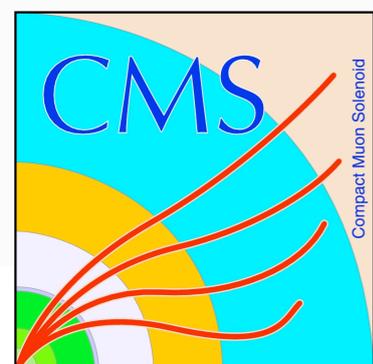


# CMS Computing Operations

Scientific Computing Division Projects Meeting # 6  
19. December 2012

Oliver Gutsche  
for  
CMS Computing Operations





- ▶ CMS Computing Operations
  - ▶ Central support for all CMS sites
  - ▶ Operation of all central workflows: Processing & MC production
  - ▶ Operation of central services
    - ▶ Meta data catalog (DBS)
    - ▶ Transfer System (PhEDEX)
    - ▶ Access to conditions (Frontier/Squid)
    - ▶ SAM tests/HammerCloud tests/Site Readiness/
    - ▶ ...
  - ▶ Team of scientists and engineers distributed worldwide
    - ▶ 60 individuals
    - ▶ 30 FTE

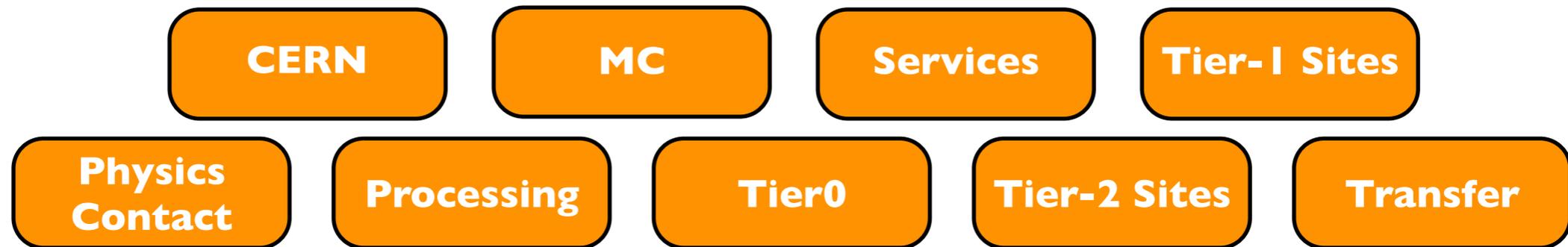
Tier-Level	US Sites	Non-US Sites
Tier-0		1
Tier-1	1	6
Tier-2	8	46

L2

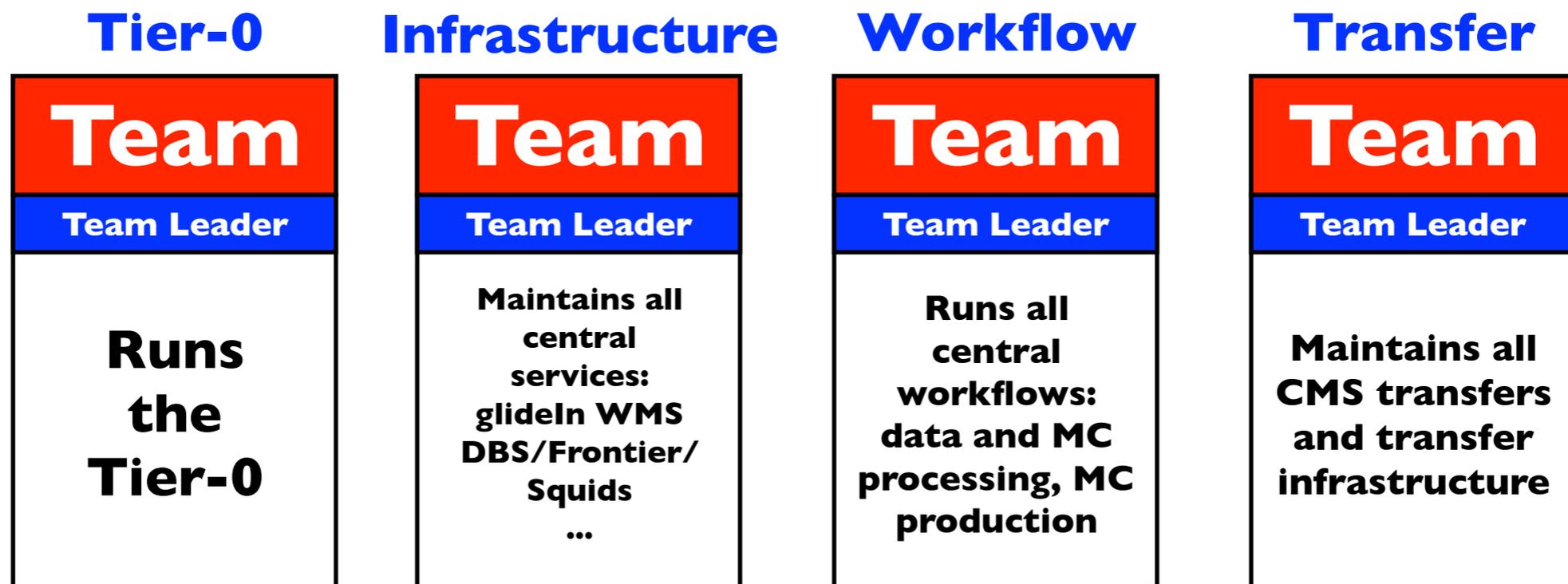


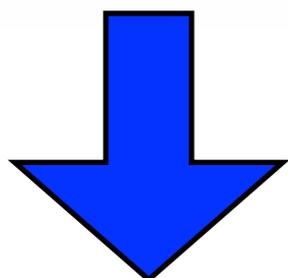
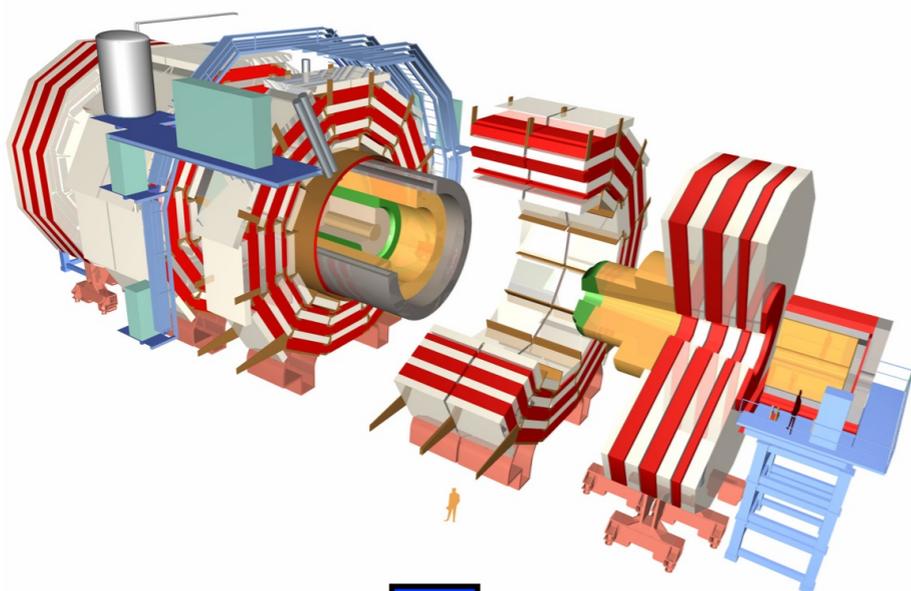
L3

