CMS

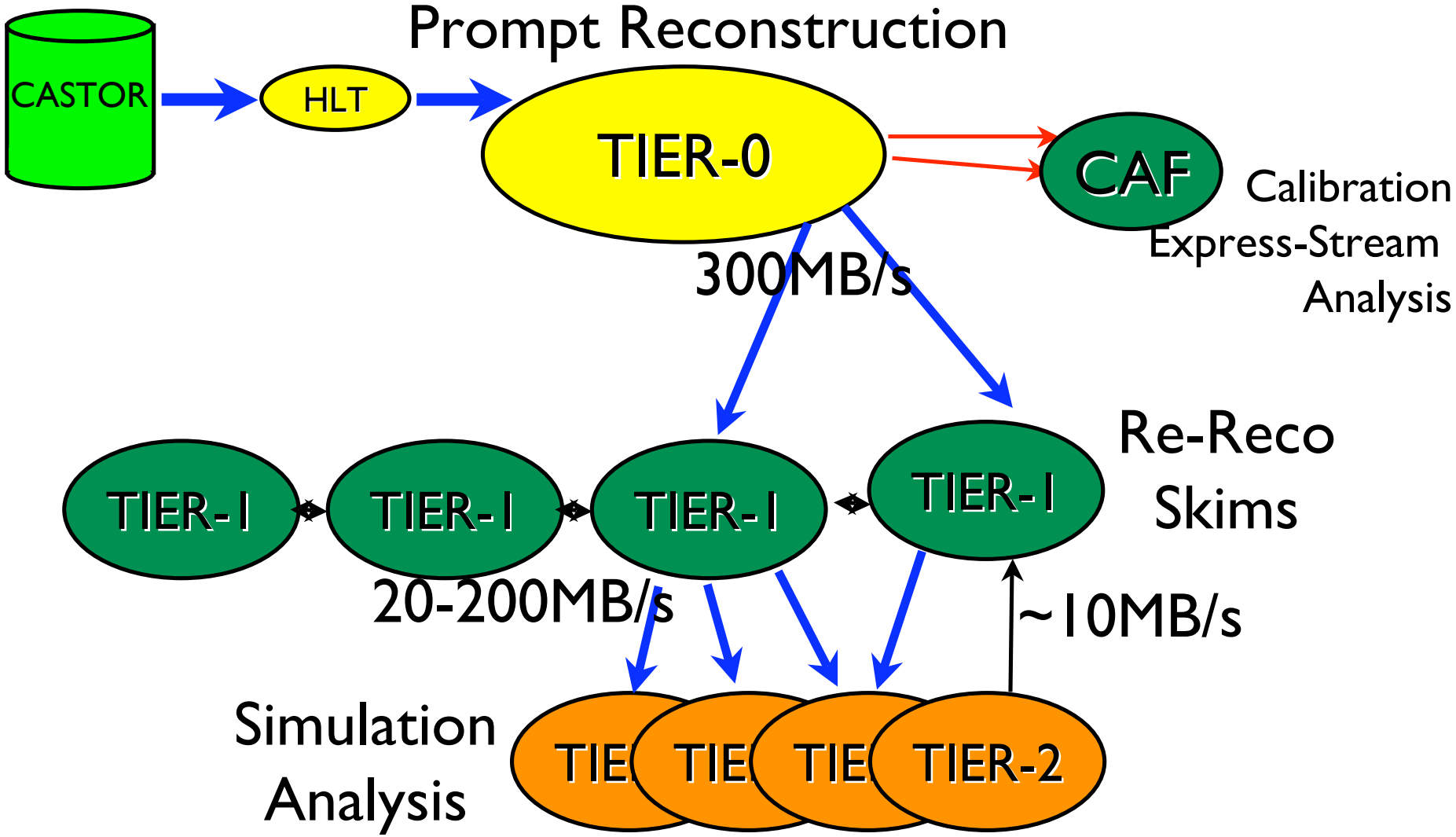
Computing Report

- CSA07 performance & summary
- PADA Taskforce
- CCRC08-February tests

Matthias Kasemann

WLCG MB 26.2.08

CSA07 Workflows

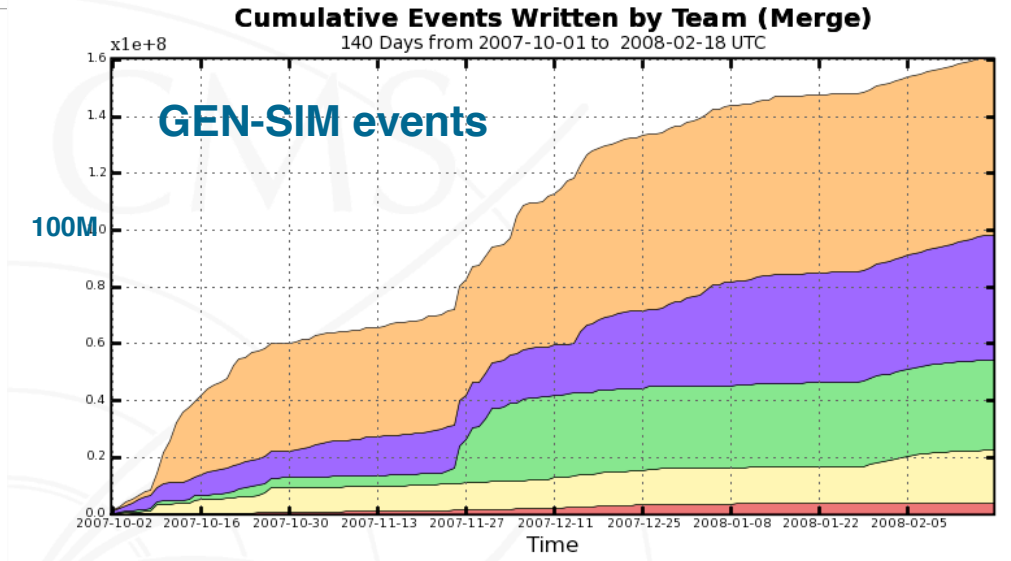


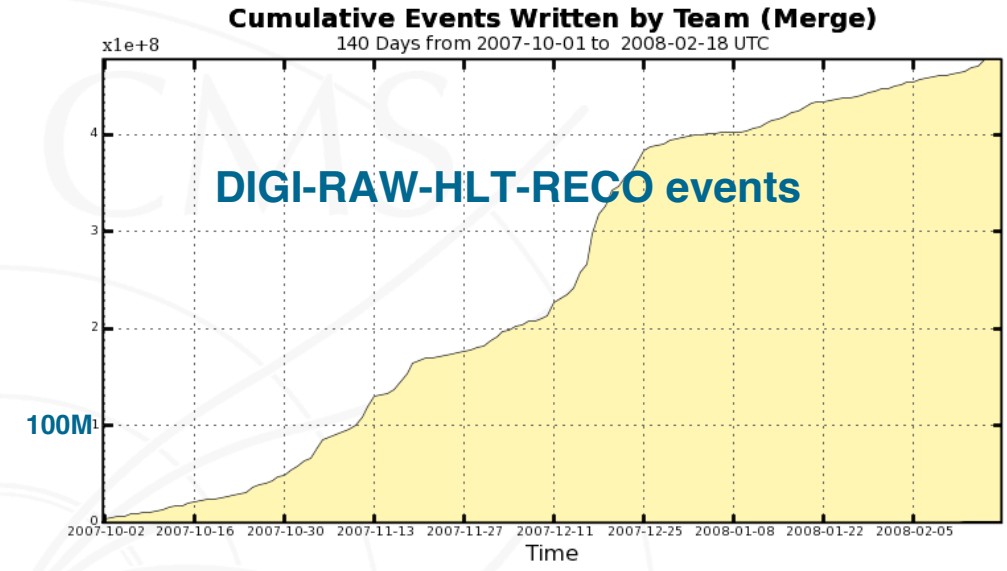
Production Summary 10/'07-02/'08

- **160M Monte Carlo events produced since October 07**
  - On request of Physics, DPG and HLT groups

- **Total CSA07 event counts:**
  - **80M GEN-SIM**
  - **80M DIGI-RAW**
  - **80M HLT**
  - **330M RECO**
  - **250M AOD**
  - **100M skims (mixed RECO/AOD)**
  - **----------**
  - **920M events**

- **Events were processed + reconstructed in several steps, several times**





DataOps coordinated by C.Paus, L.Bauerdick

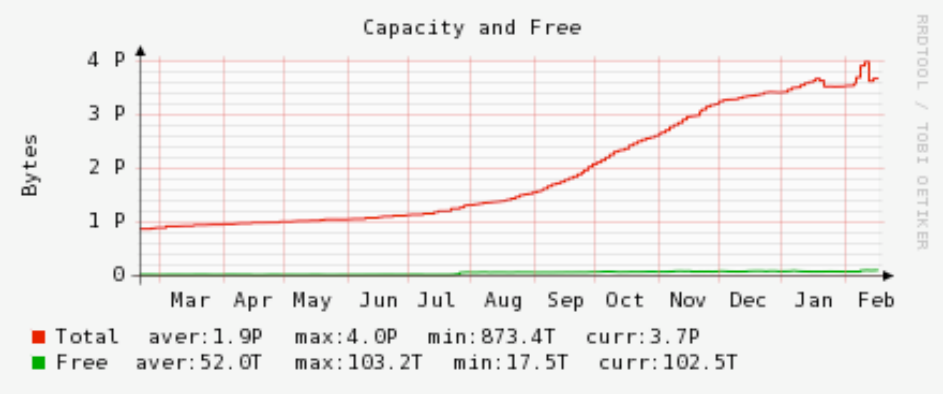
CSA07 event samples

- CSA07 soups: 250M RECO, 250M AOD, 100M skims
  - Three calibrations applied: 10/pb^{-1},100/pb^{-1} 0/pb^{-1}
  - Events produced: RECO, EXPRESS, AOD, skim, ALCARECO

- The CSA07 signal samples really evolved over time. We started from 50M and went up to 85M by now (not a real problem)

- Data volume of CSA07 samples right now: 1.9 PB (without counting repetitions)

- Delivery of the samples is mostly done with small remainders pending.



