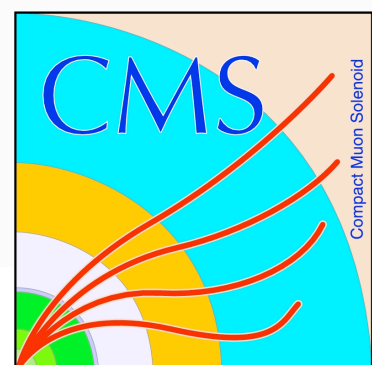


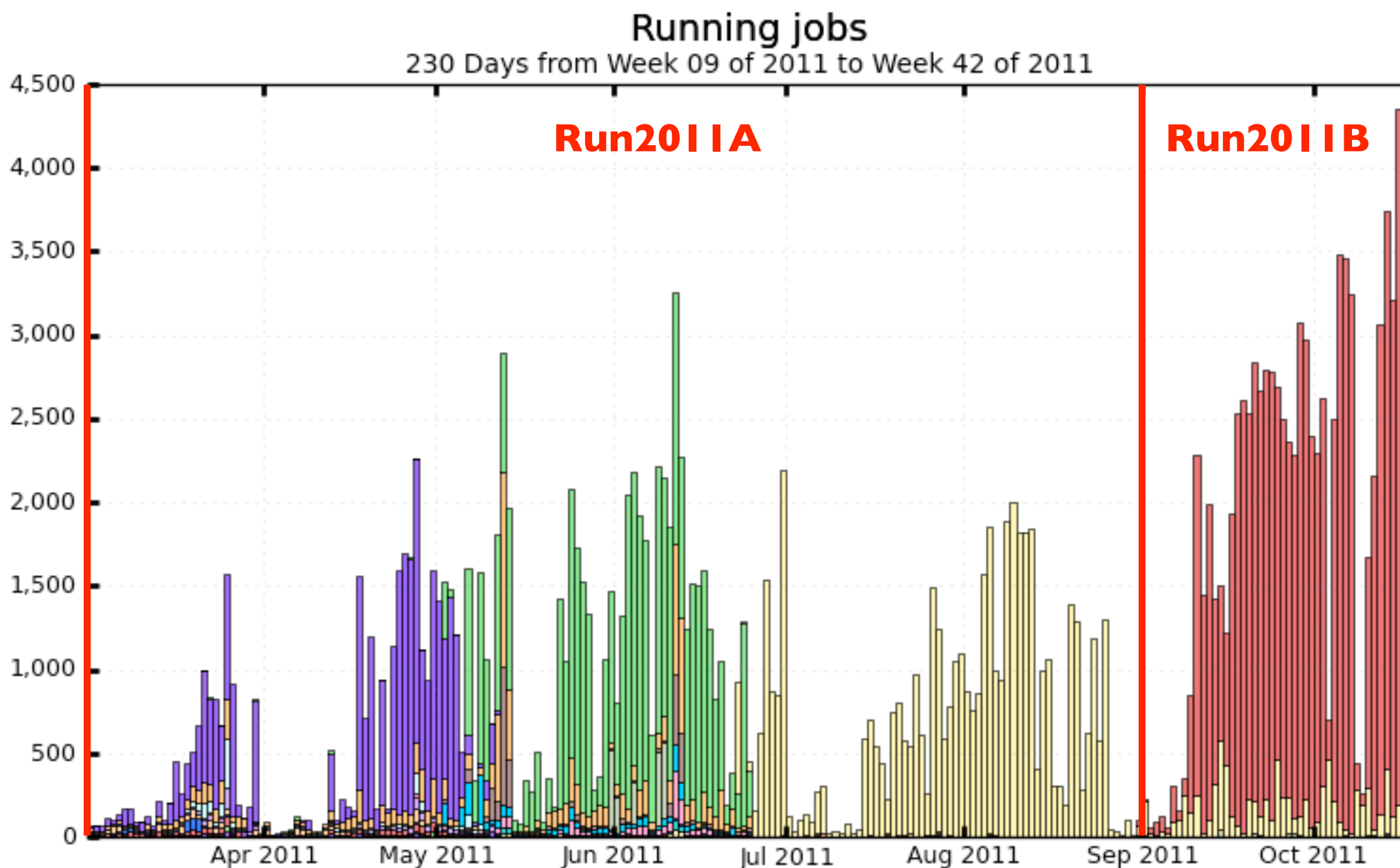
Data Processing Status

CMS Offline & Computing Week October 2011
17. October 2011

Oliver Gutsche
for
Data Operations