Coordination

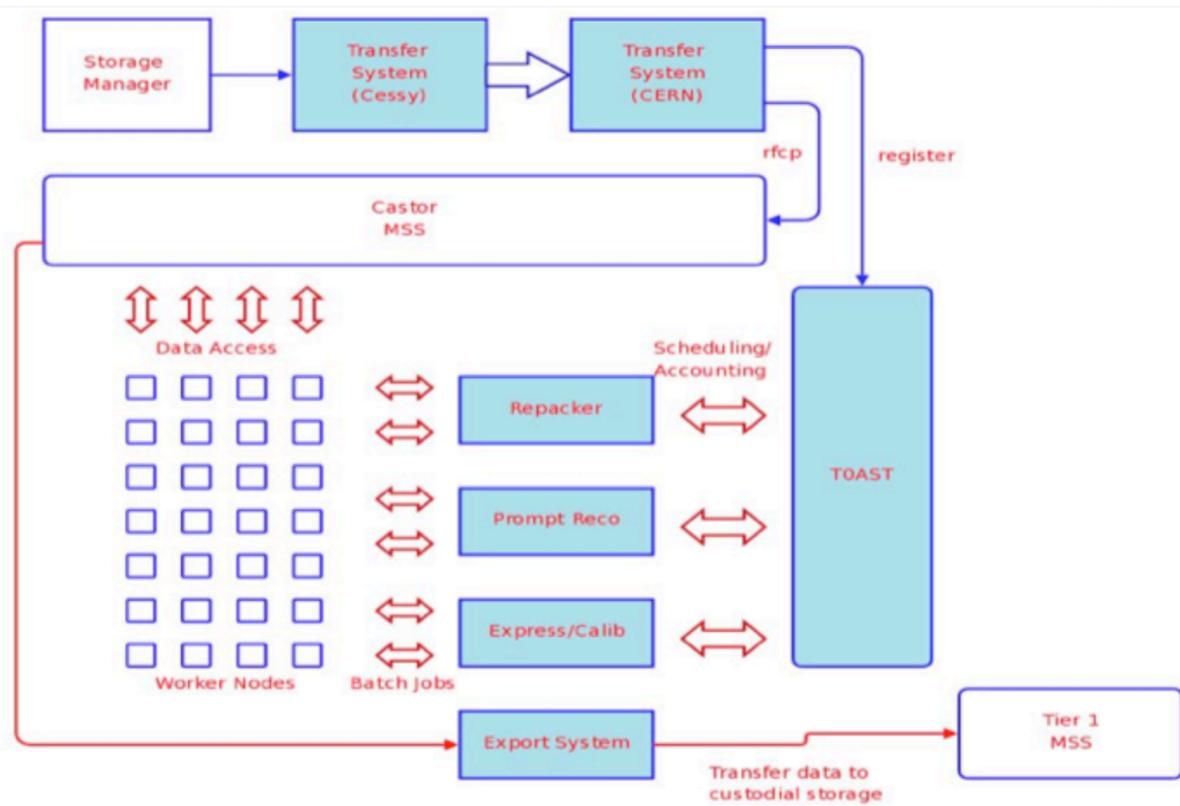


CORE



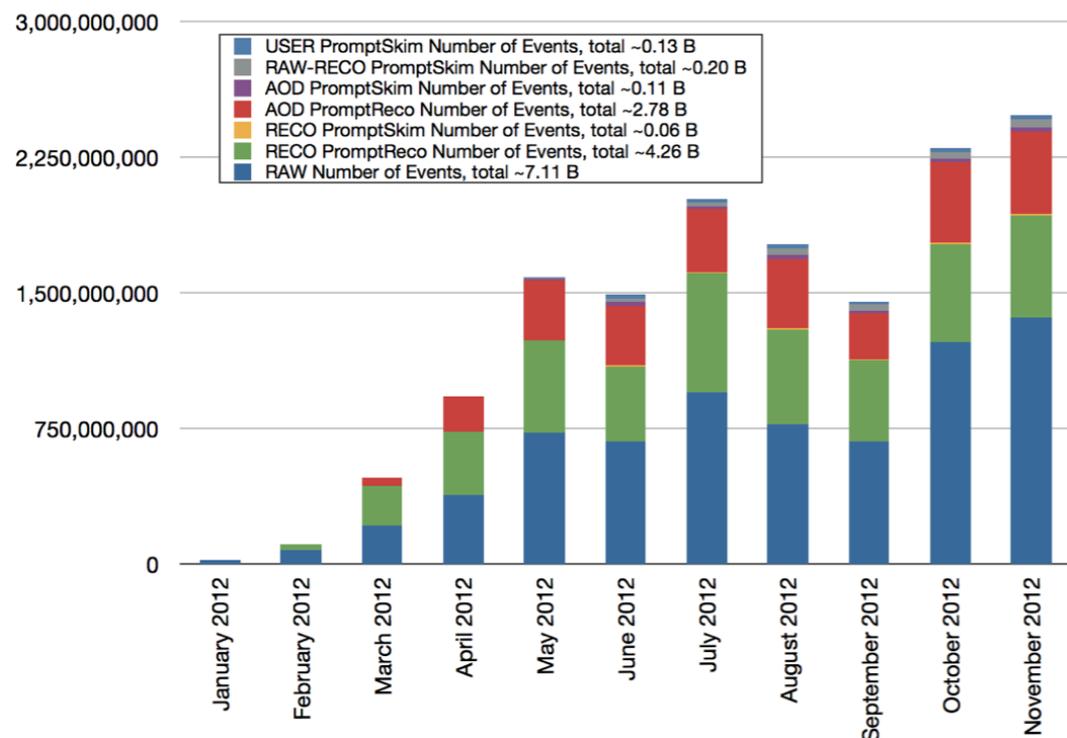


Recorded events

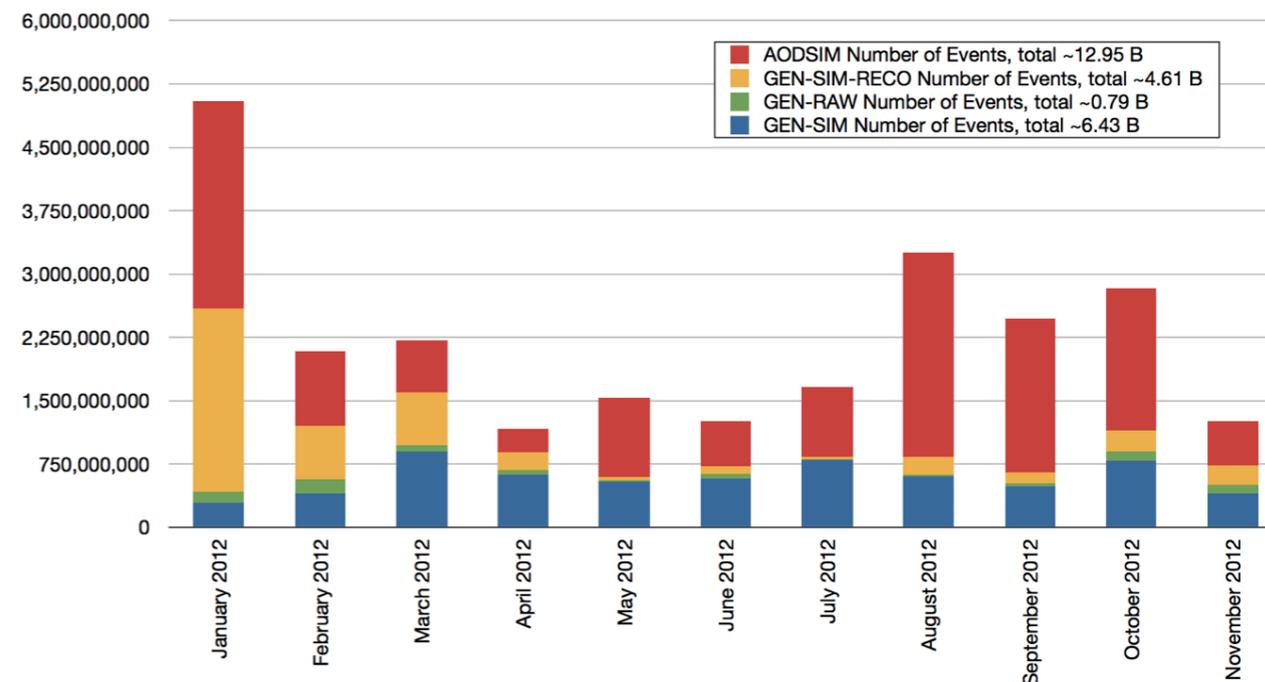


- ▶ Time critical processing of data coming from the detector
- ▶ Express: repack and reconstruct 10% of data with 1 hour latency
  - ▶ Used for monitoring and prompt calibration and alignment workflows
- ▶ Repacking: convert binary format of RAW data into ROOT files
- ▶ PromptReconstruction: after 48 hour waiting time to allow for prompt alignment & calibration, reconstruct data
