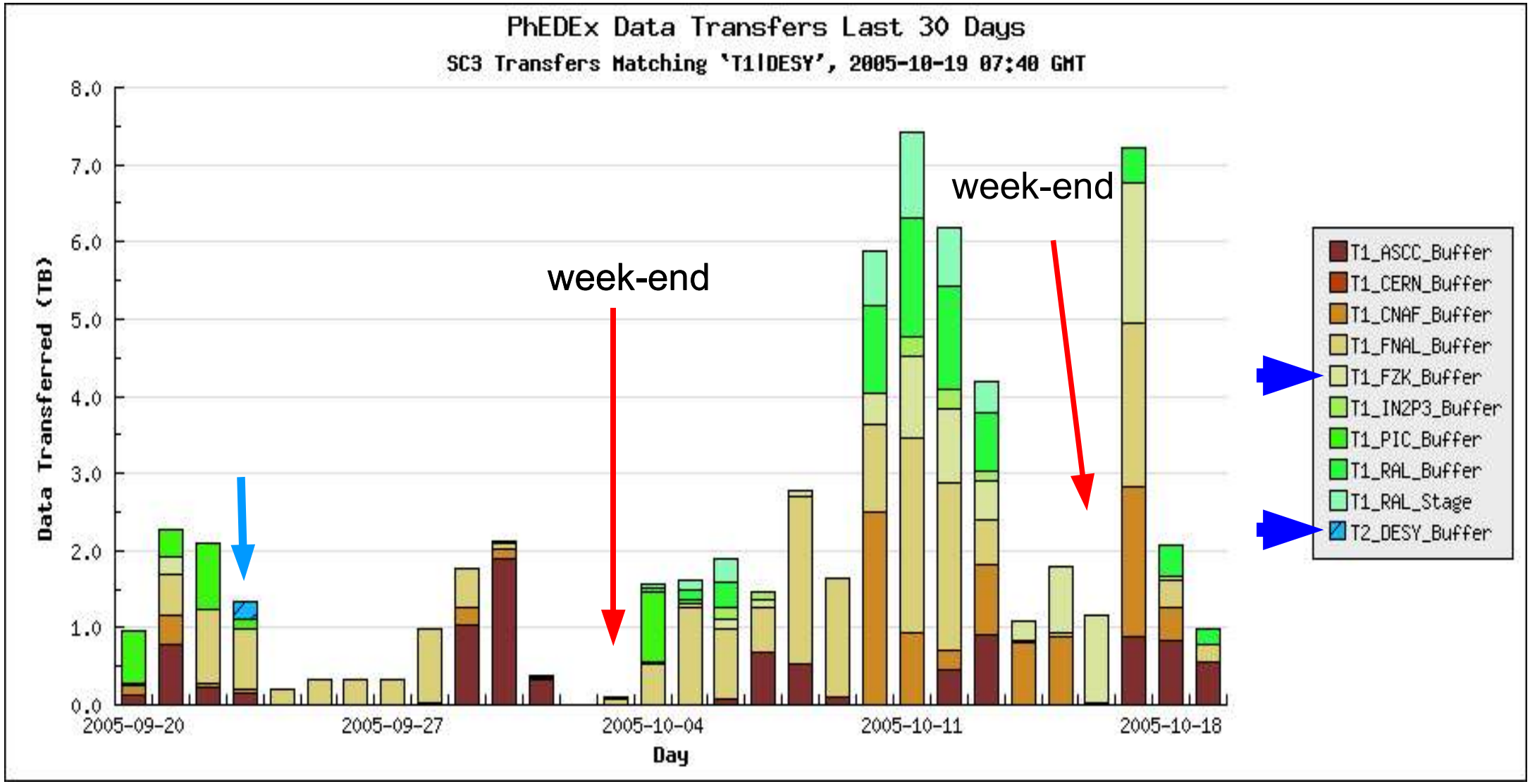
SC3: PhEDEx data transfers

Throughput (MB/s) in the last 30 days:



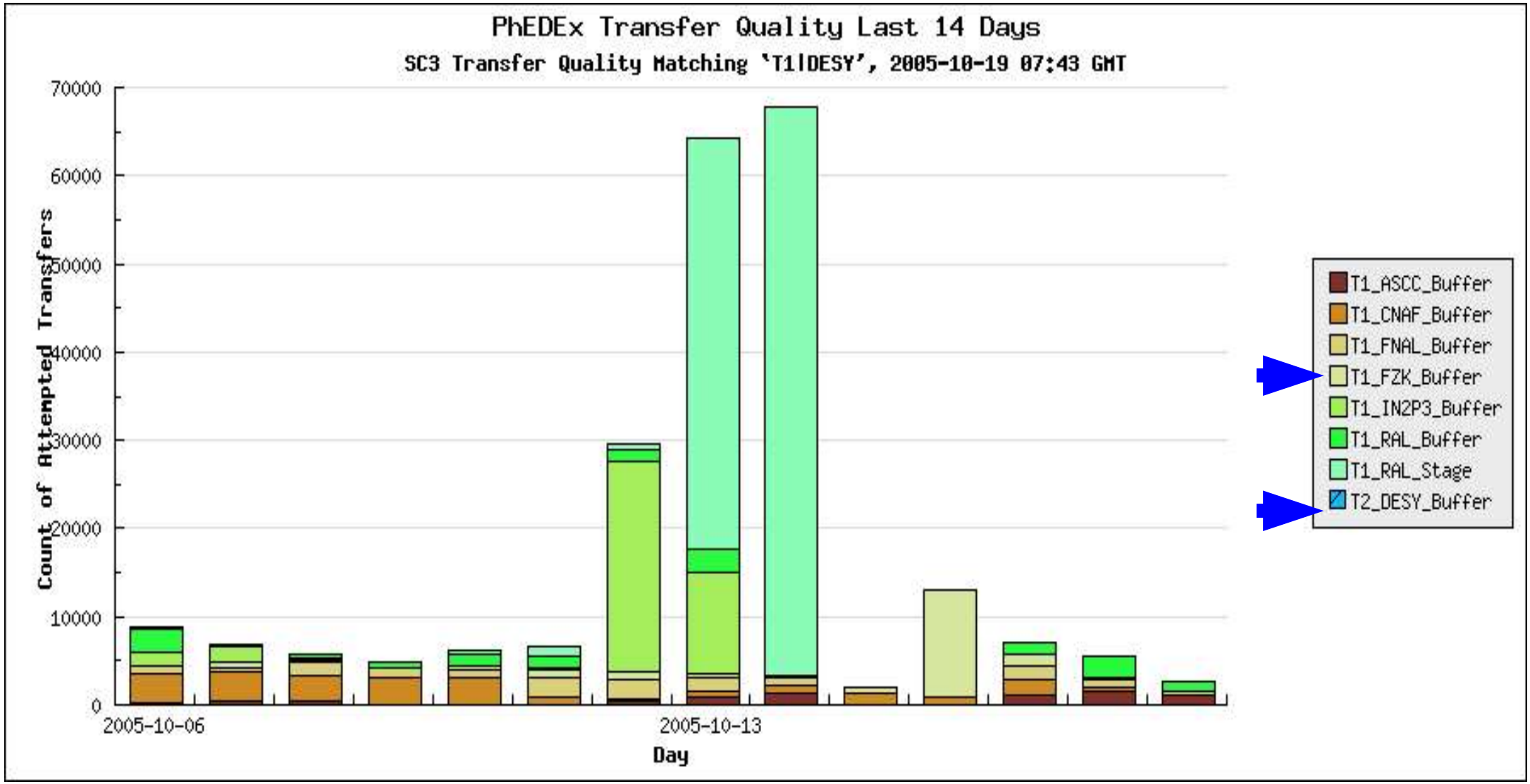
SC3: PhEDEx data transfers

Data transferred (TB) in the last 30 days:



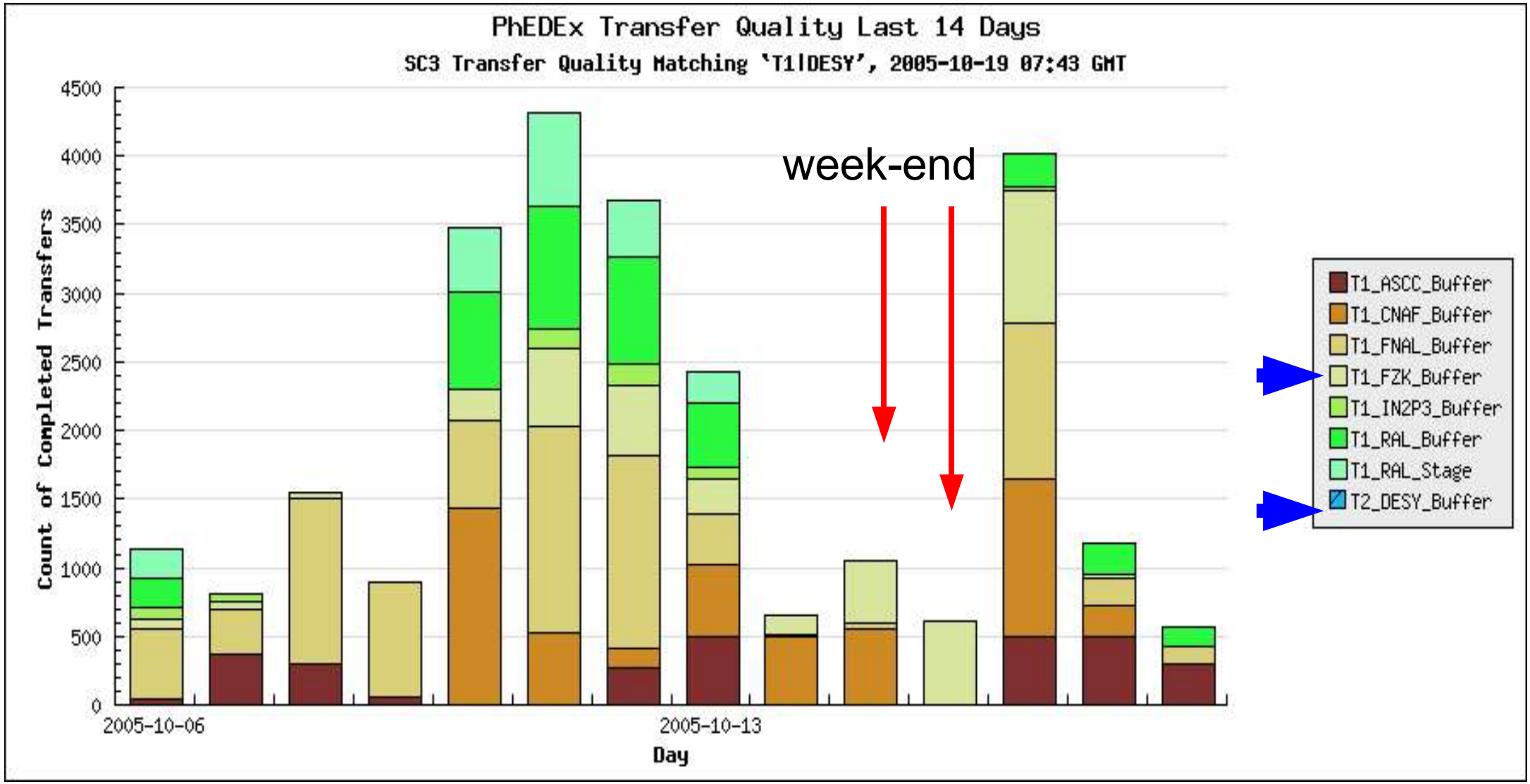
SC3: PhEDEx transfer quality

Attempted transfers in the last 14 days:



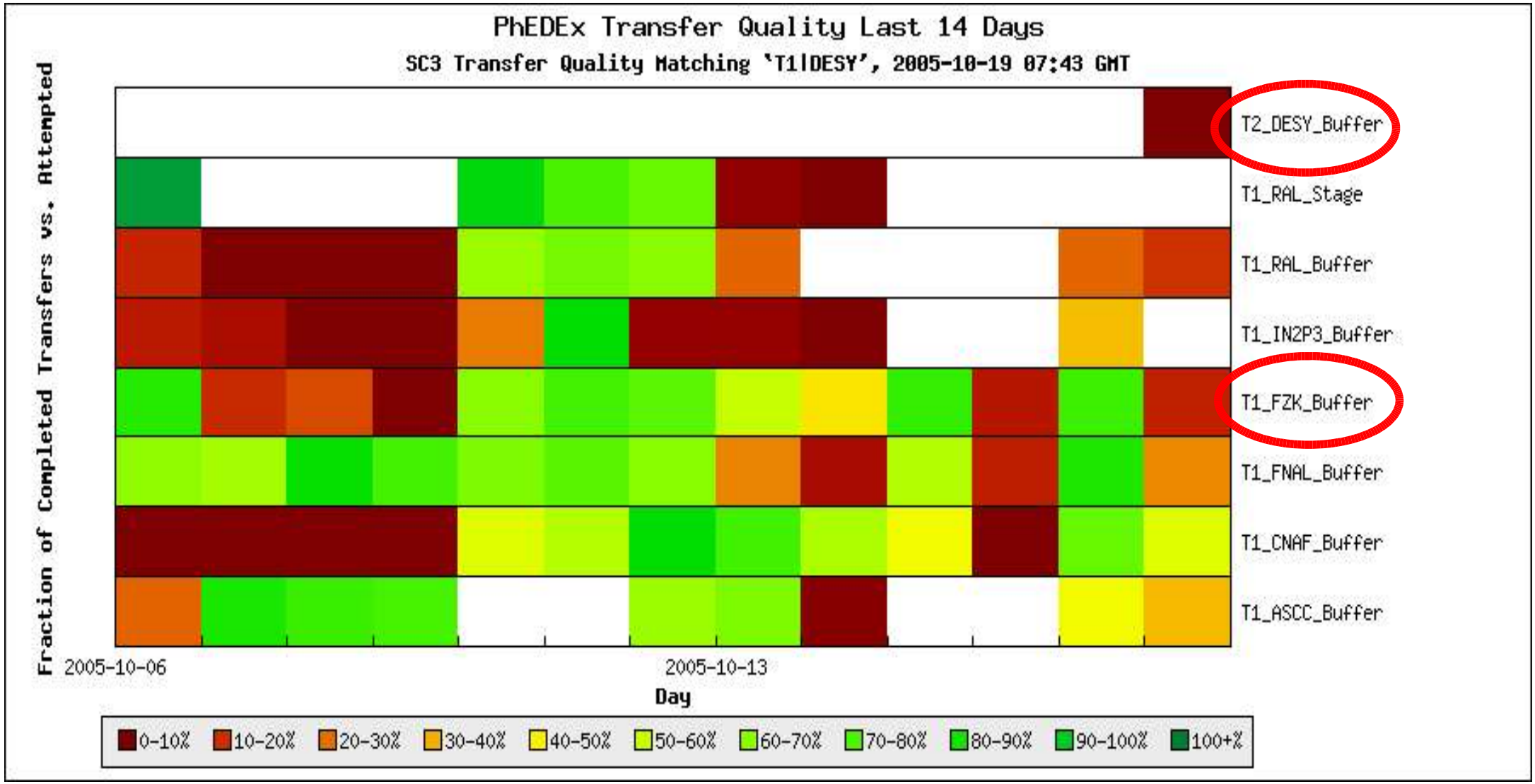
SC3: PhEDEx transfer quality

Completed transfers in the last 14 days:



SC3: PhEDEx transfer quality

Completed transfers / Attempted transfers in the last 14 days:



Non SC3 CMS activities at GridKa

Support for PTDR analyses has highest priority until April 2006!

- Grid based user analyses
  - at FZK for a long time
  - at DESY for a short time
- "Custodial responsibility" of GridKa:
  (CMS computing model)

  dCache based servicing of ~ 10 – 20 TB data volume
- Presently working on GridKa (T1) ↔ FNAL (T1) data transfers
- CMS Monte Carlo production:
  running again very efficiently since ~July