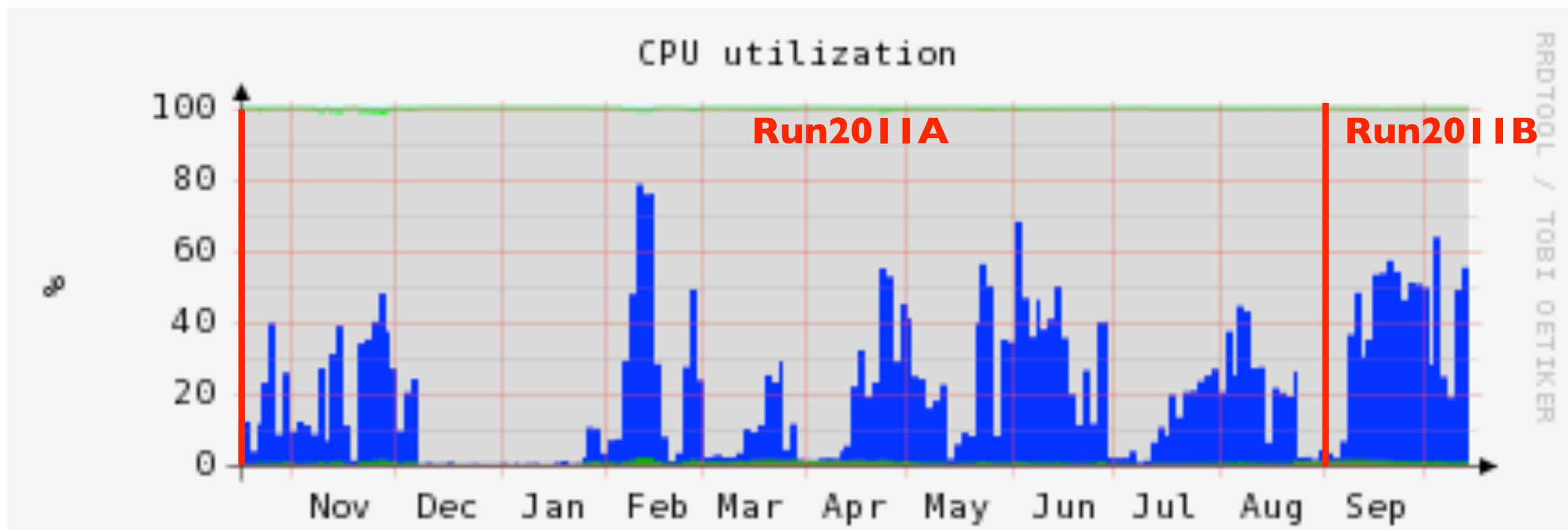


- ▶ Data Taking at the Tier-0 site
- ▶ Data & MC Processing at the Tier-1 sites
- ▶ MC production at the T1/T2/T3 sites
- ▶ Processing plans for the last months of 2011