- ▶ Transfer all RAW and RECO/AOD to Tier-I sites for tape archival (hot copy)
  - ▶ Store 2<sup>nd</sup> copy at CERN on tape (cold copy)
- ▶ System developed by Dirk Hufnagel (FNAL)
- ▶ Operated by visiting scientist at FNAL: Diego Ballesteros (He sits in the ROC!)

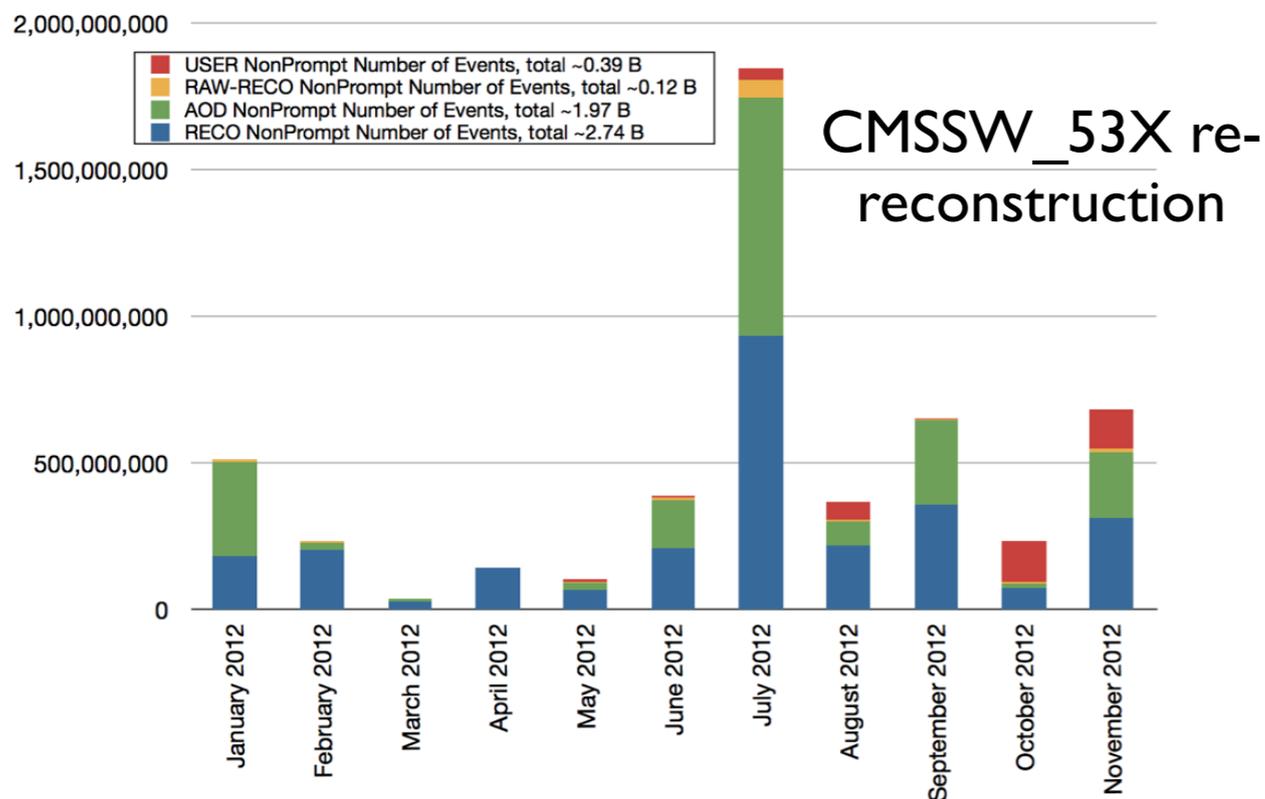
Prompt Data in 2012: Number of Events per Month



MC in 2012: Number of Events per Month



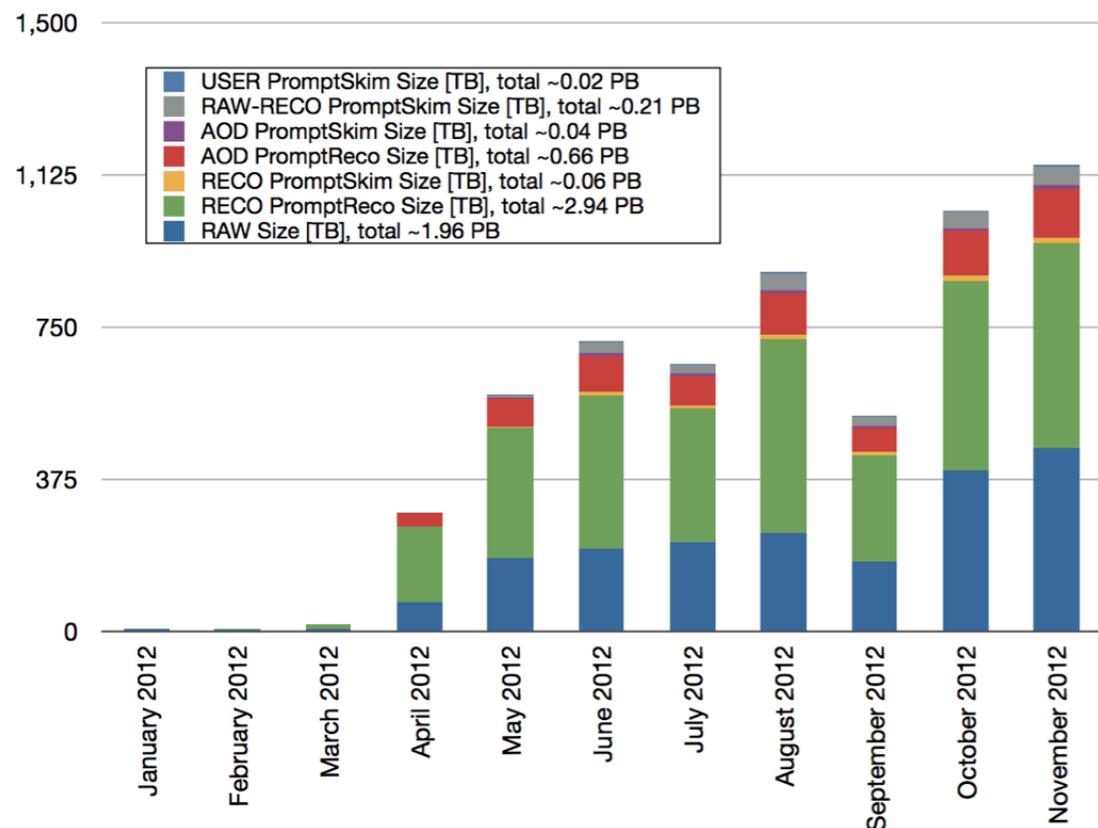
Re-reconstruction of Data in 2012: Number of Events per Month