CMS user statistic at GridKa

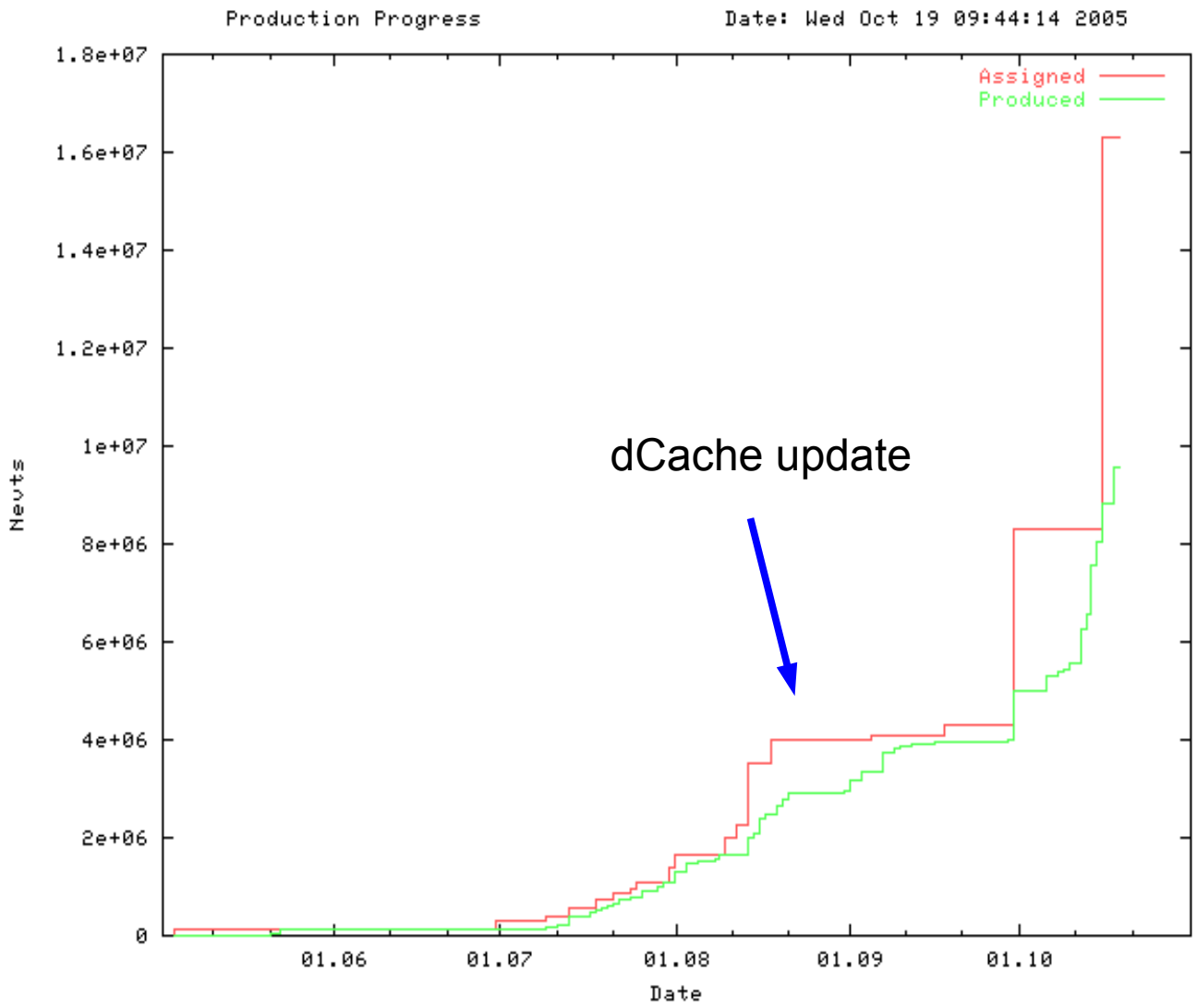
Example for user statistic at GridKa (grid based and local):

2005-10-09 - 2005-10-15
USER STATISTIC:

user	group	#jobs	WaitTime [hours]			
			sum	min	ave	max
cms001	cms	26	12.09	0.00	0.46	10.09
cms002	cms	55	1.70	0.00	0.03	0.22
cms004	cms	10	2.56	0.00	0.26	2.41
cms005	cms	960	266.65	0.00	0.28	3.07
cms007	cms	1	0.03	0.03	0.03	0.03
cms008	cms	260	486.95	0.00	1.87	47.62
cms009	cms	30	8.86	0.01	0.30	7.41
cmsadmin	cms	2417	4894.61	0.00	2.03	28.20
cmssgm	cms	5	0.09	0.01	0.02	0.03
jfernan	cms	95	10.72	0.00	0.11	1.41
schmidt	cms	418	1144.93	0.06	2.74	6.14
zhukov	cms	72	9.33	0.01	0.13	0.52