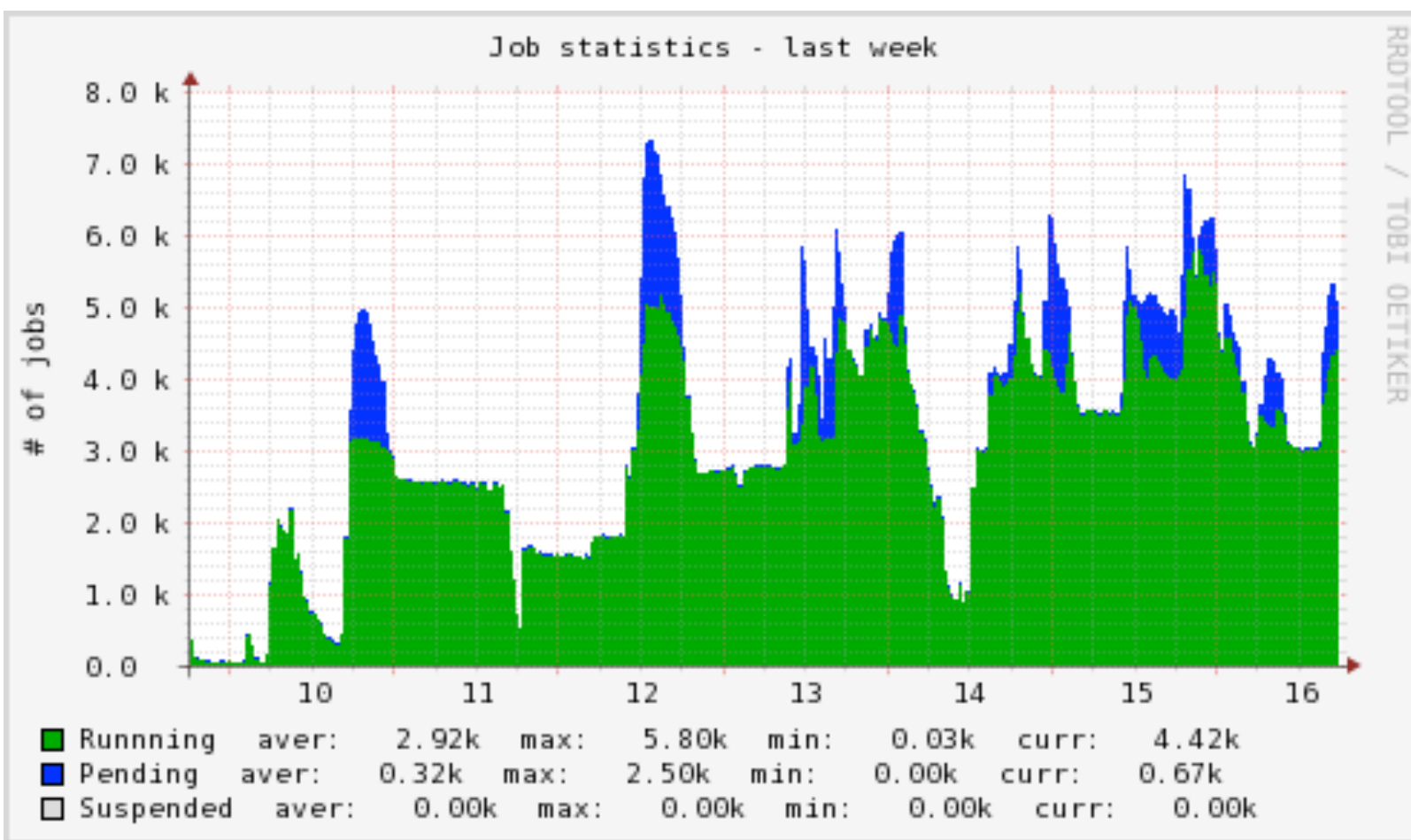
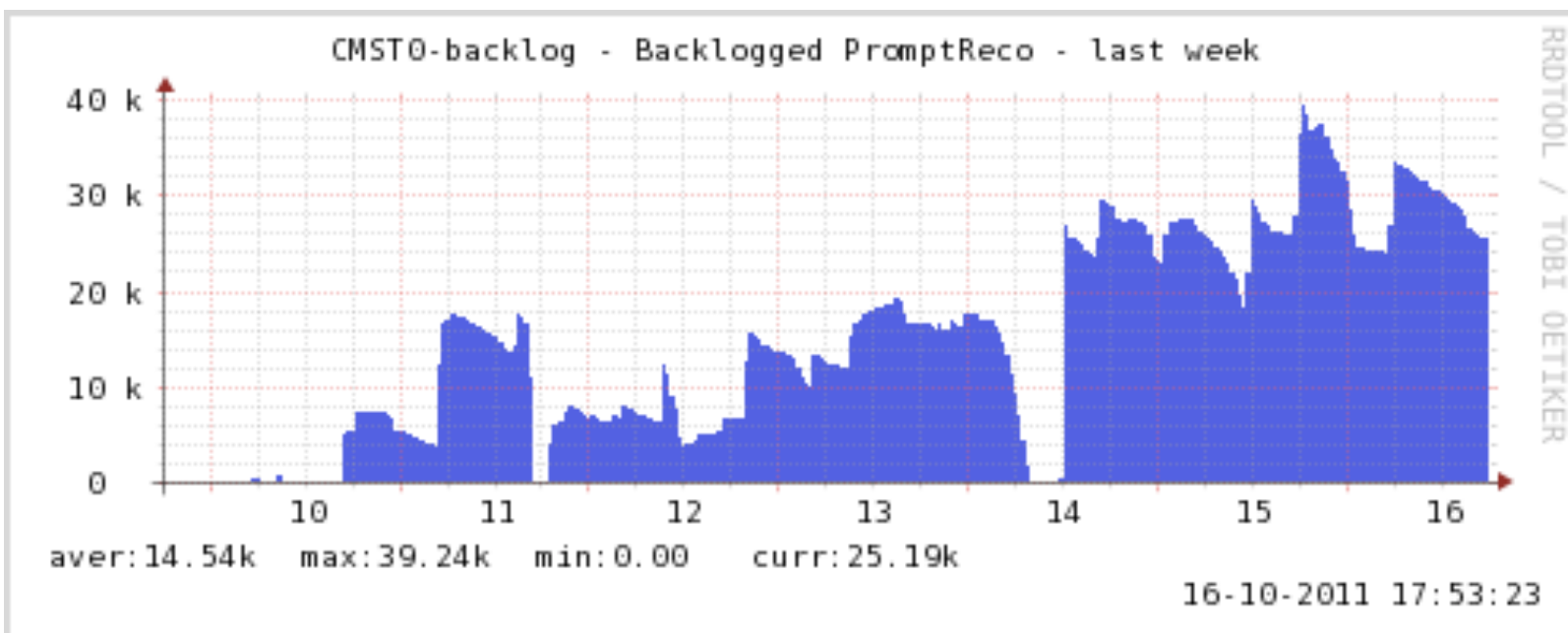
► **2nd half of Run2011A:** first signs of memory strain of CMSSW_4_2_X while PileUp is increasing, less running jobs possible

