



CMS

Compact Muon Solenoid

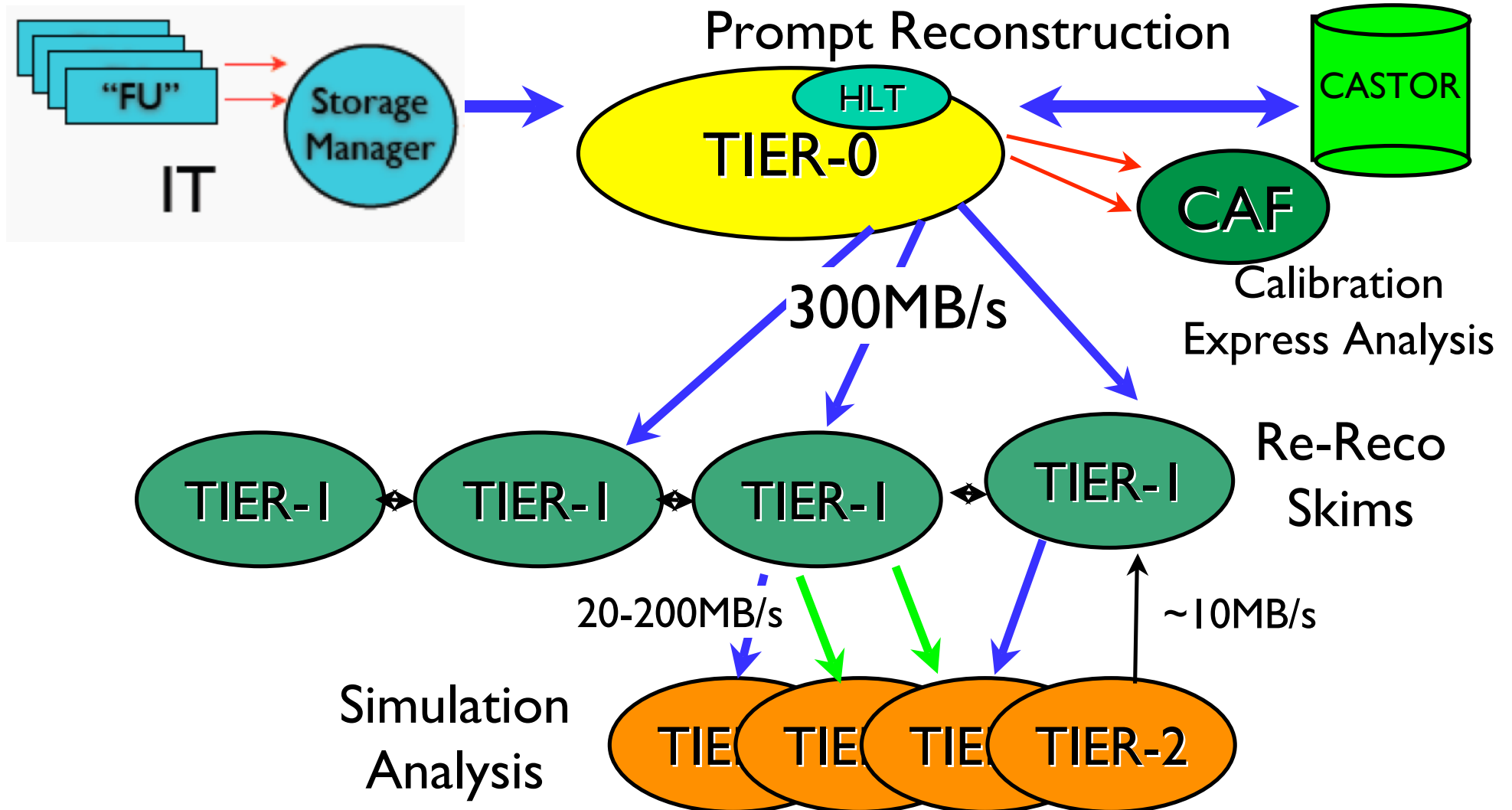
First CSA07 experiences

Matthias Kasemann CERN/DESY
(on behalf of I.Fisk & N.Geddes)