CMS Monte Carlo production at GridKa

- LCG based MC production at a huge scale:
  ~ 100 M events worldwide
- At GridKa:
  MC production (local)
  of ~ 5 – 10 % of all events
  - ~ 8 M events produced in less than 4 months
  - 9 M events assigned yesterday
  - 4 – 5 M events produced in first part of SC3
  - local MC production changes to LCG based MC production in SC4
- MC production partially at T2's in SC4



Summary

- Data transfers GridKa (T1) ↔ DESY (T2) established with max. 100 MB/s
- Transfers CERN (T0) to GridKa (T1) dCache/disk done (~ 100 MB/s)
  Transfers CERN (T0) to GridKa (T1) dCache/tape ready in SC3 service phase
- CMS achieved no sustained stable transfer rate of good quality due to technical problems, e.g. with CASTOR
- SC3 service phase running
- Non SC3 activities at GridKa:
  - Working on data transfers GridKa (T1) ↔ FNAL (T1)
  - Users are running grid based analyses
  - MC production restarted at high speed since ~July
- Next step: SC4

Backup slides

PhEDEx transfer rate

SC3 Transfer Rate: PhEDEx Status - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://cms-project-phedex.web.cern.ch/cms-project-phedex/cgi-bin/browser?db=sc;page=rates

PhEDEx SC3 Status

Transfer Rate

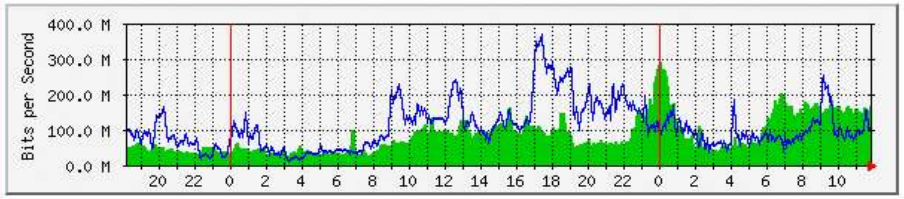
2005-10-18 09:52:06 GMT

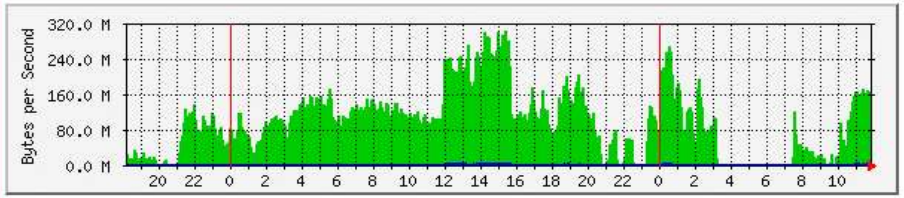
Database: Production | SC3 | Dev | Testbed

Monitor Options

- Component Status
- Transfer State
- Transfer State Details
- Replica State
- Subscriptions
- File Size Stats
- Transfer Rate
- Transfer Rate Plots
- Transfer Queue Plots
- Transfer Quality Plots
- Agent Status
- Daily Reports
- Daily Report