CMS data in CASTOR@CERN: 3.7PB

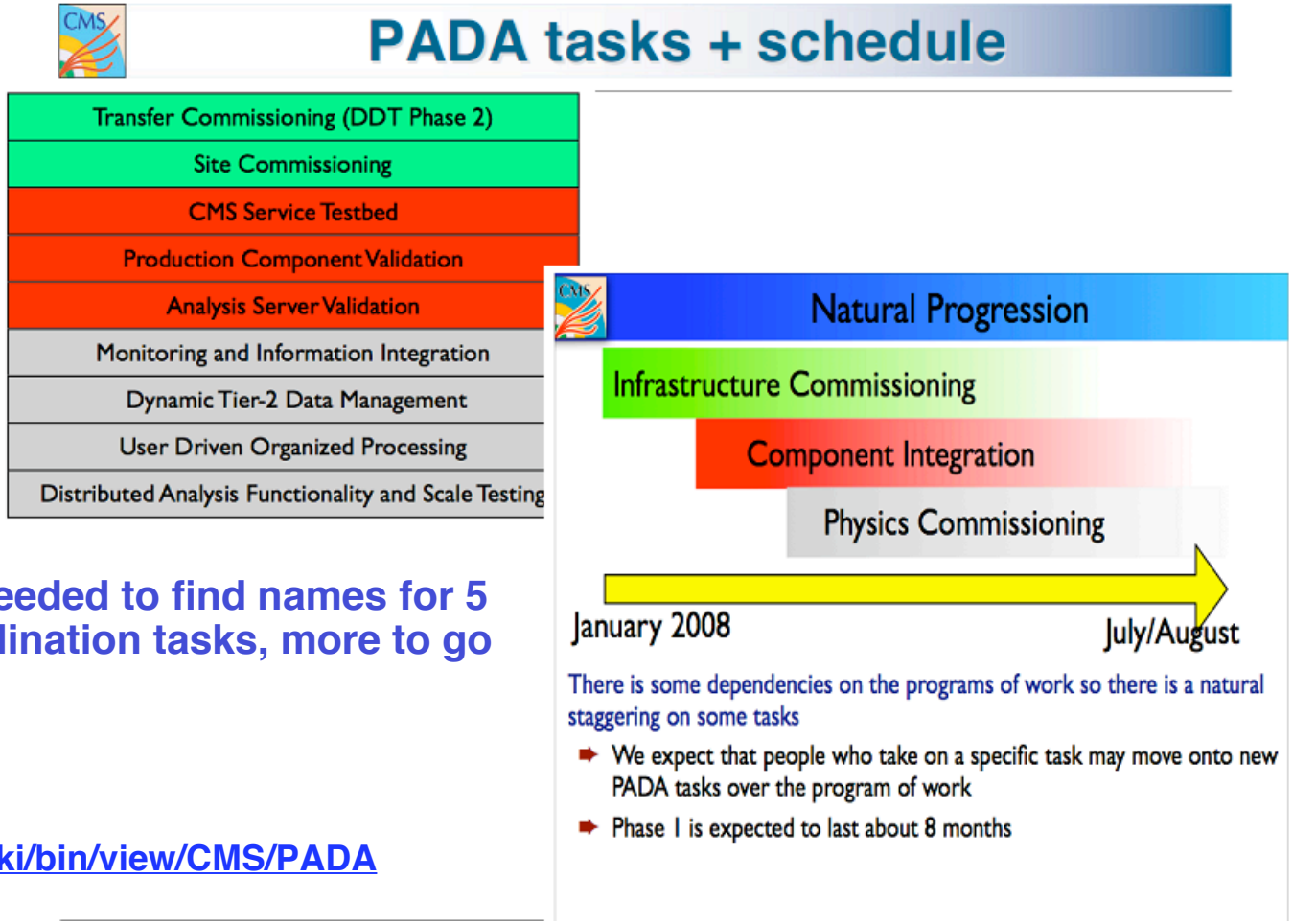
CSA07 Analysis Summary

- **There is a full list of lessons on the Twiki for Offline and Computing**
  - twiki.cern.ch/twiki/bin/view/CMS/CSA07
- **In CSA07 a lot was learned and a lot was achieved…**
- **The production infrastructure is in full operations**
- **CSA07 analysis identified tasks to be addressed**
  - **Two strategies derived for Computing:**
    - A new Task Force:
      Integrating development, deployment and commissioning
      Processing **A**nd **D**ata **A**ccess (**PADA**)
      - coordinated by I.Fisk and J.Hernandez
    - Testing the computing infrastructure in **CCRC08/CSA08** in February and prepare scope for May '08

This is the focus for Computing during this CMSweek

Processing and Data Access: PADA

- **The Processing and Data Access Task Force is a series of tasks and programs of work**
  - designed to bring the Computing Program into stable and scalable operations.



Succeeded to find names for 5 coordination tasks, more to go

See twiki:
https://twiki.cern.ch/twiki/bin/view/CMS/PADA

PADA Activities

- **Distributed production commissioning (Jose Hernandez)**
  - Integration, commissioning and scale testing of the organized production workflows at Tier-1 (reprocessing and skimming) and Tier-2 (MC production) sites.
  - Improve the level of automation, reliability, efficiency of resource use and scale of the production system, reducing at the same time the number of operators required to run the system.
  - Commission new components of the production system.
  - Perform functionality, reliability and and scale tests.

- **Monitoring activities (Stefano Belforte, Artem Trunov)**
  - Integration of monitoring tools,
  - gather needs and input from users,
  - provide feedback to developers, testing/evaluation,
  - help in defining user/site monitoring views.

PADA Activities

- **Site commissioning (Francisco Matorras, Stijn de Weirdt)**
  - Demonstrate that CMS can access the resources that are pledged to CMS.
  - Test scalability of CEs and storage for CMS-style workflows.