GDB meeting 071010



CSA07 Workflows





CSA07

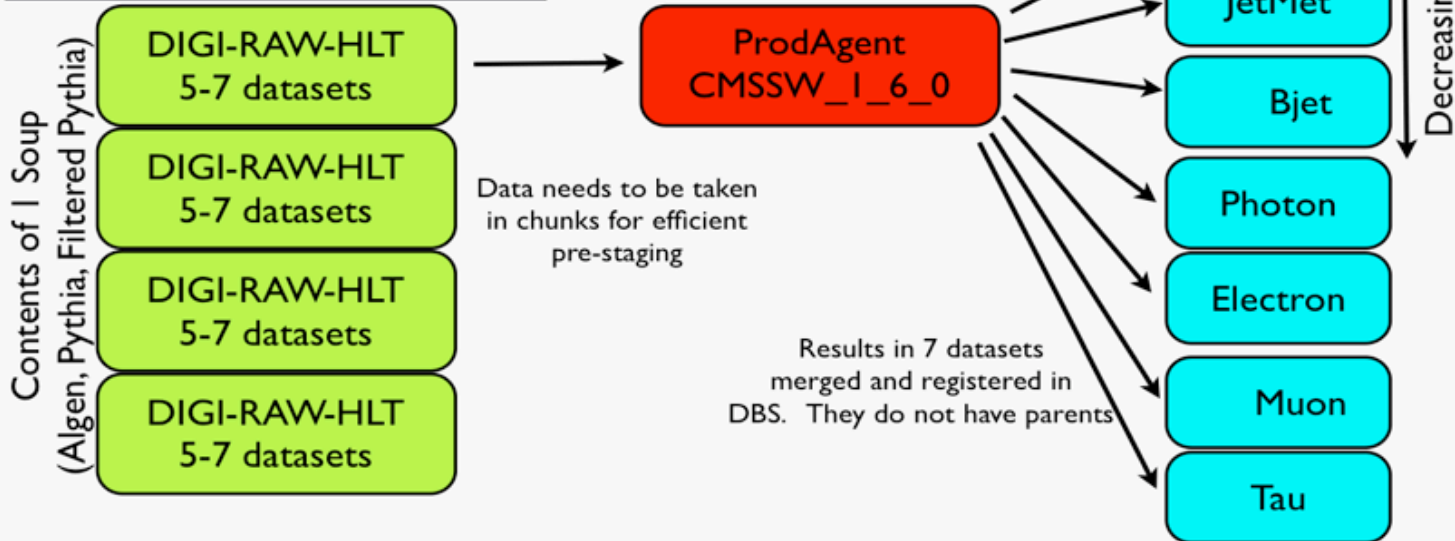
- Computing, Software and Analysis (CSA) challenge (Sep-Oct 2007)
 - **Mimic in detail what is needed in 2008 (at a scale of 50%)**
- 200 MeV mis-calibrated/mis-aligned (for 10-100pb⁻¹)
- Making (at the T0) and distributing (to all T1 centers) the AOD data.
 - **Placement of data in the Tier-1 centers**
- Running of skims at Tier-1 centers
- Re-reconstruction at the T1 centers
- Re-making of the full AOD samples after a re-reconstruction step.
- Copying of the skimmed datasets to Tier-2 centers and execution of analysis exercises at these centers.
 - **Migrate the bulk of analysis activities to Tier-2 centers.**
- Parallel (with the processing of the CSA07 data at the Tier-0) Monte Carlo production of signal events at the Tier-2 centers
- CAF (CERN Analysis Facility): first test of “Express Line” for fast turnaround of a few analyses.



CSA07 Primary Datasets & Allocations

- Splitting Workflow:

Splitting Workflow (CERN Based)



Primary Dataset	CSA size (TB)	Tier 1 Streams						
		ASGC	IN2P3	CNAF	FZK	PIC	RAL	FNAL
Tau	3.4		1					1
Photon	14.6		1				1	1
Muon	10			1		1		1
Electron	14.5		1	1				1
Bjet	27.5				1			1
JetMET	31.2	1						1
AllEvents	83.4		0.2	0.2	0.2		0.2	0.2
Total data volume (inc. copies)	346	31	58	51	53	10	40	127