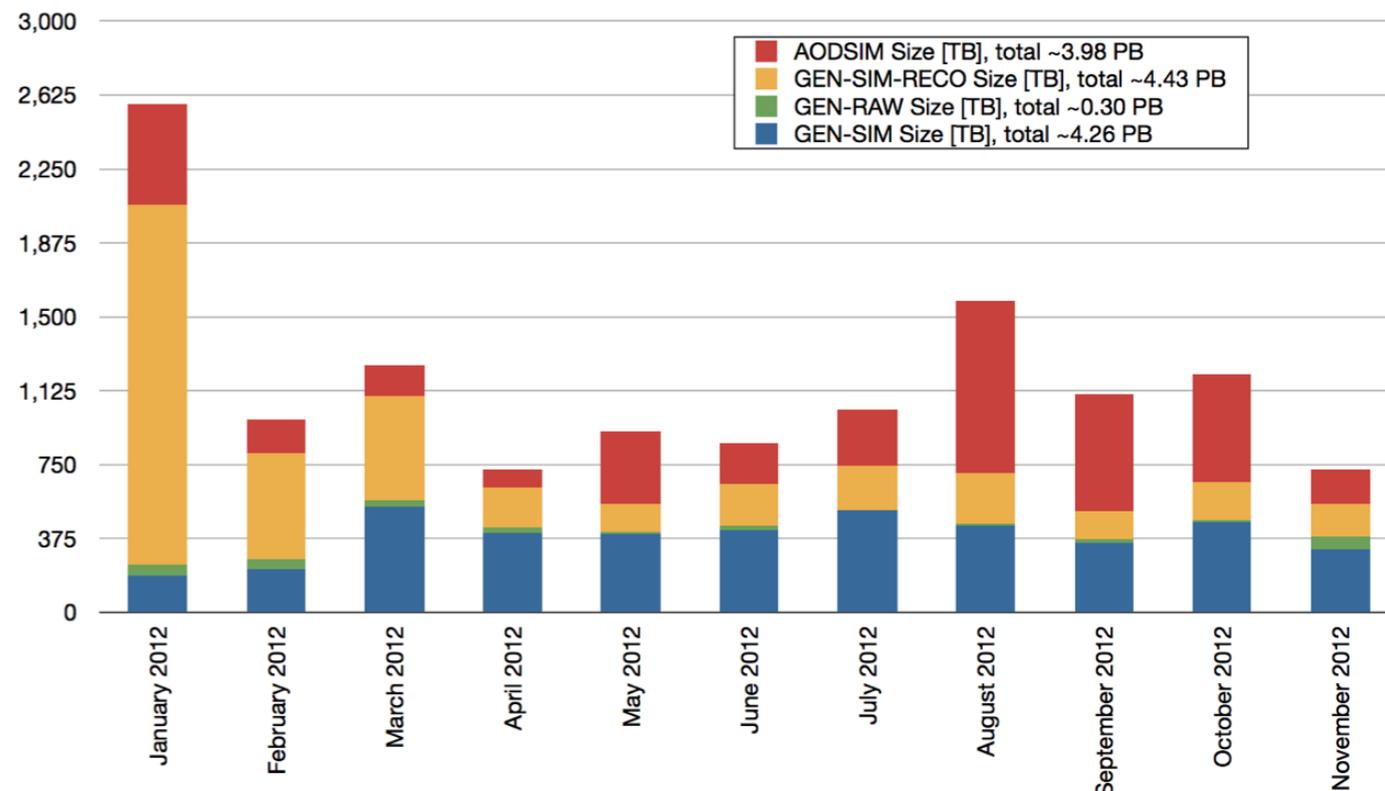
## ► Successful year 2012

- Recorded over **7 Billion RAW** physics events (includes parked PDs)
  - Change of trigger in October clearly visible on plot
  - More parked data = only record and store RAW, reconstruct later in 2013
- Re-reconstructed close to **2 Billion AOD** events
  - CMSSW\_53X re-reconstruction clearly visible on plot
- Simulated close to **6.5 Billion GEN-SIM** events

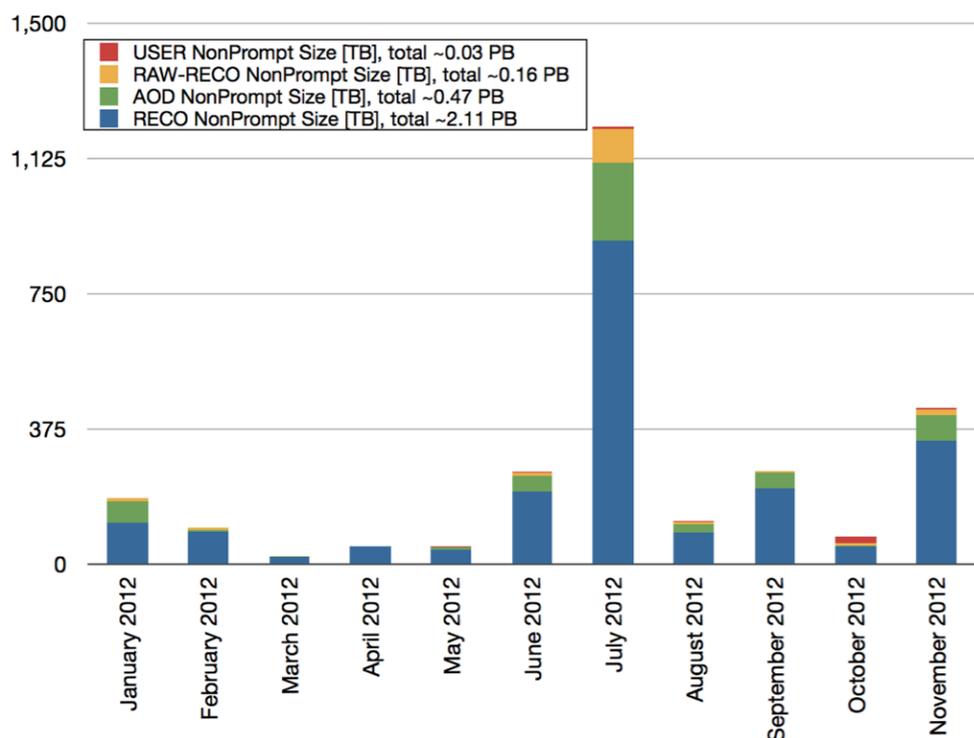
Prompt Data in 2012: Size in TB per Month