  - Site commissioning is a step before demonstrating that the CMS workflow tools can be scaled.
  - Verify that the workflows don't interfere,
  - Verify that analysis and productions jobs are shared on Tier-2s
  - Find the stable operating points of skimming and reconstruction for the Tier-1 sites.

- **Analysis activities (Coordinated by Alessandra Fanfani)**
  - User feedback: Collect inputs from the user community and provide feedback to developers.
  - Organize integration and Testing of new functionalities of the analysis tools.
  - Deployment of CRAB server.

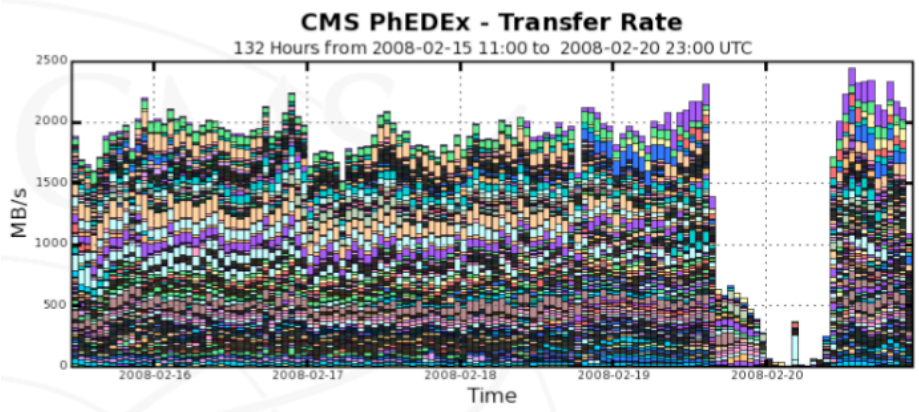
PADA Activities

- **Data transfer commissioning (DDT) (James Letts, Nicolo Magini)**
  - Demonstrate that the Tier-1 and Tier-2 sites are capable of utilizing the networking as specified in the Computing TDR.
  - Demonstrate that data management tools, networking and storage configuration at sites are adequate for data transfers at the required scale.
  - Perform link commissioning + testing following new DDT metrics.

Status:

- New DDT metric (incl. regular exercising) in place since February 11
- 311 commissioned links:
  - 52/56 T[01]-T1
  - 162/362 T1-T1
  - 90/352 T2-T1

- **During CCRC'08, total throughput in Debug is 2x what it was in CSA07, almost 20Gbps.**



This week Feb 15-20

CMS CCRC'08 Schedule

- **Phase 1 - February 2008:**
  - **blocks of functional and performance tests**
    - Verify (not simultaneously) solutions to CSA07 issues and lessons
    - Attempt to reach '08 scale on individual tests at T0, T1 and T2
    - Cosmics run and MC production have priority if possible
    - Tests are independent from each other
    - Tests are done in parallel

- **Phase 2: - May 2008:**
  - **Full workflows at all centers executed simultaneously by all 4 LHC experiments**
  - **Duration of challenge: 1 week setup, 4 weeks challenge**

  - **CMS scope defined these days**

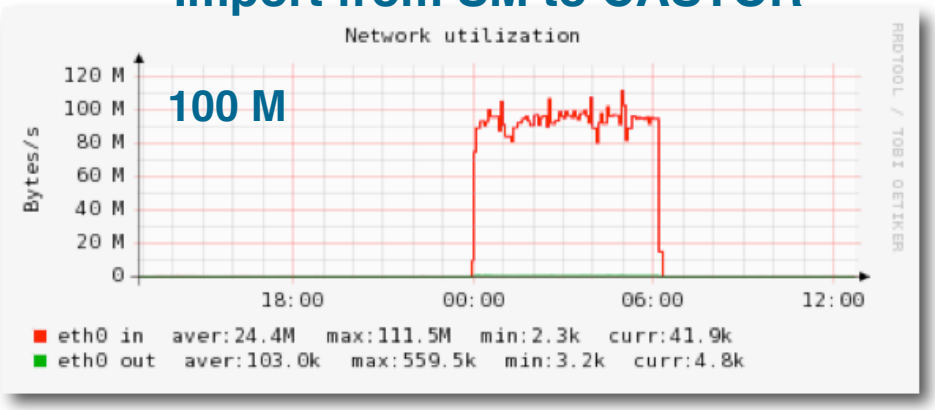
CCRC08 February tests

Data recording at CERN

1a) readout from P5, use HLT, w. stream definition, use Storage Manager, transfer to T0, perform repacking, write to CASTOR (D.Hufnagel)