Tier-0 Workflow

- A number of workflows are being attempted at production scale for the first time in CSA07
 - Production use of the Higher Level Trigger (HLT) with prompt reconstruction
 - Dividing the data into primary dataset streams
- Starting sample of events was roughly 150TB of data
 - HLT events (including all the debugging information) are larger than anticipated
 - Choice of primary datasets is very different from what was calculated in the computing model
 - Much fewer streams
 - Larger overlap between streams
 - For calibration there was a request for an AllEvents stream, which will not exist in the real data
 - Output data will be over 300TB from CERN



T0 reconstruction highlights

(M.Miller/DataOperations)

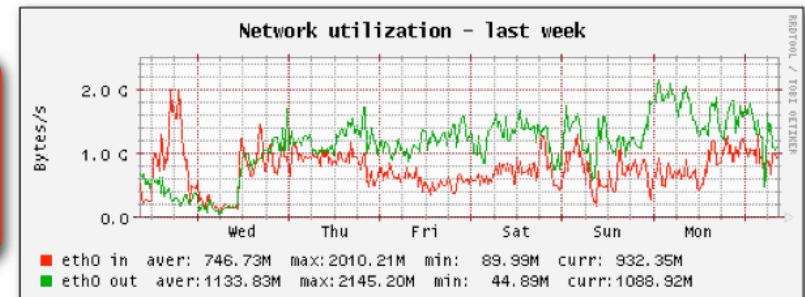
Take home messages

- CASTOR and LSF have performed smashingly at high rate
 - superb support from cern IT
- We can sustain I/O rates up to x4 beyond spec of 500 MB/s
- We can export data from cern at x2 csa07 goal

Executive Summary

- Smooth operations since last October 3
 - Not running at full speed, no new workflows to commission at T0
 - over 100k jobs run
 - over 200 TB data created
 - Should finish Alpgen soup by Wed
 - Dynamic: 4 simultaneous meta-workflows
 - Split => Reco => AOD, express skim

7 day average
Write: 750 MB/s
Read: 1130 MB/s

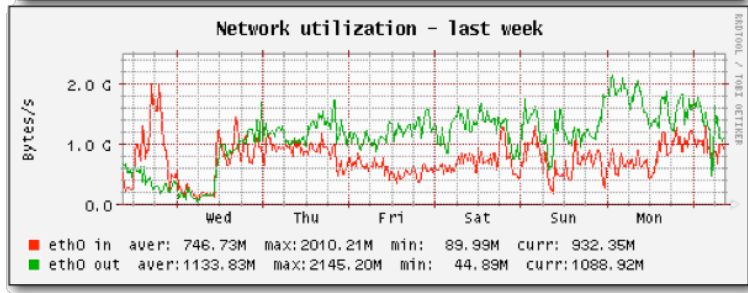
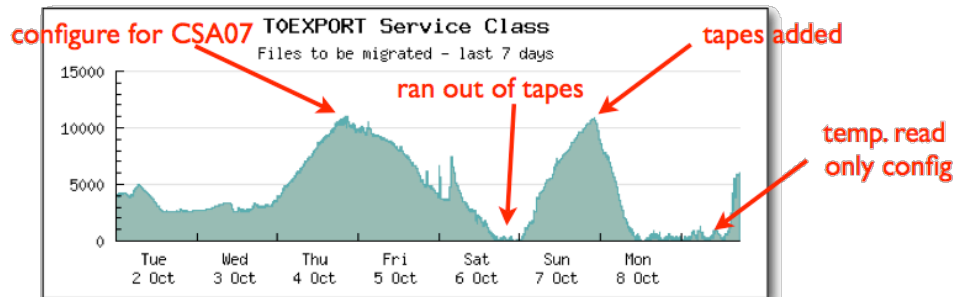




T0 reconstruction highlights

(M.Miller/DataOperations)

Can we write this to tape?