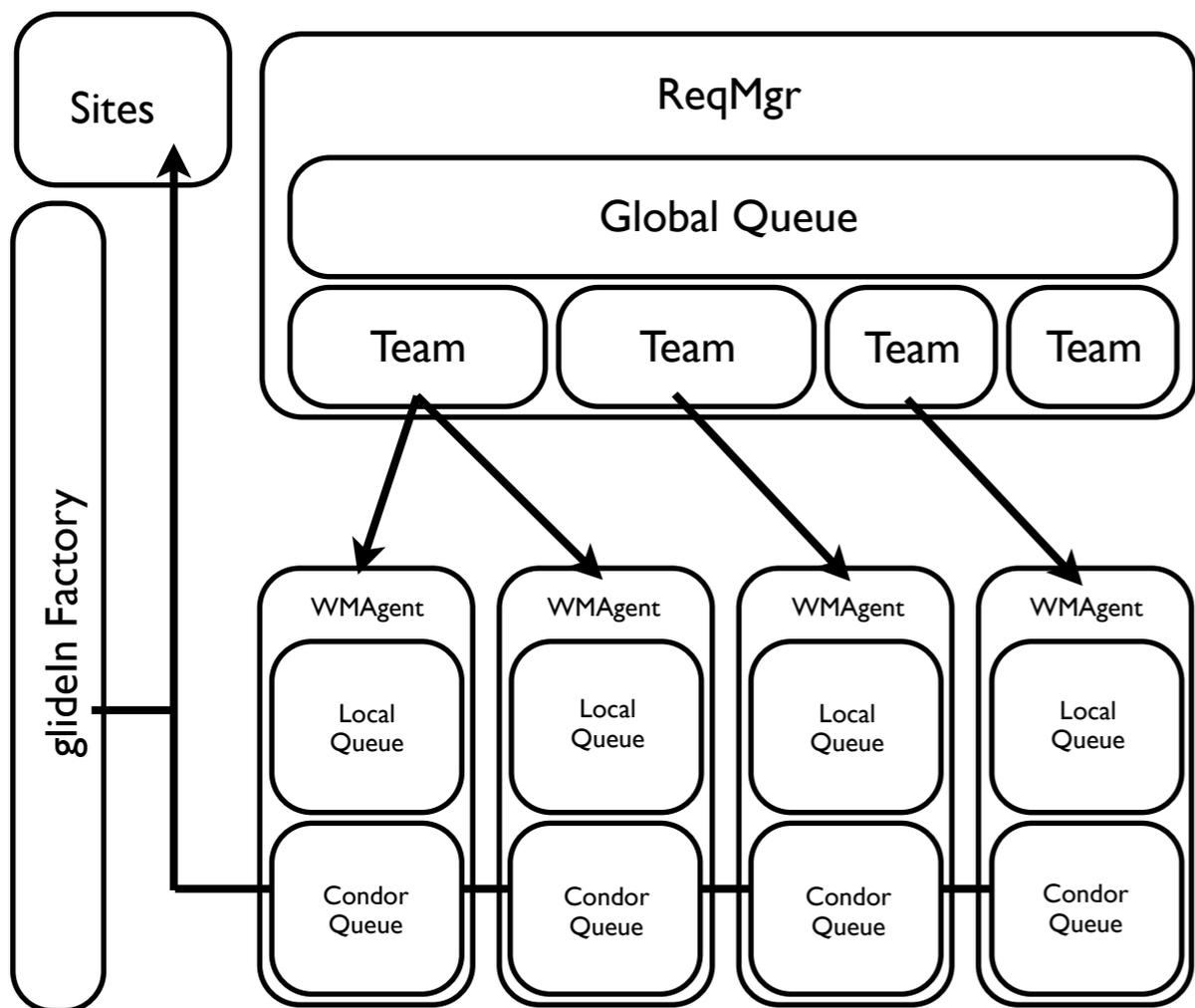
MC in 2012: Size in TB per Month



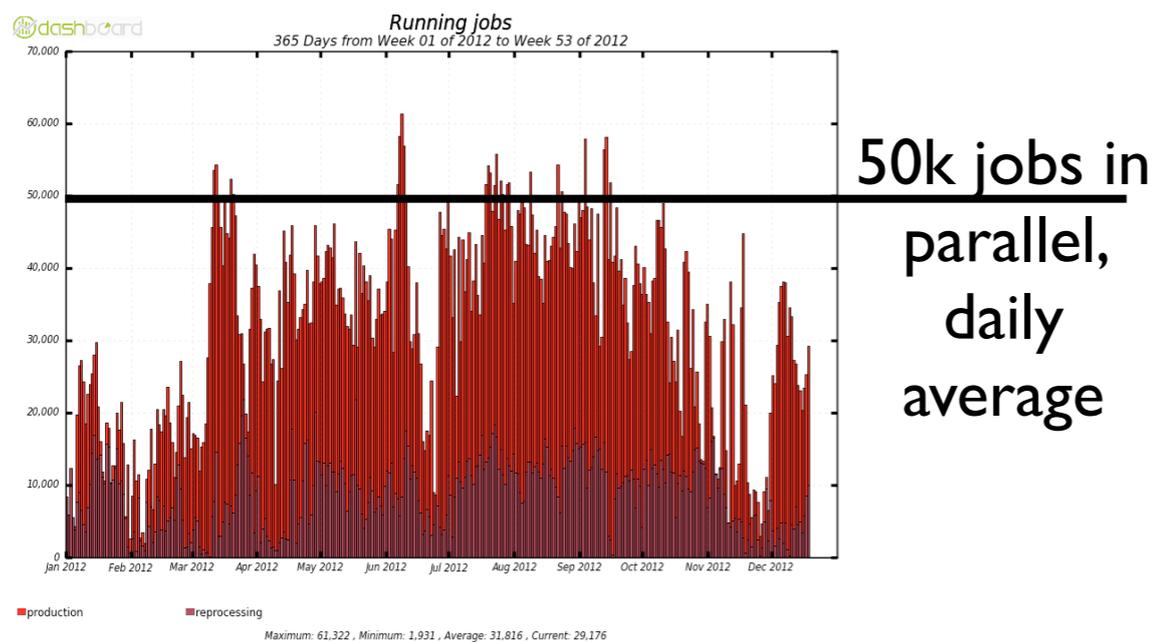
Re-reconstruction of Data in 2012: Size in TB per Month