- **Goal: verify dataflow for CMS, commission the new 10GB fiber**
  - 1 GB fiber used for Global runs since long
- **Status:**
  - 13.2.08: First successful transfer on new 10 Gb fibre at 100MB/s (limited by transfer node)
  - Next step:
    - integrate into transfer system
    - run in parallel to normal data transfers

Import from SM to CASTOR

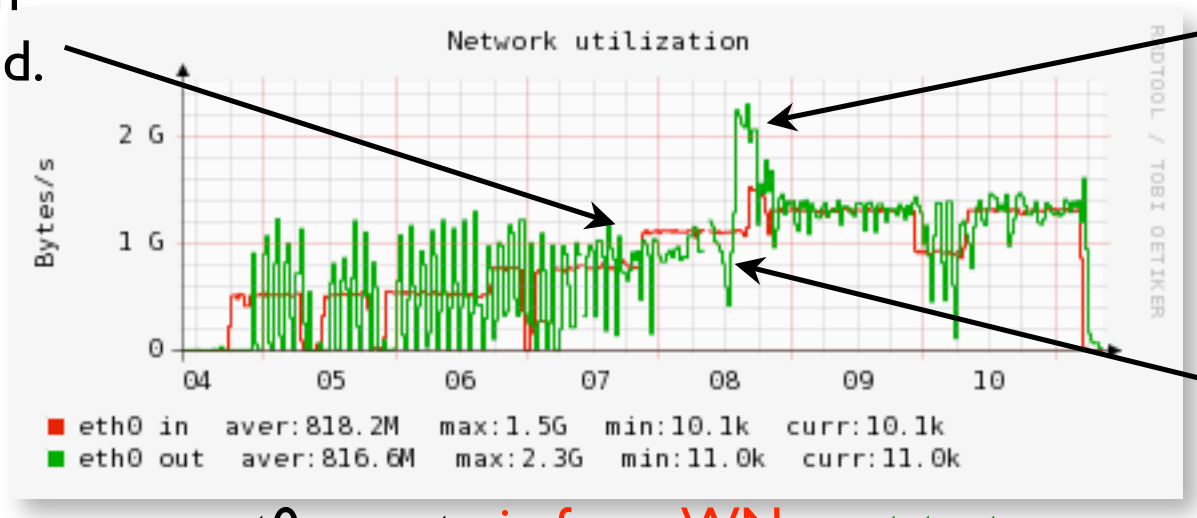


CCRC08 test: Data recording at CERN

1b) CASTOR data archiving test (M.Miller / DataOps team)

- **Goal: verify CASTOR performance at full CMS and ATLAS rate**
- **Status:**
  very successfully completed, reached rate of 1.5 GB/s
  - Good coordination with CERN-IT, quick response
  - Test at all-VO rate, other VO's didn't stress the system

Migration policy mod.

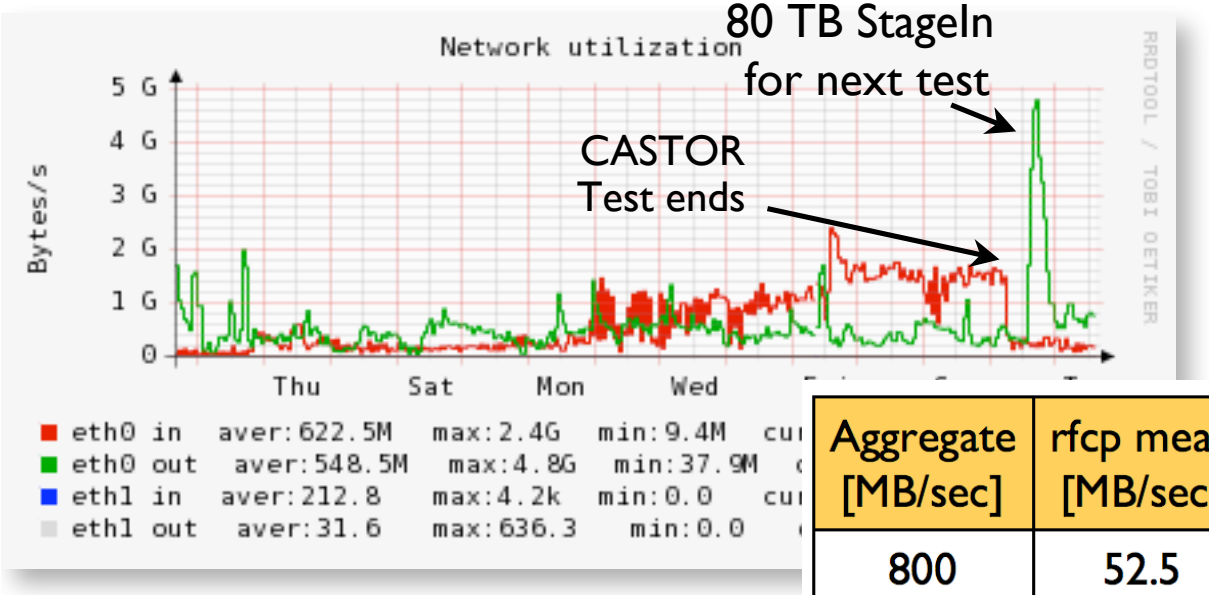


Recovered

Ran out of tapes

t0export: in from WN, out to tape

Last 2 weeks: integrated tape system usage



Aggregate [MB/sec]	rfcp mean [MB/sec]	rfcp sigma [MB/sec]	#jobs	t_{lumi} [sec]
800	52.5	17.4	~30	37
1100	40.5	15.4	~50	27
1300	31.7	13.0	~90	22
1500	18.1	11.5	100-infinity	18

Rates ultimately limited by 1Gbs on 13 t0input servers

CMS CASTOR TEST - Performance observed:
Averaged 633 MB/sec write (1.1 GB/sec during test)
Averaged 548 MB/sec read (~400 MB/sec during test)
Read Spike: regular stagein, 101 drives => 5 GB/sec