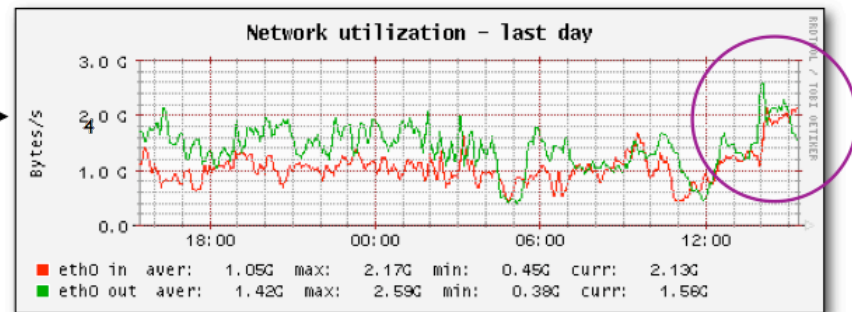
- Can we hit (sustain?) t0export limit of 2 GB/sec read and write to disk?

=> Yes, but must verify with 4 experiment load



Michael Miller

Data Operations, October 9, 2007



Sustained aggregate 4 GB/sec for more than 1 hour





Early Lessons

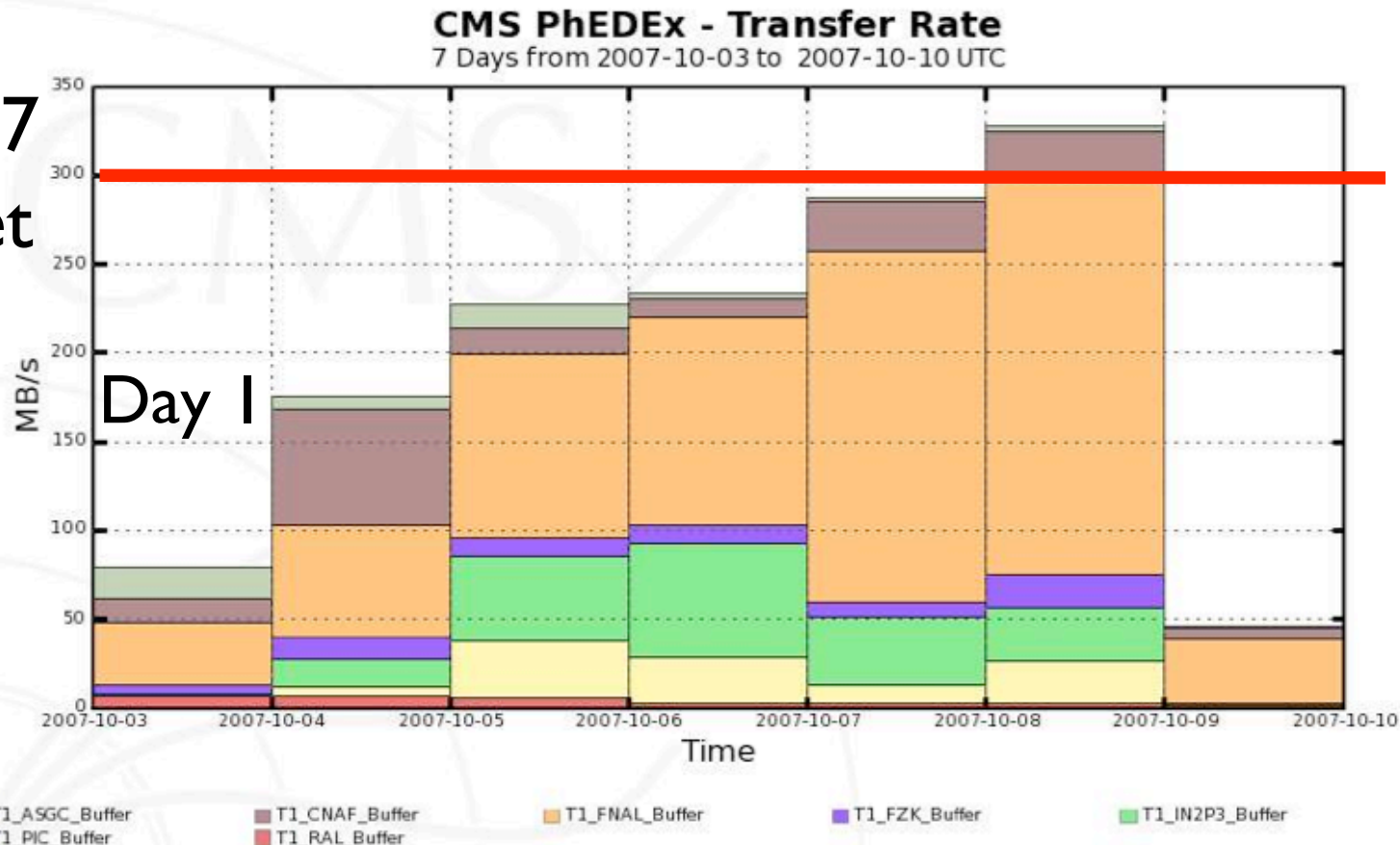
- Working with the HLT in the application has exposed some issues in the data model chosen for storing the HLT information
 - The memory footprint of the application grew beyond the constraints of the computing model and the batch systems
 - Speed issues with registering the branches into the data bookkeeping service
 - Data Model is under review, and the application and service issues have been worked around for the challenge.
 - Dropping information to speed the bookkeeping and dividing the workflow to keep the memory within the physical limits
- Need to make the primary datasets with the HLT from a “soup” of input datasets exposed a number of issues in the workflow
 - Tremendous load on Castor: The pre-challenge preparations will write more than 1PB of data into mass storage
 - Except for issues on specific Castor servers and CMS using all its media the Castor facility has performed remarkably well