CERN Daily External Internet Traffic

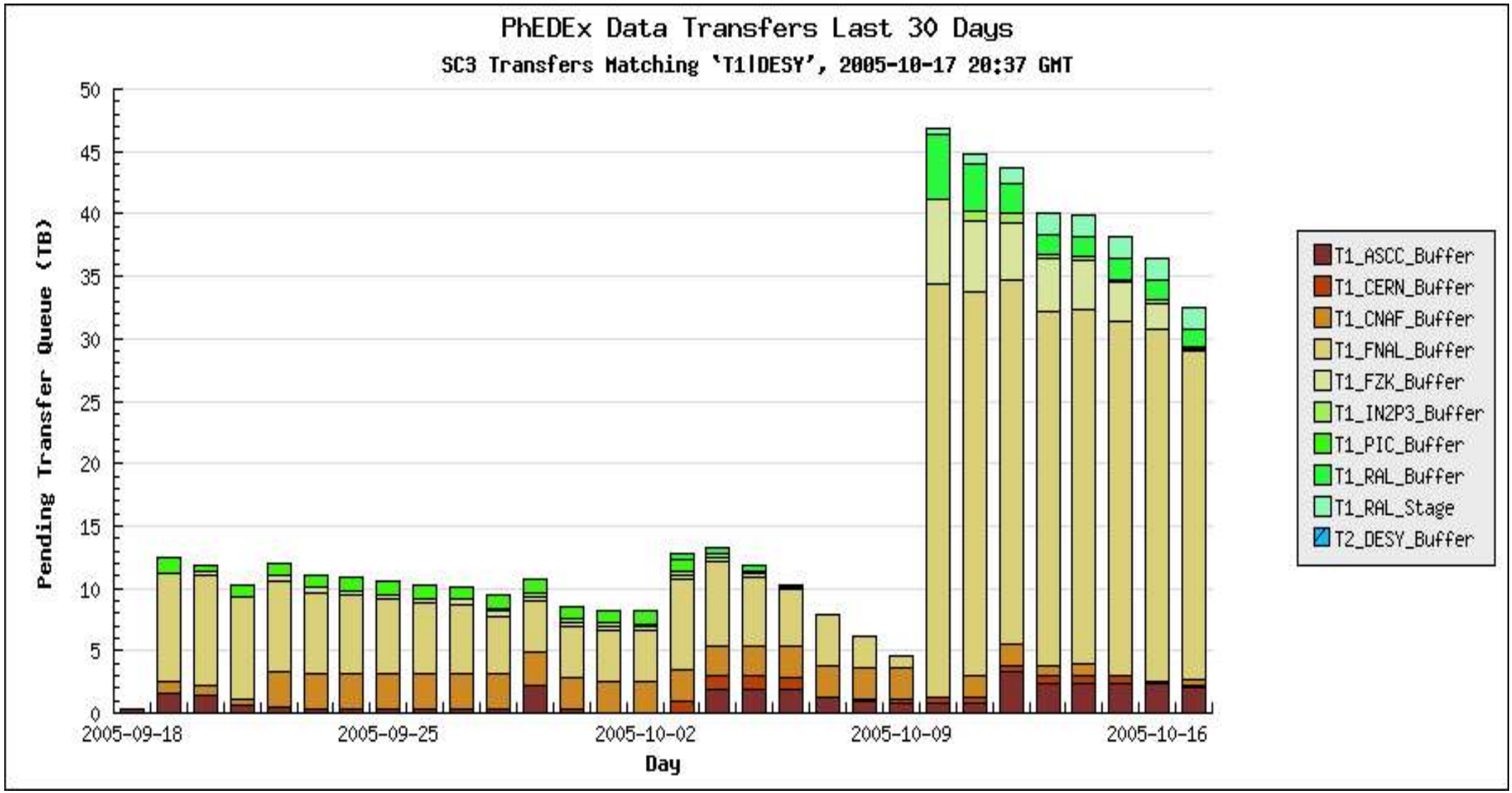




Last hour

Age	From	To	Files	Total Size	Aggregate Rate
0h00	T1_CERN_Buffer	T1_CNAF_Buffer	104	193.3 GB	55.0 MB/s
0h00	T1_CERN_MSS	T1_CERN_Buffer	49	66.2 GB	18.8 MB/s
0h00	T1_FNAL_Buffer	T2_Nebraska_Buffer	14	27.3 GB	7.8 MB/s
0h00	T1_FNAL_Buffer	T1_FNAL_MSS	4	7.0 GB	2.0 MB/s
0h00	Total		171	293.9 GB	83.6 MB/s

Last day

Age	From	To	Files	Total Size	Aggregate Rate
0h00	T1_CERN_Buffer	T1_CNAF_Buffer	1225	2.1 TB	25.4 MB/s
0h00	T1_CERN_Buffer	T1_ASCC_Buffer	520	941.4 GB	11.2 MB/s
0h00	T1_FNAL_Buffer	T1_FNAL_MSS	476	908.9 GB	10.8 MB/s
0h00	T1_CERN_Buffer	T1_FNAL_Buffer	460	877.8 GB	10.4 MB/s
0h00	T1_CERN_MSS	T1_CERN_Buffer	490	868.7 GB	10.3 MB/s
0h00	T1_CERN_Buffer	T1_FZK_Buffer	291	559.6 GB	6.6 MB/s

SC3 PhEDEx data transfers