CCRC08 test: High Rate Processing at T0

Coordinated by M.Miller / DataOps

Goal:

- "high-rate" processing of cpu/RAM limited jobs
- Originally: measure interaction with other VO's on same WN
  BUT: CMS does not share WN with other VO's @ CERN (for now)

Setup:

- regular operations (physics requests)
- ReReco with 0pb^{-1} conditions of Stew and Gumbo

Status:

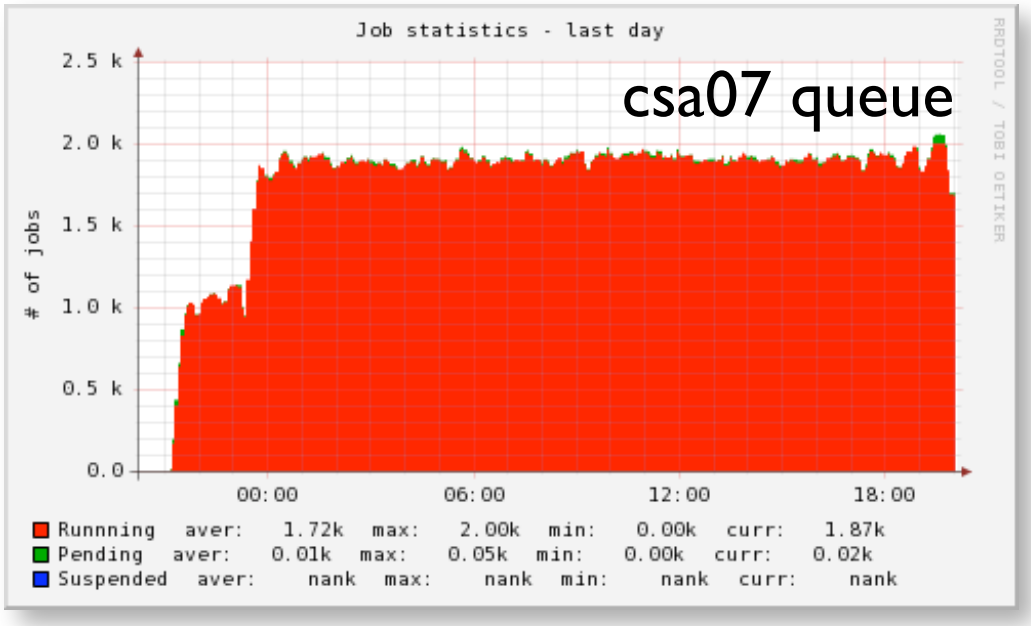
- started with 41k jobs of the 80 TB Stew AllEvents
- Finished in expected time
- Not much action from other VO's, no sign of WN problems
- Again turning into a CASTOR I/O test

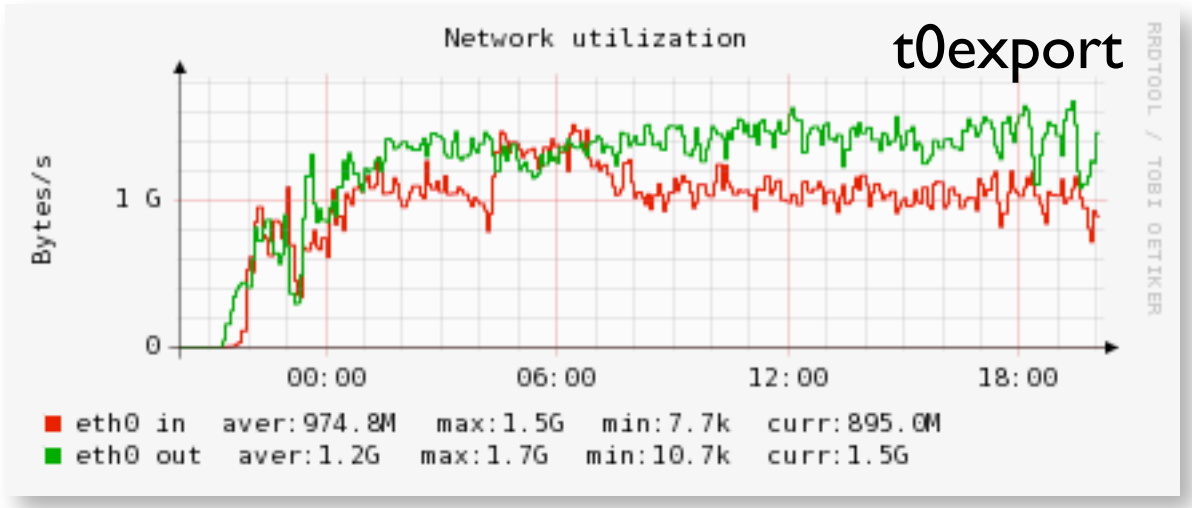
CCRC08 test: High Rate Processing at T0

Wednesday snapshot

Summary:

- processing runs routinely
- Small level of IO errors (2%), cured by retries
- Will test: copy files to local disk