Transfers from CERN

- Average transfer rate from CERN to Tier-1 centers
 - The presence of the primary dataset splitting workflow has prevented the reconstruction rate from reaching the challenge target. Not all sites have enough data subscribed to hit the challenge targets currently

CSA07
Target

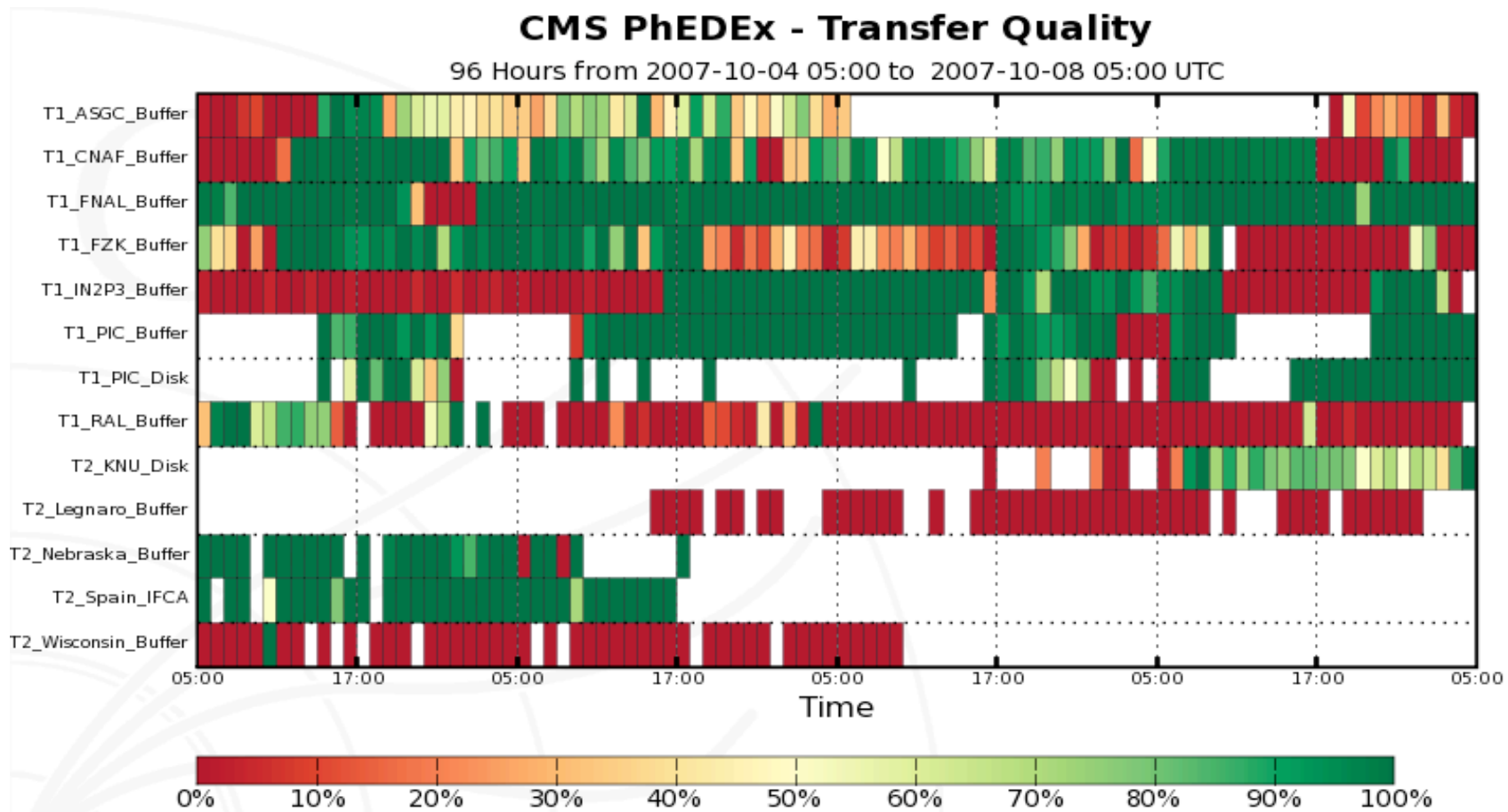


Maximum: 328.15 MB/s, Minimum: 45.98 MB/s, Average: 196.69 MB/s, Current: 45.98 MB/s



Transfer Quality

- Transfer quality from CERN, room for improvement
 - In CMS transfer quality is defined as the number of times a transfer is attempted before it succeeds. Nearly all transfers eventually succeed.





Data Transfer highlights