► **Run2011B:** more resources available to the Tier-0, manual overflow to public queues significantly helps to cope with the further increasing PileUp



- ▶ Changes in data taking conditions (PileUp) also visible in CPU utilization in the T0.
- ▶ **2nd half of Run2011A:** PileUp is increasing and CPU utilization is decreasing as more and more memory has to be reserved for the CMSSW_4_2_X jobs
- ▶ **Run2011B:** New hardware allowing 3 GB RSS per core allowing higher CPU utilization

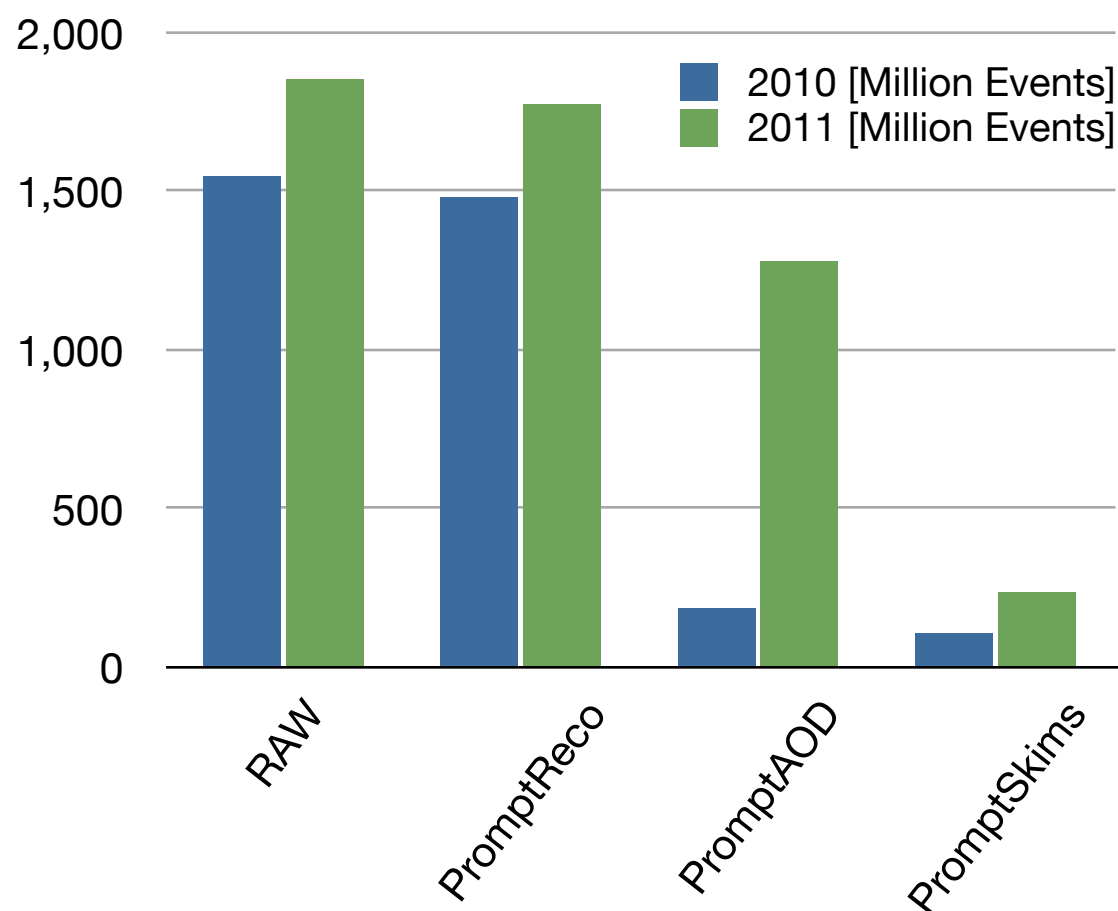
- ▶ Backlog in the beginning of the week at ~20k jobs when new runs pass the 48 hour prompt reco delay
- ▶ T0 caught up reasonably quickly using especially the manual overflow to the public queues (more than 5000 jobs running)
- ▶ End of the week, several very long fills with high data taking efficiency
- ▶ Backlog more than 20k jobs with new runs coming into the queue to keep the backlog stable
- ▶ Expect to catch up when the data taking situation was not as good 48 hours ago