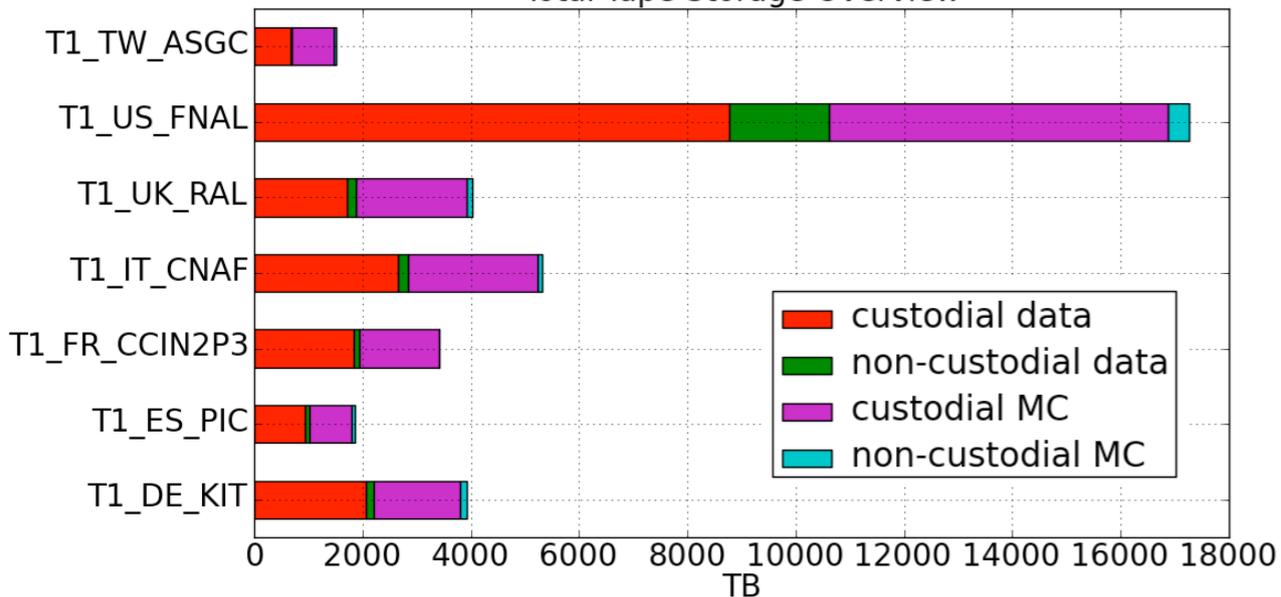
- ▶ Recorded close to **6 PB** of data on tape
- ▶ Re-reconstruction of data produced close to **2.8 PB** on tape
- ▶ MC simulation produced close to **13 PB** on tape



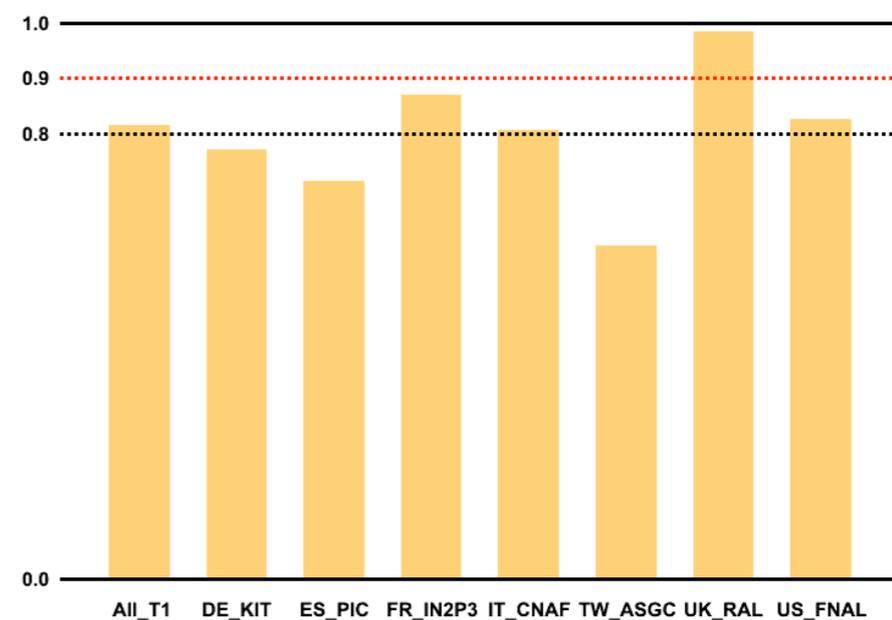
- ▶ Run by teams of operators
- ▶ Jen Adelman-McCarthy (FNAL) leads the operator team (She sits in the ROC!)
- ▶ Infrastructure mainly developed by Fermilab people
- ▶ Current lead developer: Steve Foulkes (FNAL)
- ▶ Concentrating currently on monitoring: Seangchan Ryu (FNAL)
- ▶ Operational support & development:
  - ▶ Visiting scientist at FNAL: Diego Ballesteros (He sits in the ROC!)



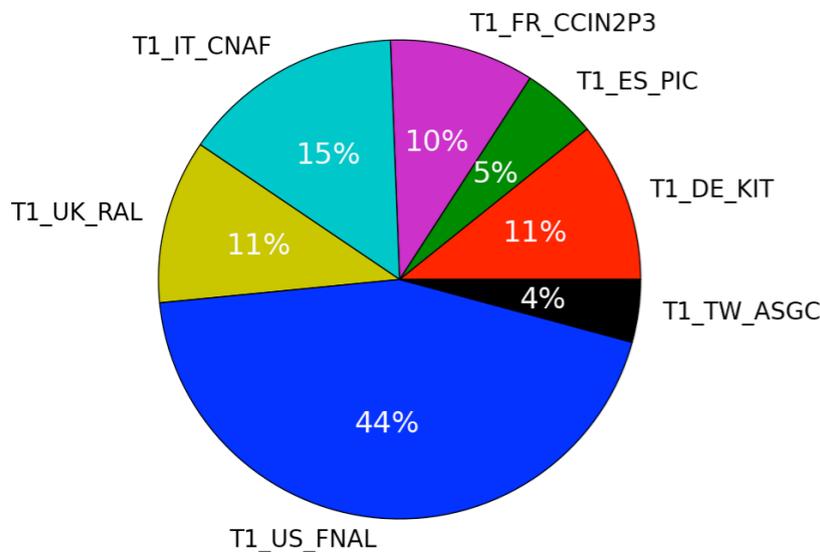
Total Tape Storage Overview



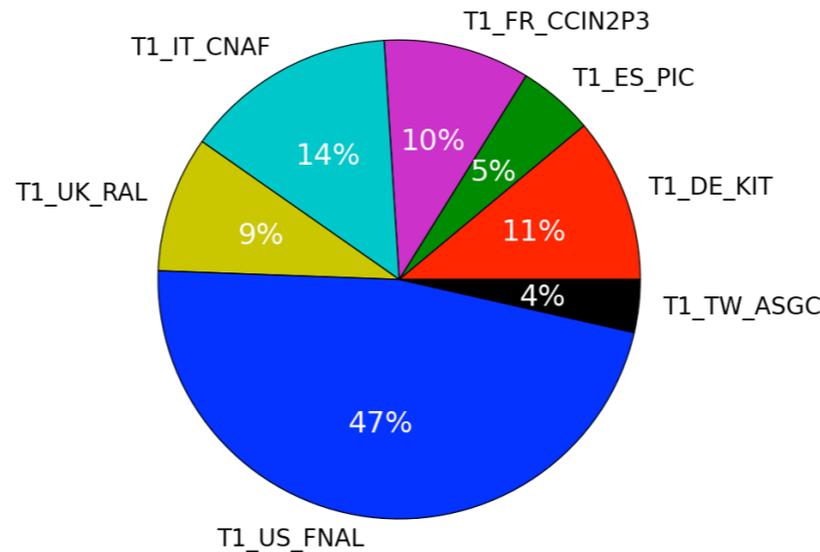
Tier-1 Storage Utilization (w.r.t pledge) CMS Computing Operation Dec. 2012