(Nicolò Magini/DDT)



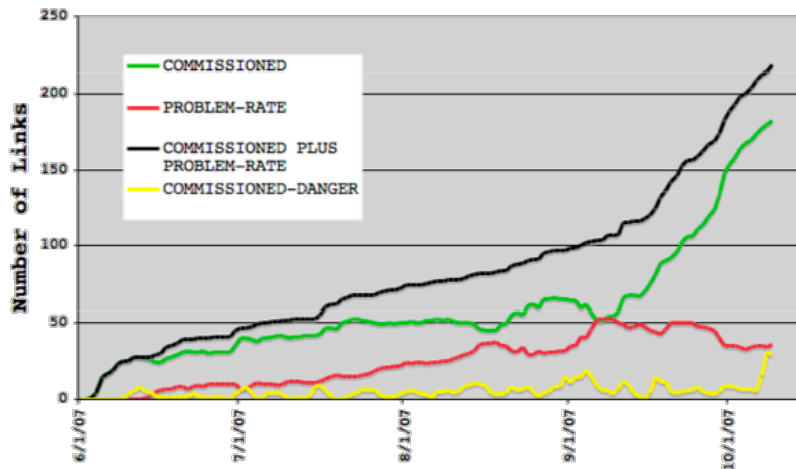
Summary of Commissioned Links

- More than 180 COMMISSIONED LINKS
- But: 23 links in DANGER yesterday!



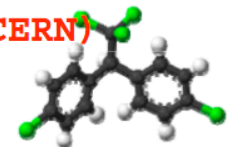
T1 Cross-Link Commissioning

COMMISSIONED LINKS



	ASGC	CERN	CNAF	FNAL	FZK	IN2P3	PIC	RAL
ASGC		Overide						
CERN	Overide							
CNAF								
FNAL								
FZK								
IN2P3								
PIC								
RAL								

- Almost all links exercised now
- All CERN ↔ T1 links COMMISSIONED
- Of the 42 T1-T1 links (not CERN):
 - 17 are COMMISSIONED
 - 8 more on the way to COMMISSIONING with good rates
- 4/7 T1s can import from another T1 (+CERN)
 - CNAF, RAL, IN2P3 only from CERN