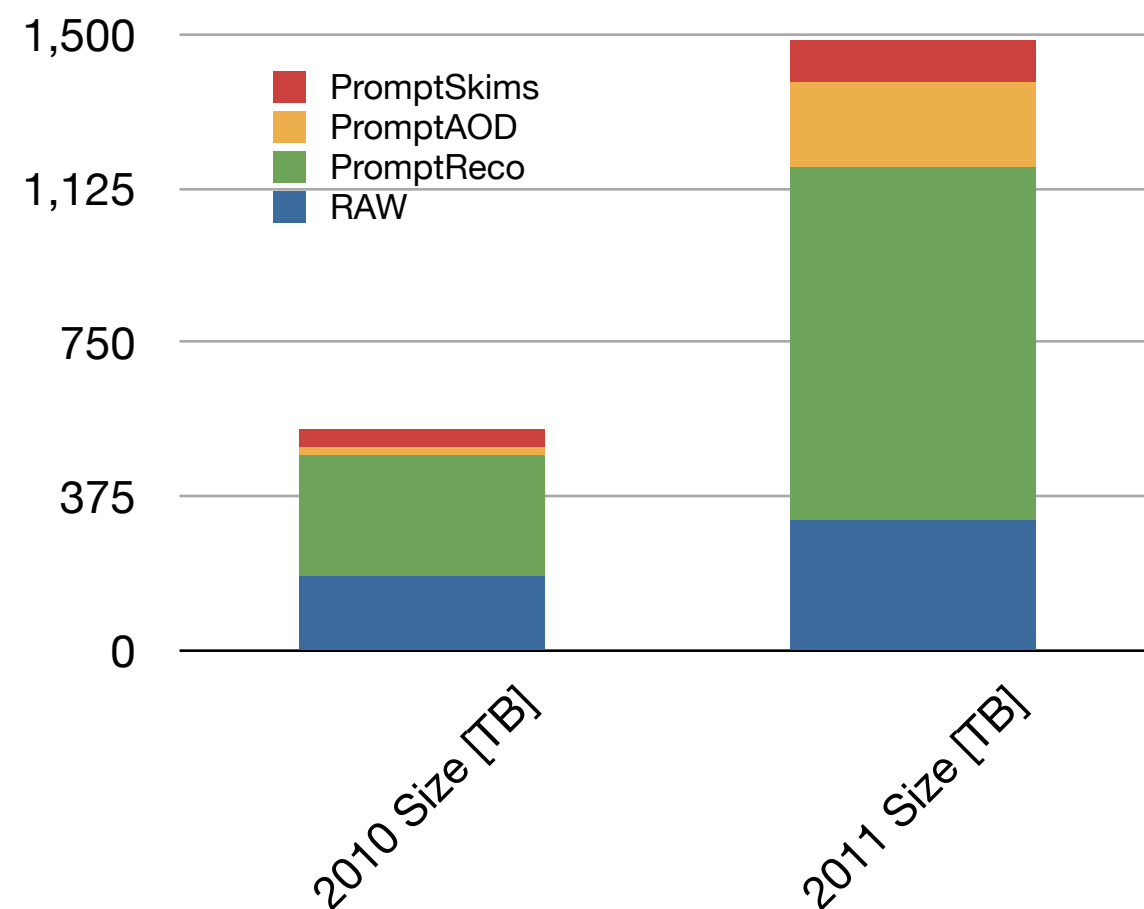
	RAW	PromptReco	PromptAOD	PromptSkims
2010 [Million Events]	1,546	1,482	184	107
2011 [Million Events]	1,852	1,774	1,280	236
Total [Million Events]	3,398	3,256	1,464	343