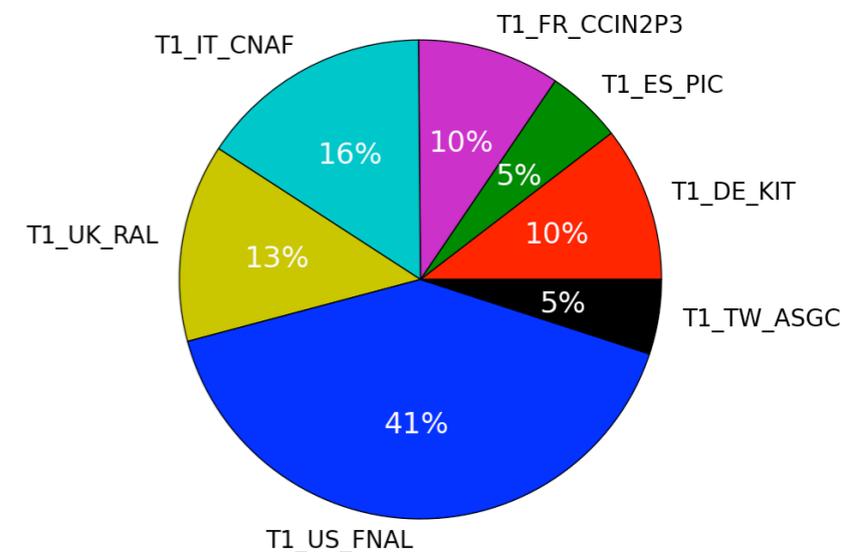
Custodial Tape Storage Overview



Custodial Data Tape Storage Overview



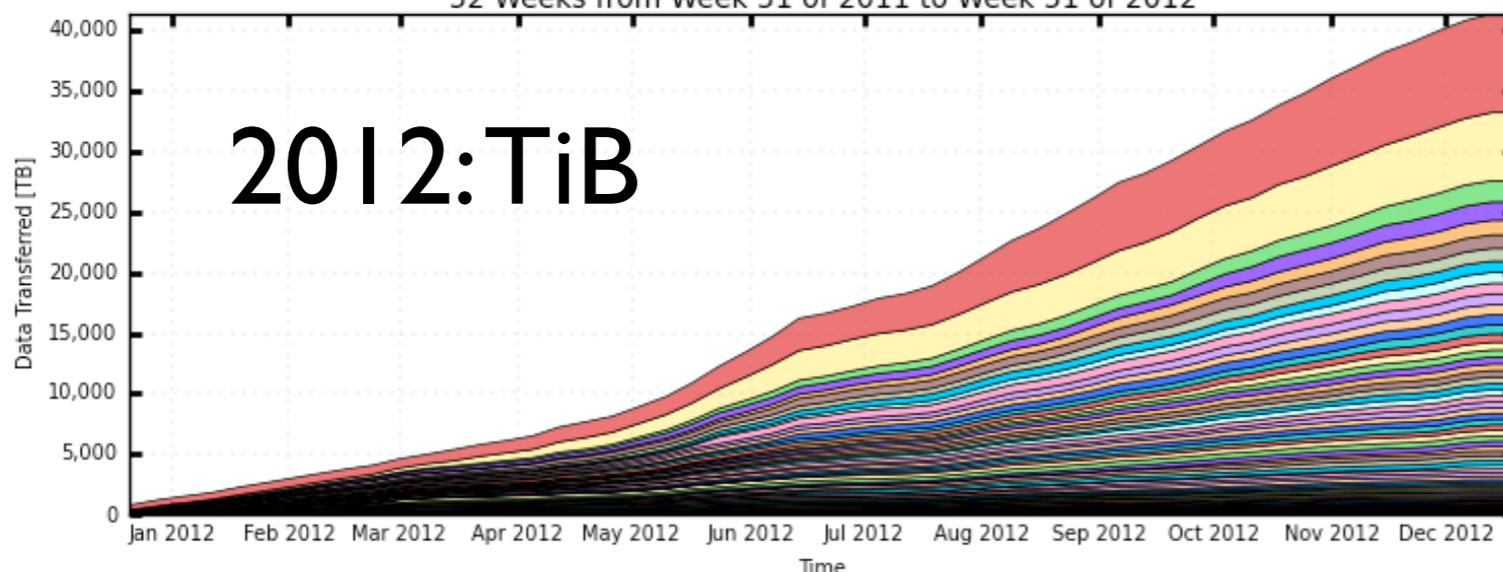
Custodial MC Tape Storage Overview



▶ We are currently using 35 PB out of 45 PB pledged.

▶ We are preparing to delete over 8 PB mostly RECO datasets that are not used by the end of the year

**CMS PhEDEx - Cumulative Transfer Volume**  
52 Weeks from Week 51 of 2011 to Week 51 of 2012



2012:TiB

- T1\_US\_FNAL\_Buffer
  - T2\_UK\_London\_IC
  - T2\_US\_Florida
  - T2\_EE\_Estonia
  - T2\_IT\_Bari
  - T2\_BE\_UCL
  - T2\_DE\_RWTH
  - T2\_BR\_SPRACE
  - T3\_US\_TAMU
  - T3\_US\_Minnesota
  - T2\_CH\_CERN
  - T2\_US\_Wisconsin
  - T2\_US\_Nebraska
  - T2\_US\_UCSD
  - T2\_IT\_Rome
  - T3\_US\_Colorado
  - T2\_ES\_IFCA
  - T2\_CN\_Beijing
  - T1\_IT\_CNAF\_Buffer
  - T1\_FR\_CCIN2P3\_Buffer
  - T2\_US\_MIT
  - T2\_ES\_CIEMAT
  - T2\_FR\_IPHC
  - T1\_ES\_PIC\_Buffer
  - T2\_US\_Vanderbilt
  - T2\_TW\_Taiwan
  - T2\_RU\_JINR
  - T2\_AT\_Vienna
  - T2\_DE\_DESY
  - T1\_DE\_KIT\_Buffer
  - T3\_US\_FNALLPC
  - T2\_FR\_GRIF\_LL
  - T2\_US\_Caltech
  - T2\_FR\_GRIF\_IRFU
  - T2\_UK\_London\_Brunel
  - T2\_CH\_CSCS
  - T3\_CH\_PSI
  - T2\_UK\_SGrid\_Bristol
  - T1\_UK\_RAL\_Buffer
  - T2\_BE\_IHE
  - T2\_US\_Purdue
  - T2\_UK\_SGrid\_RALPP
  - T1\_TW\_ASGC\_Buffer
  - T2\_IT\_Pisa
  - T2\_IT\_Legnaro
  - T2\_IN\_TIFR
  - T2\_FR\_CCIN2P3
  - ... plus 47 more
- Total: 41,408 TB, Average Rate: 0.00 TB/s

- ▶ Transfers for CMS in 2012
- ▶ Average 0.9 PB per week
- ▶ Total transfer volume between all sites: 46 PB
- ▶ Teams of operators maintain the transfer system
- ▶ Lead by visiting scientist at FNAL: Marek Zielinski
- ▶ Taking especially care of FNAL transfer issues and subscriptions to the T3\_US\_FNALLPC: visiting scientist at FNAL: Si “Lucy” Liu