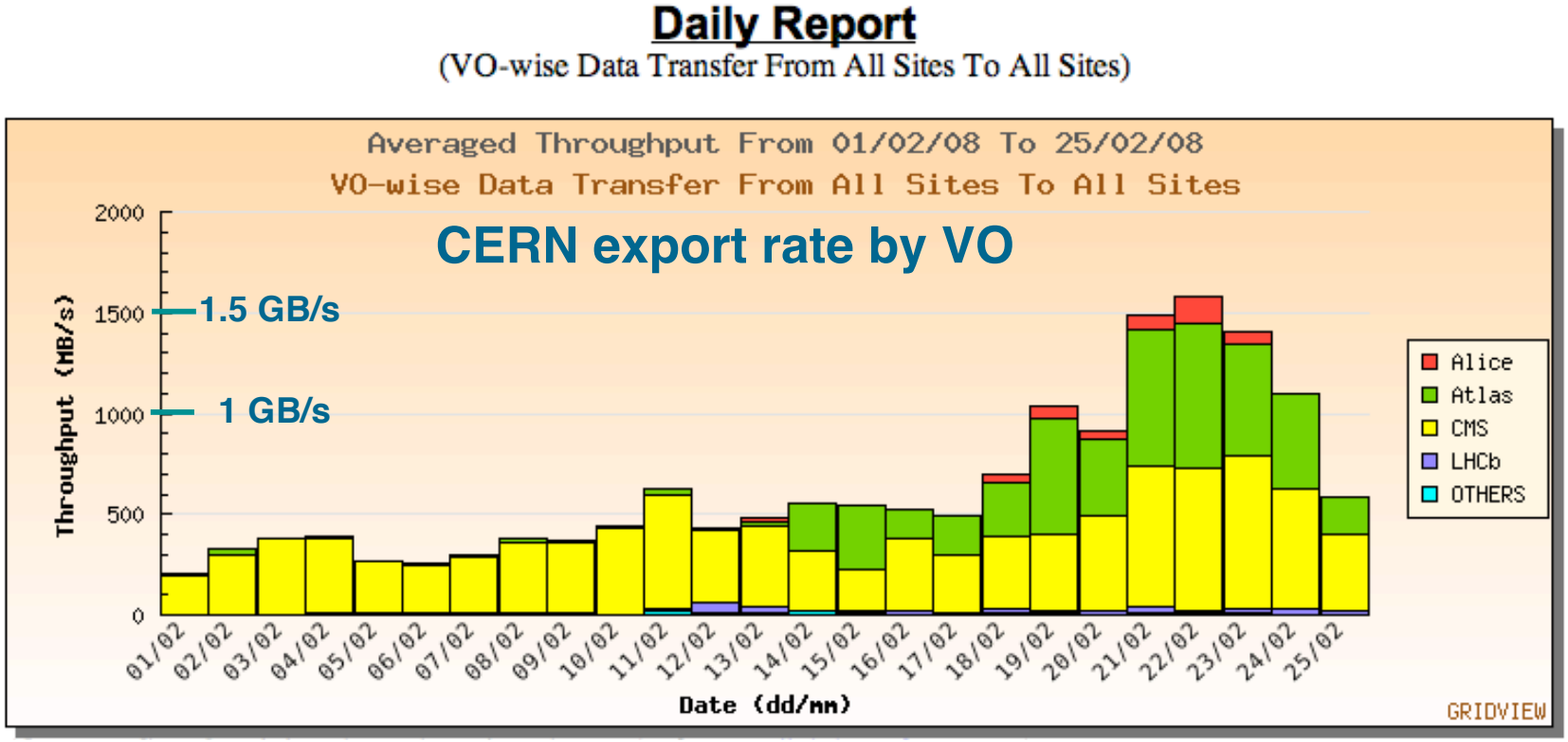


CCRC08 Transfer tests

Coordinated by D.Bonacorsi / FacOps

Goal: verify performance under CMS + ATLAS load

- CERN export and T1 import
- T1/T2 export + import



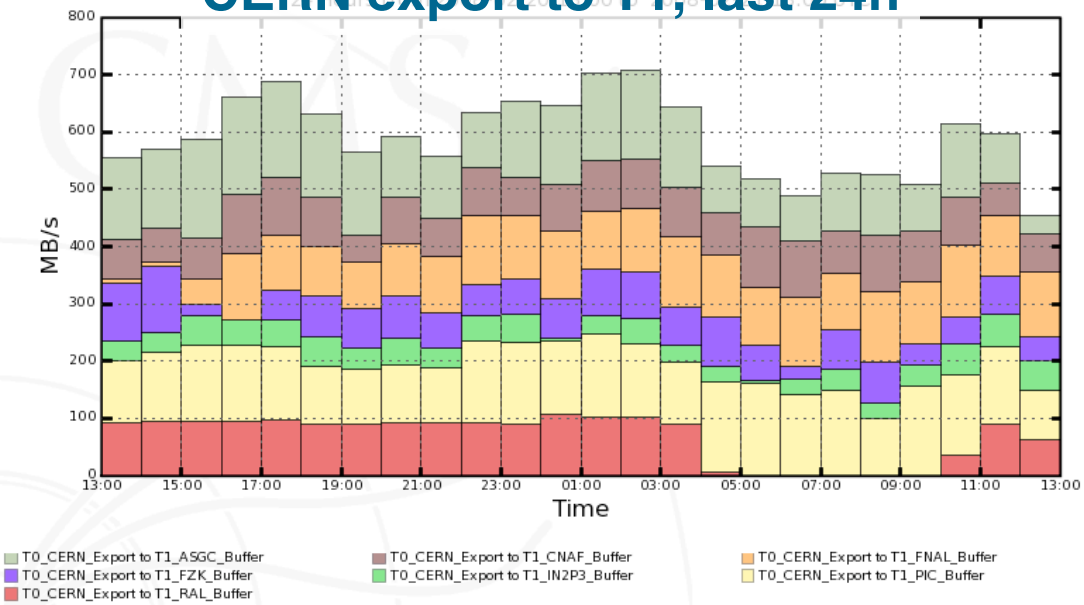
CCRC08 Transfer tests

Goal: use SRMv2 data transfers where possible

Target rates:
- **T0-T1: 25/40/50% of full 2008**
- **T1-T1: 50% in+outbound**
- **T1-regional-T2: full/high rate**
- **T2-regional-T1: full/high rate**

A detailed Plan worked out:
- **cycle through different parts of all link combinations per week**

CERN export to T1, last 24h

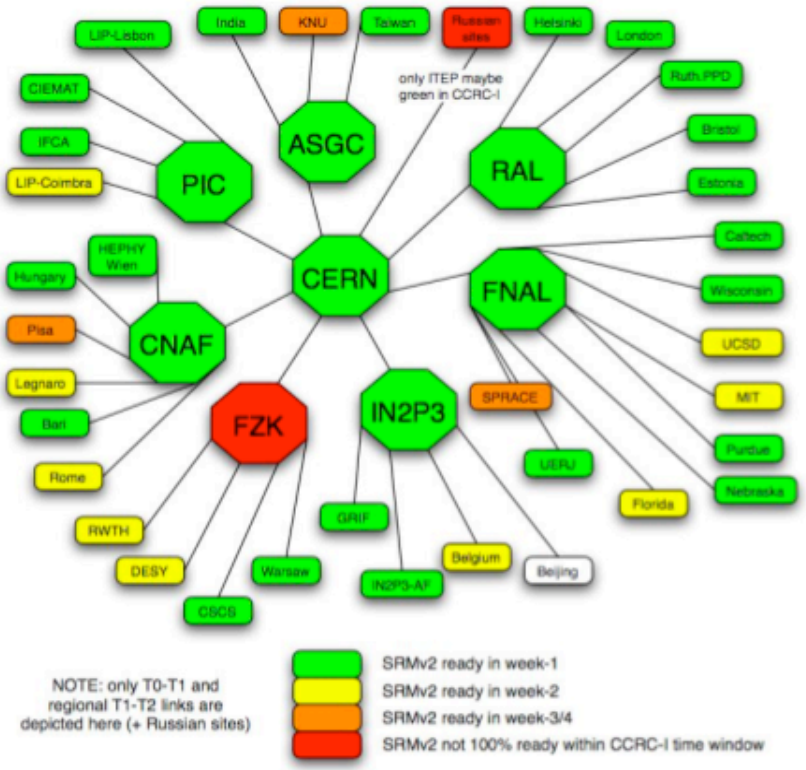


Tests are progressing well
- **T0-T1 metric goal by all all T1's**
- **5 out of 7 T1's reached T1-T1 goal**
- **individual problems are being addressed and result in delayed testing**
- **More detailed analysis available at the end of February**