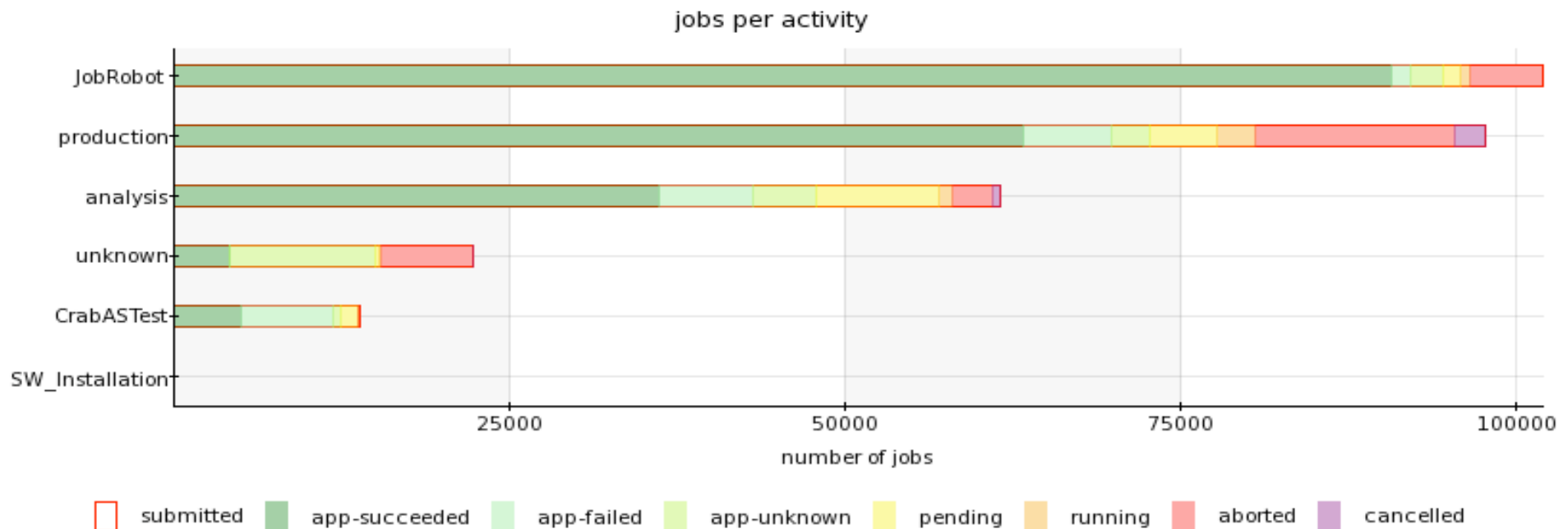
Tier-1 Tape

- An early concern was the ability for the Tier-1 sites to successfully archive the incoming data to mass storage
 - So far 7 of 7 Tier-1 sites are successfully archiving to tape
 - A backlog of data to migrate exists at CNAF and to a lesser extent IN2P3. Other Tier-1 sites are largely keeping current with the incoming data
- The data rate will increase as the reconstruction rate ramps up
 - Need to continue to watch this and monitor the stability over the month
 - Already promising that more sites are participating with tape libraries than CSA06