2012:TB

Exported from	Volume [PB]
Tier-0	10
Tier-1	23
Tier-2	13

- ▶ PerfSONAR: toolkit to monitor and debug underlying network infrastructure
- ▶ Each site has PerfSONAR instances to run latency (packet loss) and throughput tests
- ▶ Marek and Si established PerfSonar in CMS
- ▶ Started with USCMS cloud
- ▶ Final setup: regional clouds around Tier-I centers and one inter-cloud connecting the regions
- ▶ Si is also heavily involved in the development of the new DashBoard, writing the GUI
- ▶ Modular approach, better performance, for both Atlas and CMS

USCMS Latency Matrix

	---	0	1	2	3	4	5	6	7	8	9
0: Caltech (perfsnar2.ultraflight.org)	---	0.0	0.0	41.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1: FNAL (cmsperfsnar01.fnal.gov)	0.0	0.0	8.0	0.0	31.0	1.0	65.0	1.0	40.0	1.0	0.0
2: Florida (perfsnar1.lhepa.ufl.edu)	0.0	1.0	0.0	1.0	1.0	3.0	1.0	1.0	2.0	3.0	0.0
3: MIT (perfsnar01.cmsaf.mit.edu)	0.0	0.0	33.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4: Nebraska (hcc-ps01.unl.edu)	0.0	0.0	32.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5: Purdue (mannperf.itns.purdue.edu)	0.0	0.0	37.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6: SPRACE (perfsnar-lt.sprace.org.br)	0.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7: UCSD (perfsnar-2.t2.ucsd.edu)	0.0	0.0	34.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8: Vanderbilt (psndt2.accre.vanderbilt.edu)	0.0	7.0	34.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9: Wisconsin (perfsnar01.hep.wisc.edu)	0.0	0.0	42.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0

USCMS Throughput Matrix

	---	0	1	2	3	4	5	6	7	8	9
0: Caltech (perfsnar.ultraflight.org)	---	0.62	0.18	0.71	0.67	0.58	0.32	0.91	0.64	0.51	0.91
1: FNAL (cmsperfsnar01.fnal.gov)	0.75	---	0.60	0.92	0.75	0.93	0.34	0.90	0.72	0.92	0.92
2: Florida (perfsnar2.lhepa.ufl.edu)	0.00	0.52	---	0.32	0.54	0.19	0.15	0.45	0.41	0.24	0.24
3: MIT (PERFSNAR02.CMSAF.MIT.EDU)	0.75	0.46	0.34	---	0.88	0.90	0.00	0.76	0.84	0.90	0.90
4: Nebraska (hcc-ps02.unl.edu)	0.00	0.64	0.19	0.91	---	0.86	0.40	0.85	0.89	0.92	0.92
5: Purdue (mannperf2.itns.purdue.edu)	0.73	0.82	0.26	0.87	0.71	---	0.34	0.74	0.00	0.83	0.83
6: SPRACE (perfsnar-bw.sprace.org.br)	0.00	0.44	0.55	0.42	0.45	0.34	---	0.33	0.50	0.33	0.33
7: UCSD (perfsnar-1.t2.ucsd.edu)	0.89	0.00	0.00	0.00	0.38	0.24	0.00	---	0.48	0.45	0.45
8: Vanderbilt (psndt2.accre.vanderbilt.edu)	0.00	0.69	0.00	0.93	0.93	0.00	0.00	0.91	---	0.75	0.75
9: Wisconsin (perfsnar02.hep.wisc.edu)	0.59	0.75	0.92	0.88	0.67	0.93	0.19	0.87	0.75	---	---

### PerfSONAR Dashboard

**Navigator**

- List of Hosts
- List of Sites
- List of Clouds
- List of Matrices

**List of Clouds**

Cloud Number :2

Cloud	Actions			
<a href="#">USATLAS</a>	<a href="#">edit</a>	<a href="#">delete</a>	<a href="#">add/remove sites</a>	<a href="#">add/remove matrices</a>
<a href="#">USCMS</a>	<a href="#">edit</a>	<a href="#">delete</a>	<a href="#">add/remove sites</a>	<a href="#">add/remove matrices</a>

[create a new cloud](#)

### PerfSONAR Dashboard

**Navigator**

- List of Hosts
- List of Sites
- List of Clouds
- List of Matrices

	---	lhcmn.bn1.gov	iut2-net1.iu.edu	mwt2-ps01.campuscluster.illinois.edu	psum02.aglt2.org	mwt2-ps02.campuscluster.illinois.edu
lhcmn.bn1.gov	---	0.00	0.00	2.00	2.52	1.68
iut2-net1.iu.edu	0.00	0.00	---	0.00	0.00	0.00
mwt2-ps01.campuscluster.illinois.edu	0.00	0.00	0.00	---	0.00	0.00
psum02.aglt2.org	0.00	0.00	0.00	0.00	---	0.00
mwt2-ps02.campuscluster.illinois.edu	1.24	0.00	0.00	2.53	0.00	---

[return to the list of matrix](#)

- ▶ During the long shutdown ending 2015, following improvements/changes to the computing infrastructure are planned
  - ▶ Global glideIn WMS condor pool for CMS
  - ▶ Opportunistic resource usage using Parrot+CVMFS and cloud through glideIn WMS
    - ▶ HLT farm usage (10% of Tier-1 level resources) is close to be working
  - ▶ Separation of Disk/Tape at Tier-1 sites through PhEDEx endpoints
    - ▶ Taking over management of disk at Tier-1 sites
  - ▶ Xrootd federation of all files on disk on all CMS sites
    - ▶ Fall back for access problems
    - ▶ Increase flexibility by accessing samples through WAN for processing at Tier-1 sites exploiting the 120 Gpbs LHCOPN
      - ▶ Will try to use the general network infrastructure and LHCONE to do the same at big T2 sites
  - ▶ Finalize PerfSONAR deployment

## From the Scientific Computing Division

### Computing matters at CMS

Lothar Bauerdick, Oliver Gutsche and Burt Holzman, members of the CMS group at Fermilab, wrote this column.



Lothar Bauerdick

On Monday, the LHC ended its collisions of protons on protons for 2012 and is not sched

runn  
when  
shutd  
upgra  
comp  
comp  
collisi  
mark  
era fo  
during  
world  
peek  
energ

With the most successful year yet for CMS winding down, the work doesn't stop. While the physicists enjoy their well-deserved rest and recharge their batteries for an equally promising 2013, the computers will dutifully process data, analyze results and prepare for the higher-energy run to start in 2015, all under the watchful eyes of the CMS computing team at Fermilab.



Burt Holzman

though the collider ran at reduced energy. In this relatively short three-year run, the LHC was able to gather a sufficiently sizeable data set to observe what looks like the Higgs boson and make a broad range of interesting physics measurements. CMS has published more than 200 refereed physics publications to date on this data set and will continue to mine this data for the coming years while it repairs and upgrades its detectors.

- ▶ 2012 was a very successful year for the LHC, for CMS and for CMS at Fermilab
- ▶ The Proton-Proton run is officially over and we will resume with higher beam energies in 2015
- ▶ Computing Operations was at the forefront of enabling CMS to do all the exciting physics that was published in 2012
- ▶ The Fermilab teams were instrumental to this success
- ▶ The work does not stop, in 2013 big reprocessing campaigns are planned and in 2014 we have to produce new MC for ~13 TeV collisions
- ▶ And all of the planned improvements will be put in place to prepare for the restart in 2015