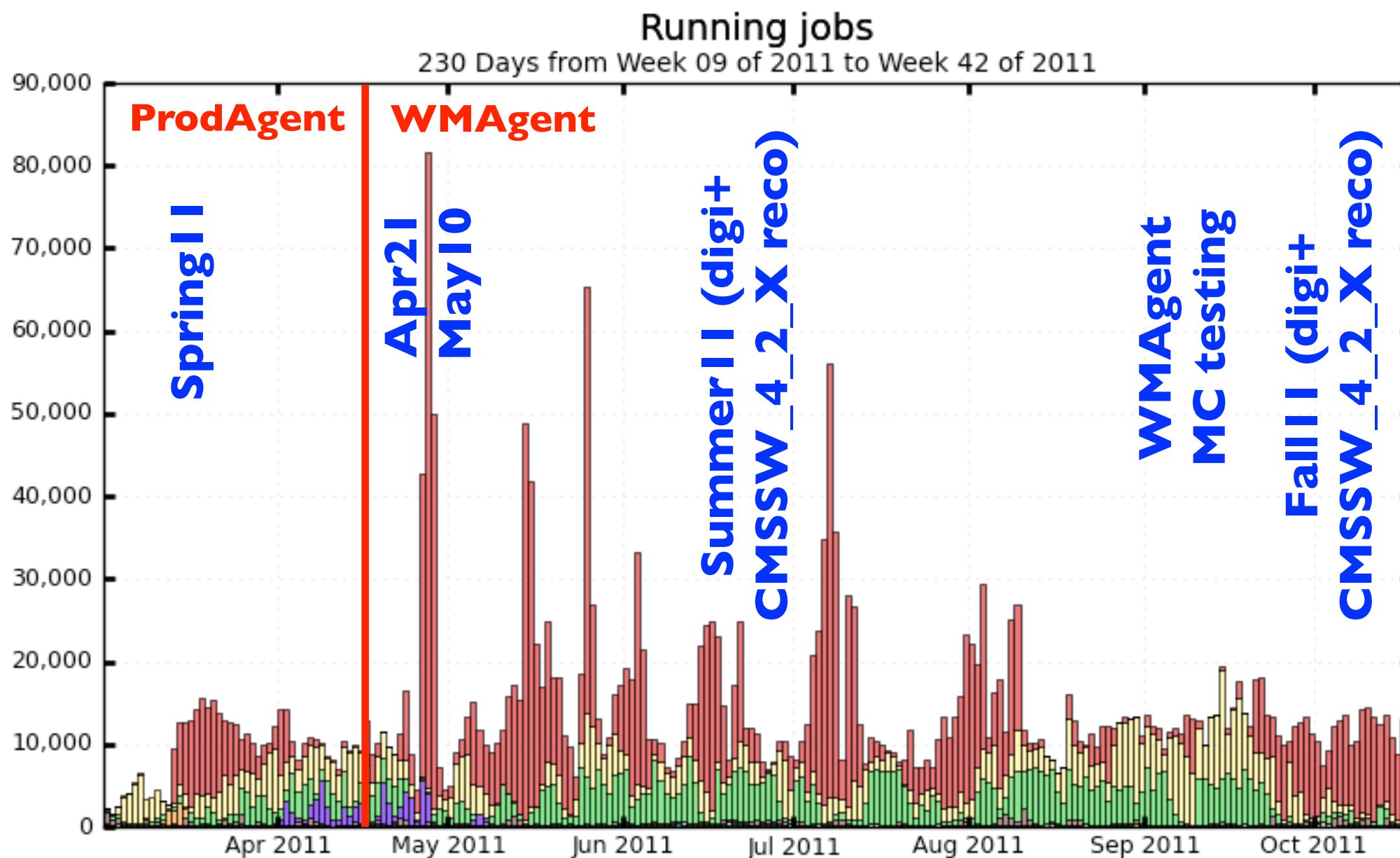
	RAW	PromptReco	PromptAOD	PromptSkims	Total
2010 Size [TB]	182	295	19	44	540
2011 Size [TB]	318	859	207	103	1,487
Total Size [TB]	500	1,154	226	147	2,027

Data: Events



Data: Size





▶ Over the year, put WMAgent into production

▶ Complete Summer I I digitization & reconstruction campaign in CMSSW_4_2_X adding PU_S4 pileup scenario

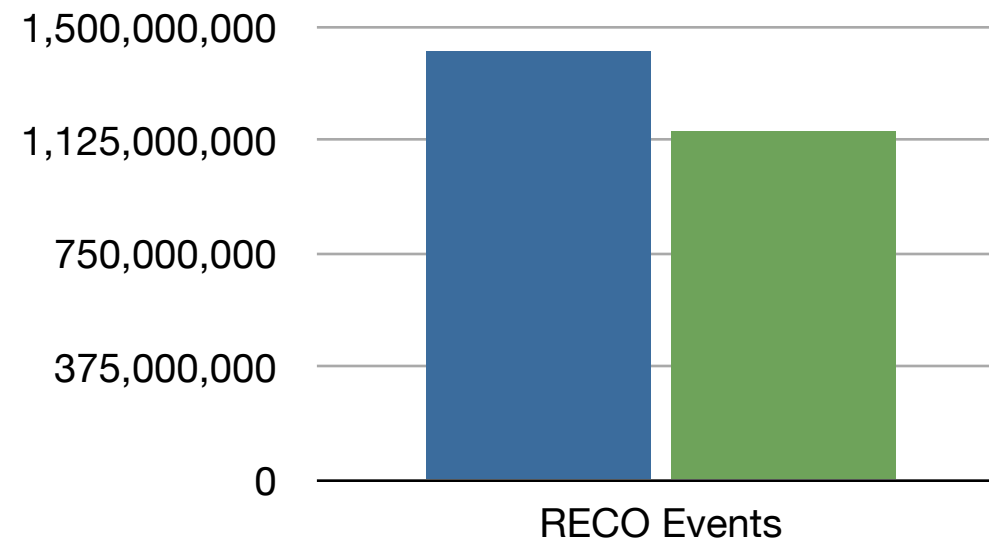
▶ Continued with high scale tests of MC production with the WMAgent till End of September