Job Submissions

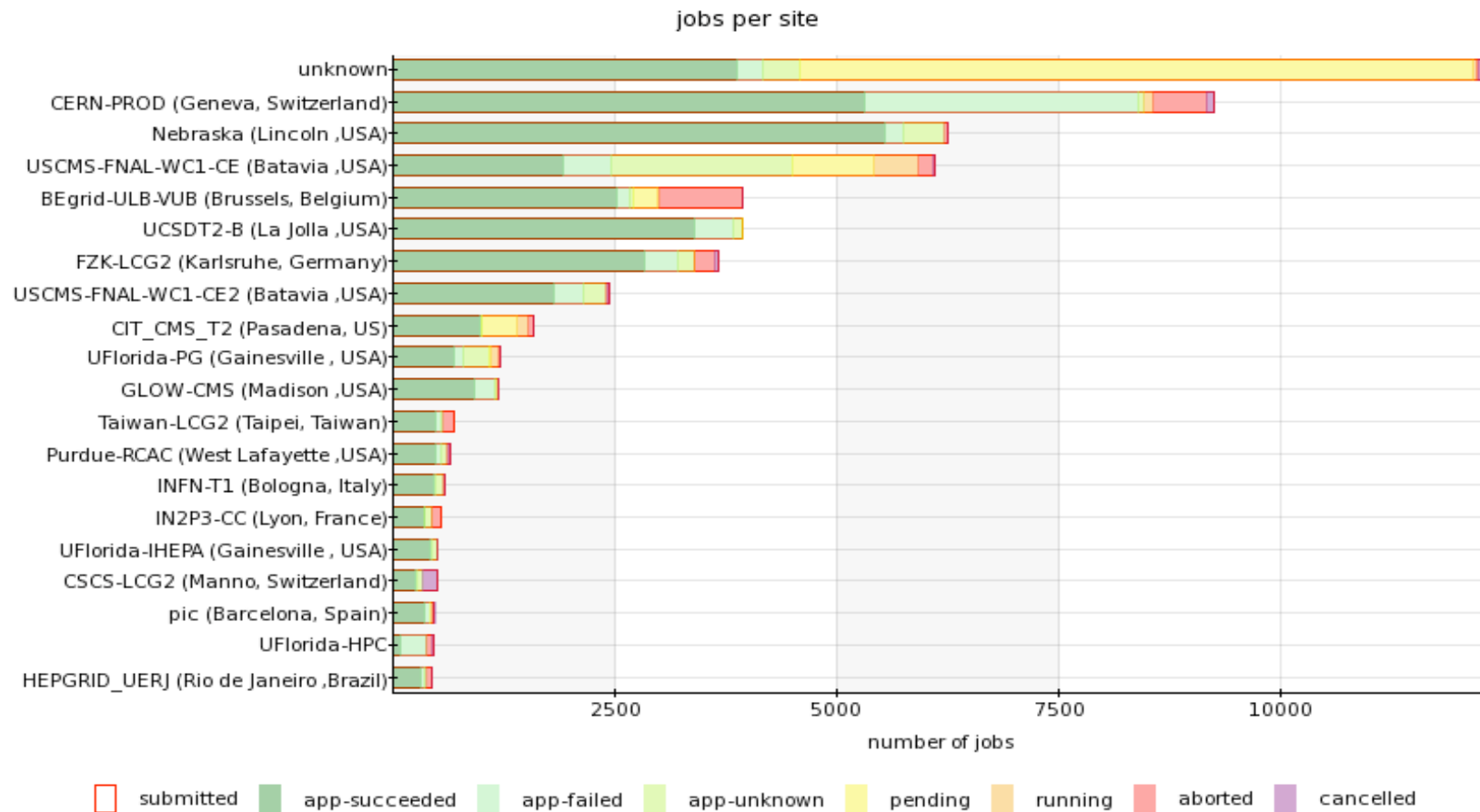
- Organized processing jobs for simulation at Tier-2 centers, reconstruction at the Tier-1 centers and soon data skimming at Tier-1s are keeping a high rate of production jobs
 - During the first 5 days of the challenge
 - Averaged nearly 20k jobs per day (close to the challenge metric of 25k)
 - With Job Robot, Production and Analysis we are exceeding the rate achieved at the final rate of CSA06





Analysis

- The analysis rate is much lower than the target metric in the challenge
 - Hopefully will increase throughout October, when new CSA07 data arrive
 - Use of Tier-2s for analysis is increasing





Next Steps

- Need to increase the reconstruction rate at CERN this week
 - The splitting workflow is very IO intensive and we hoped to have finished it before the challenge began, (it will finish today)
- Start skimming data at Tier-1 centers (started yesterday evening)
 - The output will be transferred to Tier-2 centers for analysis
 - Will bring Tier-1 to Tier-2 transfers into the challenge
 - Hopefully more analysis access as well
- Next week we need to begin re-reconstruction at Tier-1 centers
 - Will bring Tier-1 to Tier-1 transfers into the challenge
 - Will also exercise reading access of the Tier-1 mass storage systems

1) Detector Installation, Commissioning & Operation

2) Preparation of Software, Computing & Physics Analysis

Tracker Inserted

Test Magnet at low current

Last Heavy Element Lowered

Tracker cabled

CMS Cosmic Run CCR_0T
(defined periods Dec-Mar)
(Several short periods Dec-Mar)

Beam-pipe Closed and Baked-out

1 EE endcap Installed, Pixels installed

Cosmic Run CCR_4T

Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May

S/w Release 1_6 (CSA07)

CSA07
S/w Release 1_7 (CCR_0T, HLT Validation)

2007 Physics Analyses Completed

S/w Release 1_8 (Lessons of '07)

CCRC08 functional tests (in series)
S/w Release 2_0
(CCR_4T, Production of startup MC samples)
MC Production for Startup

CCRC08 = CSA08_[CMS]