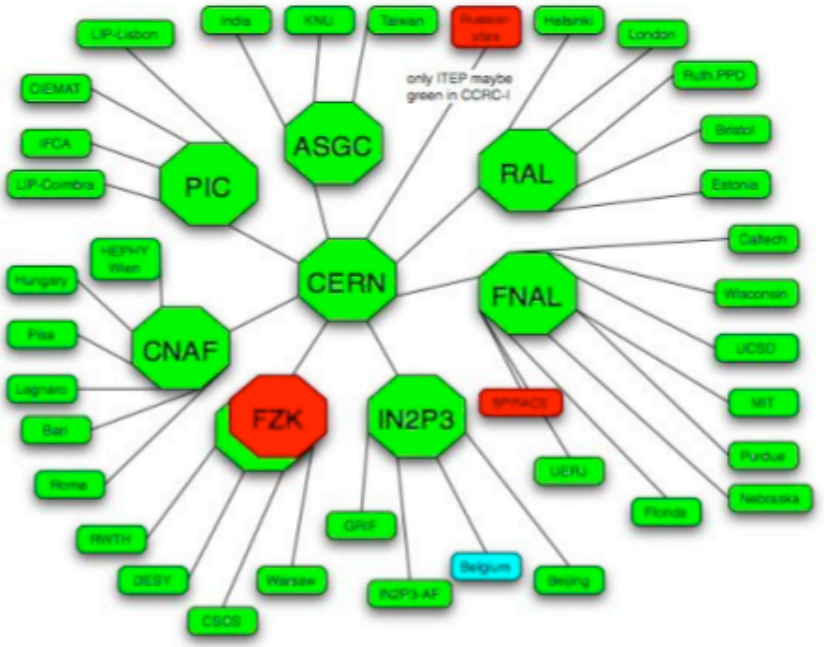
Data Transfer Tests Results

SRMv2 deployment status for CMS Tiers

At the start of CCRC:
(week-1 day-1)



End of CCRC week-3:
(week-3 day-7)



Situation much improved in all region, and faster than expected, during CCRC weeks-1/2/3:

- *Check details out at:*
  https://twiki.cern.ch/twiki/bin/view/CMS/Tier2SRM

CCRC08 Re-Reconstruction tests

Coordinated by G.Gomez-Ceballos, Josep Flix

Goal: measure performance of:

- **Migration from Tape to Buffer: pre-stage test.**
- **Reprocessing exercise: use all available CMS CPU-slots at T1s**

Plan:

- **Select one (or more) dataset(s) of ~10TB size existing at T1.**
- **Remove all the files from disk (aka, T1_Buffer).**
- **Fire the staging from Tape to Buffer of all files.**
- **Monitor the process and provide some measurements/plots**
- **Run Re-reconstruction over CSA07 data present at all T1s**
  - **Measure performance**

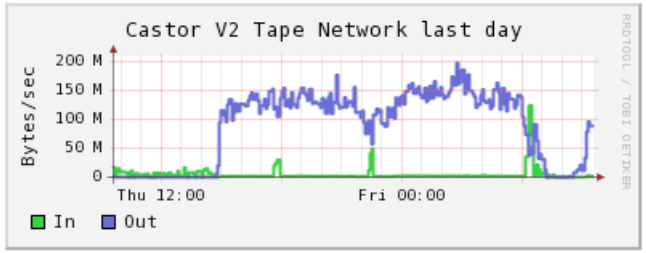
CCRC08 Re-Reconstruction tests

Status:

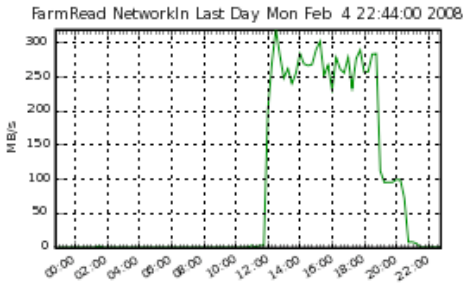
- **buffer to tape migration successfully finished at all sites**
  - **Results: total staging time 8-44h, rate: ~80-250MB/s observed**
    - Except IN2P3, performance was poor, reconfigure and redo

ASGC Recall Test: comments and plots

- Recall test data tape throughput:



- Tape->Buffer throughput: **RAL**



- **high performance processing without overlap with ATLAS**
  - **Finished at FNAL(1200 slots), CNAF(1000-1300 slots), FZK(600 slots), ASGC(300 slots)**
  - **IN2P3 and PIC, RAL: normal prosessing, no problem foreseen**
- **Processing test together with ATLAS planned at two Tier-1's:**
  - **special queue for Atlas and CMS is setup at IN2P3 and PIC**

Staging results at T1's

Migration from Tape to Buffer: pre-stage test

- **Obtained Results:**

T1 site	Data [TBs]	# Files	# Tapes	Staging req. time [min]	Staging time [h]	<MB/s> Tape->Buffer
RAL	10.5	5376	19	10'	10	290 MB/s
ASGC	13.2	5632	360	18'	22	150 MB/s
FNAL	10.0	5736	270	13'	25	110 MB/s
PIC	11.6	4744	38	300'	33	100 MB/s
FZK	10.0	4000	50	180'	27	90 MB/s
CNAF	10.8	7235	426	45'	79	40 MB/s
IN2P3	10.0	11061	68	2'	120	23 MB/s

Staging time for 10 TBs: ~24h (except RAL and IN2P3,CNAF)

T1 site	<# files>/tape	<# files>/mount	# Mounts total	# Mounts/ # Tapes	file failures [%]
RAL	283	132	41	2,2	0%
ASGC	15.6	9.4	601	1,7	0.7%
FNAL	21.2	--	--	--	0%
PIC	125	83,2	57	1,5	0%
FZK	80	2	2000	40,0	0%
CNAF	17.0	2.1	3406	8,0	7,6%
IN2P3	163	3	3687	54,2	0%

In general, rather good strategies for staging followed at sites

CCRC08 Monte Carlo tests

Coordinated by DataOps

Goal:

- **Production tests of FastSim Monte Carlo**
- **Physics groups want to use 50M of the CSA07 samples (100pb^{-1} calibration), reading AOD's.**