▶ **End of September: Fall I I**

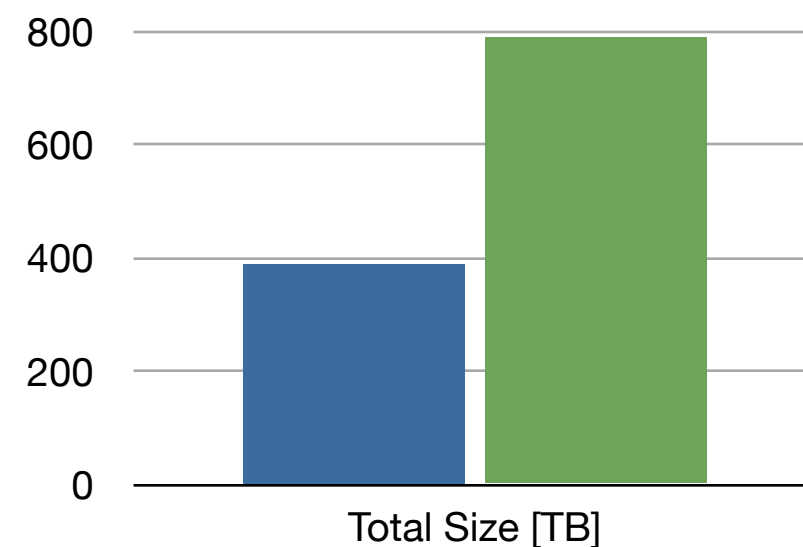
▶ started digitization of MC with PU_S6 pileup scenario (“high pileup”) in CMSSW_4_2_X

▶ Continued with reconstruction of high pileup MC with CMSSW_4_2_X

Era	Pass	RECO Events	Total Size [TB]
Run2010A+B	Apr21	1,420,077,332	388.89
Run2011A	Apr13	40,729,454	2.74
Run2011A	Apr22	10,767,224	9.59
Run2011A	May3	12,808,958	9.13
Run2011A	May7	29,085,527	21.57
Run2011A	May10	476,783,419	257.65
Run2011A	May13	18,604,900	5.59
Run2011A	16Jun	463,744	0.36
Run2011A	02Jul2011	144,560,884	130.85
Run2011A	05Jul2011	180,655,731	161.52
Run2011A	16Jul2011	56,168,211	8.55
Run2011A	05Aug2011	123,381,634	110.63
Run2011A	31Aug2011	273,081	0.65
Run2011A	03Oct2011	59,822,214	69.45
Total		2,574,182,313	1,177.16
Total Run2010A+B		1,420,077,332	388.89
Total Run2011A		1,154,104,981	788.28