Status:

- **Fast Simulation production based on CMSSW_1_6_9 completed successfully 50M on Monday**
  - because of data handling: used resources at T0/T1

CCRC08 CAF tests

Coordinated by P.Kreuzer

Goal:

- ramp-up CAF resources
- verify basics CMS use cases at scale

	CPU 70% Dual quad-core (16GB RAM) 30% Dual Dual-core (8GB RAM)	Disk	Tape
T0 2008	3000 slots (1000 slots in '07)	400 TB (420TB in '07)	3 PB
CAF 2008	1200 slots (128 slots in '07)	1600 TB (35 TB in '07)	
CAF CCRC'08	250 slots	150-200TB	

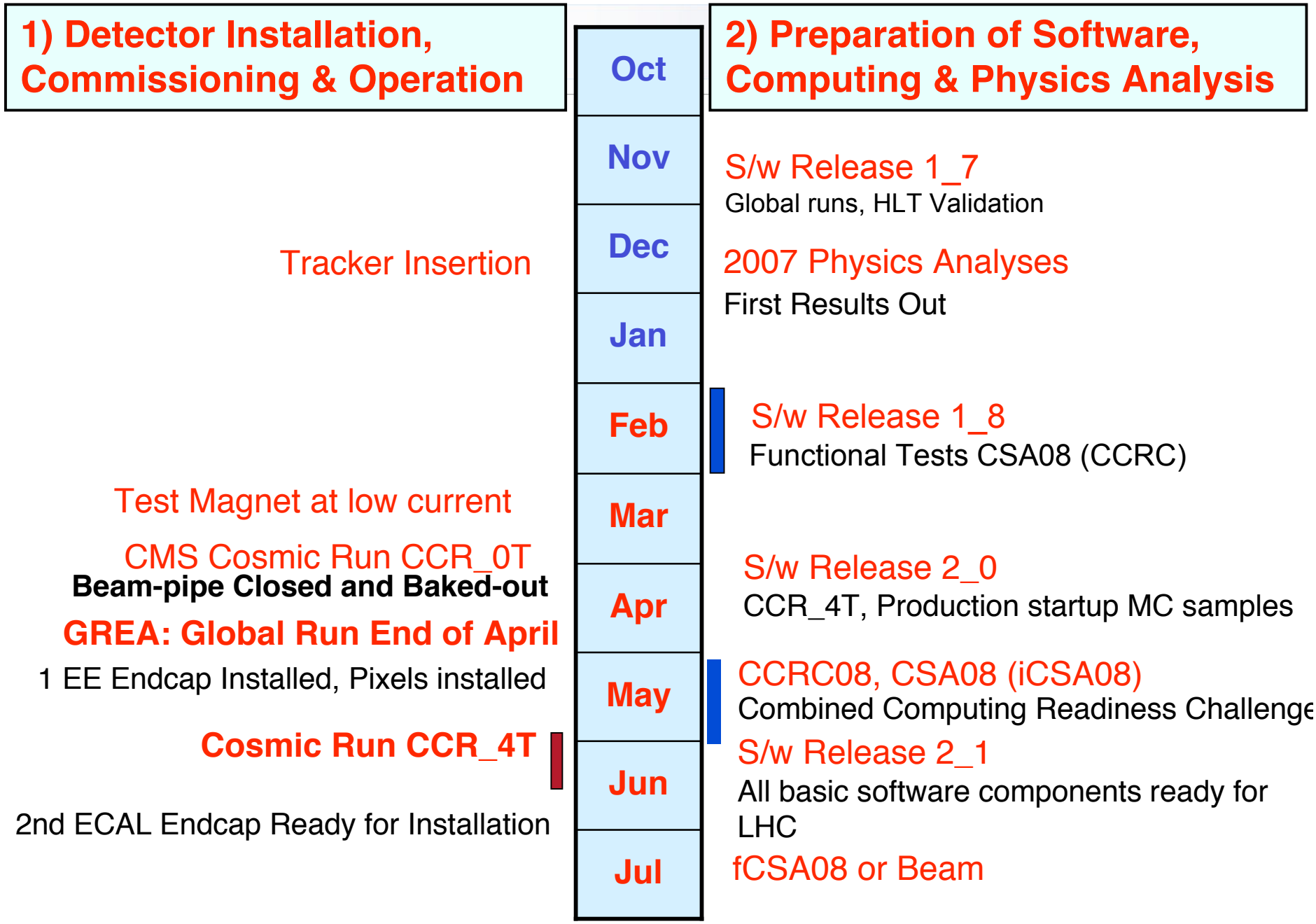
note: 3000 slots =~5.3MSI2K , 1200 slots =~ 2.1MSI2K

Status: good progress made

- resources configured according to plan
- Regular CAF meetings with user representatives (Global Run, ALCA and Physics)
- Plan for CCRC08 (week 3 and 4):
  - Transfer GR data fromT0 to CAF and populate local DBS
  - Finalize RPC workflow
  - Test/Run HcalCallsoTracks workflow
  - Test/Run Muon Alignment workflow
  - Setup and test CRAB, local submission
  - Collect list of CAF groups and users per group. Provide to IT, both for batch/interactive CAF

Computing Summary

- The Computing infrastructure is fully utilized for ongoing production
  - Finished original CSA07 production (and much more)
- Detailed analysis of CSA07 performance was performed.
  - Direct result for Computing:
    defined PADA tasks and CCRC08 functional tests
- The PADA taskforce addresses deployment, integration, commissioning and scale testing. It will bring the elements of the Computing Program into stable and scalable operations.
- The CCRC08 functional tests in February complements CSA07 and test additional functionality.
- Detailed planning of CCRC08(May), *i*-CSA08 and *f*-CSA08 is going on, expect to agree on initial scope and goals during CMS week.


1) Detector Installation, Commissioning & Operation
2) Preparation of Software, Computing & Physics Analysis
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
S/w Release 1_7
Global runs, HLT Validation
Tracker Insertion
2007 Physics Analyses
First Results Out
S/w Release 1_8
Functional Tests CSA08 (CCRC)
Test Magnet at low current
CMS Cosmic Run CCR_0T
Beam-pipe Closed and Baked-out
GREA: Global Run End of April
S/w Release 2_0
CCR_4T, Production startup MC samples
1 EE Endcap Installed, Pixels installed
CCRC08, CSA08 (iCSA08)
Combined Computing Readiness Challenge
Cosmic Run CCR_4T
S/w Release 2_1
All basic software components ready for LHC
2nd ECAL Endcap Ready for Installation
fCSA08 or Beam