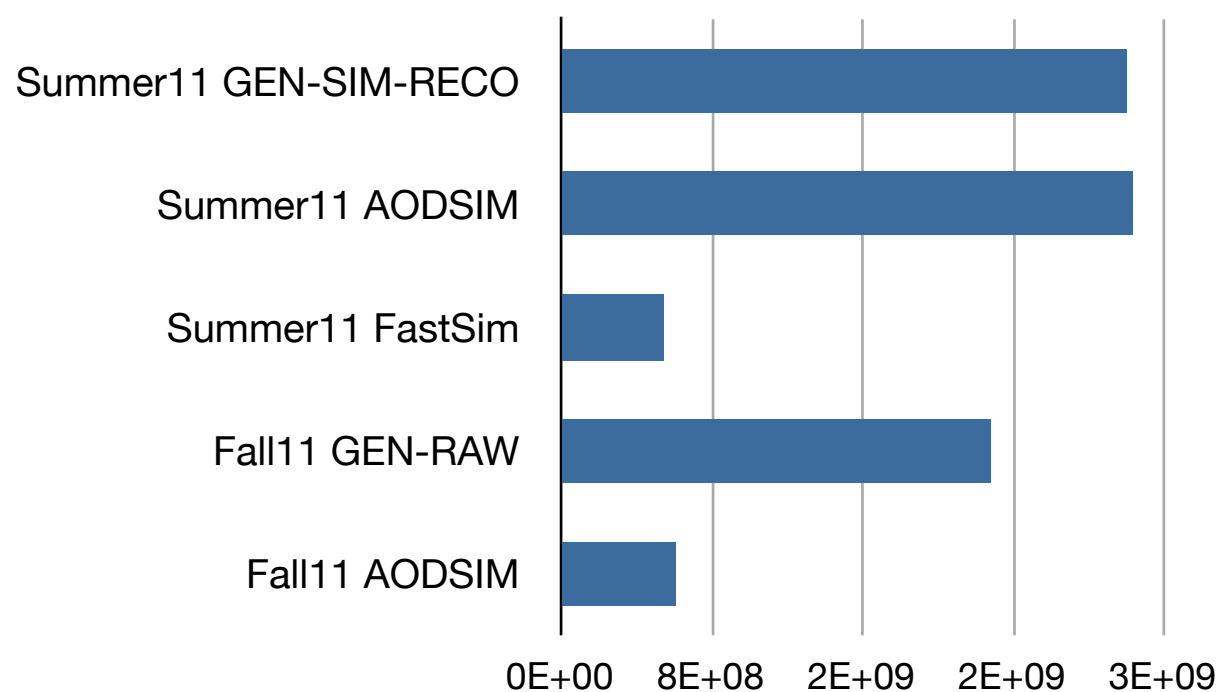
■ Total Run2010A+B ■ Total Run2011A



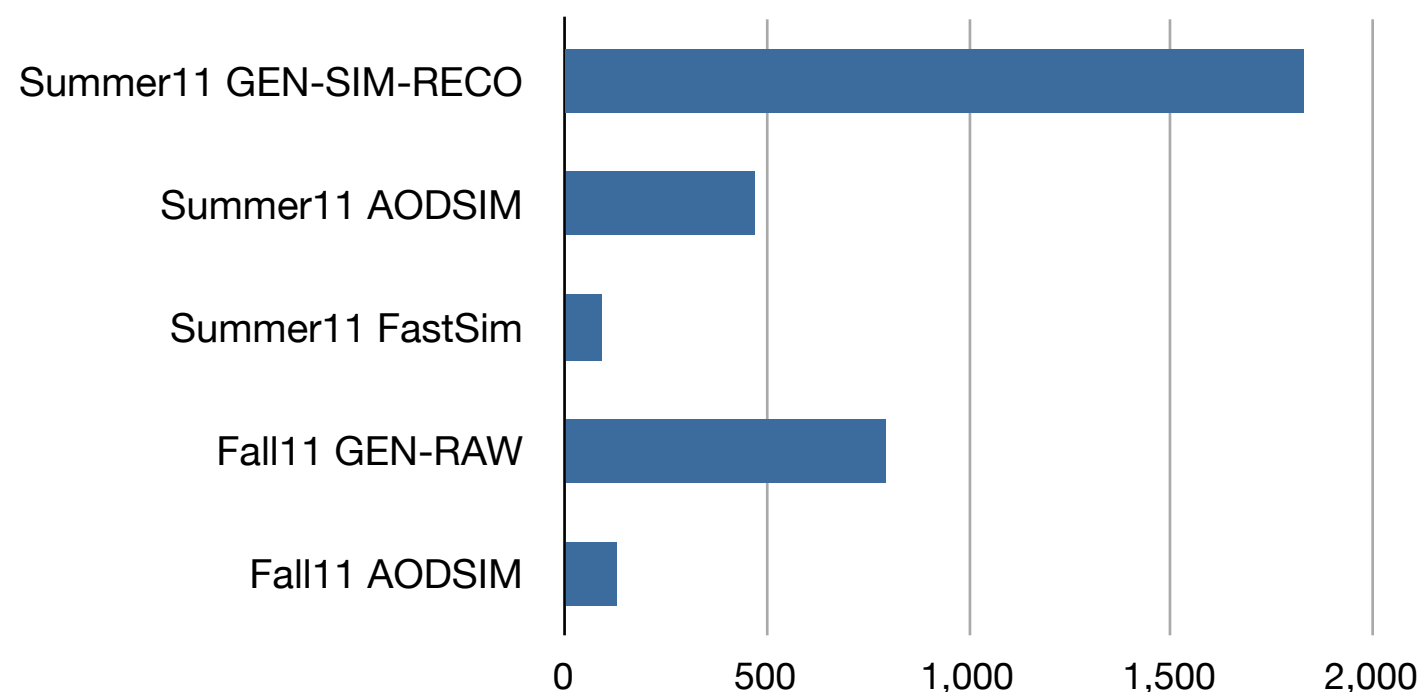
- ▶ Multiple smaller re-reconstruction passes not shown
- ▶ Number of events re-reconstructed approaching the total number of events in 2010
- ▶ Size of re-reconstruction passes in 2011 almost twice as big as for total amount of data in 2010

	Events	Size [TB]
Summer11 GEN-SIM-RECO	2,815,203,228	1,830.32
Summer11 AODSIM	2,846,026,209	469.89
Summer11 FastSim	512,499,216	90.83
Fall11 GEN-RAW	2,138,692,060	794.97
Fall11 AODSIM	572,084,724	129.17
Total	8,884,505,437	3,315.18

MC Overview: Events



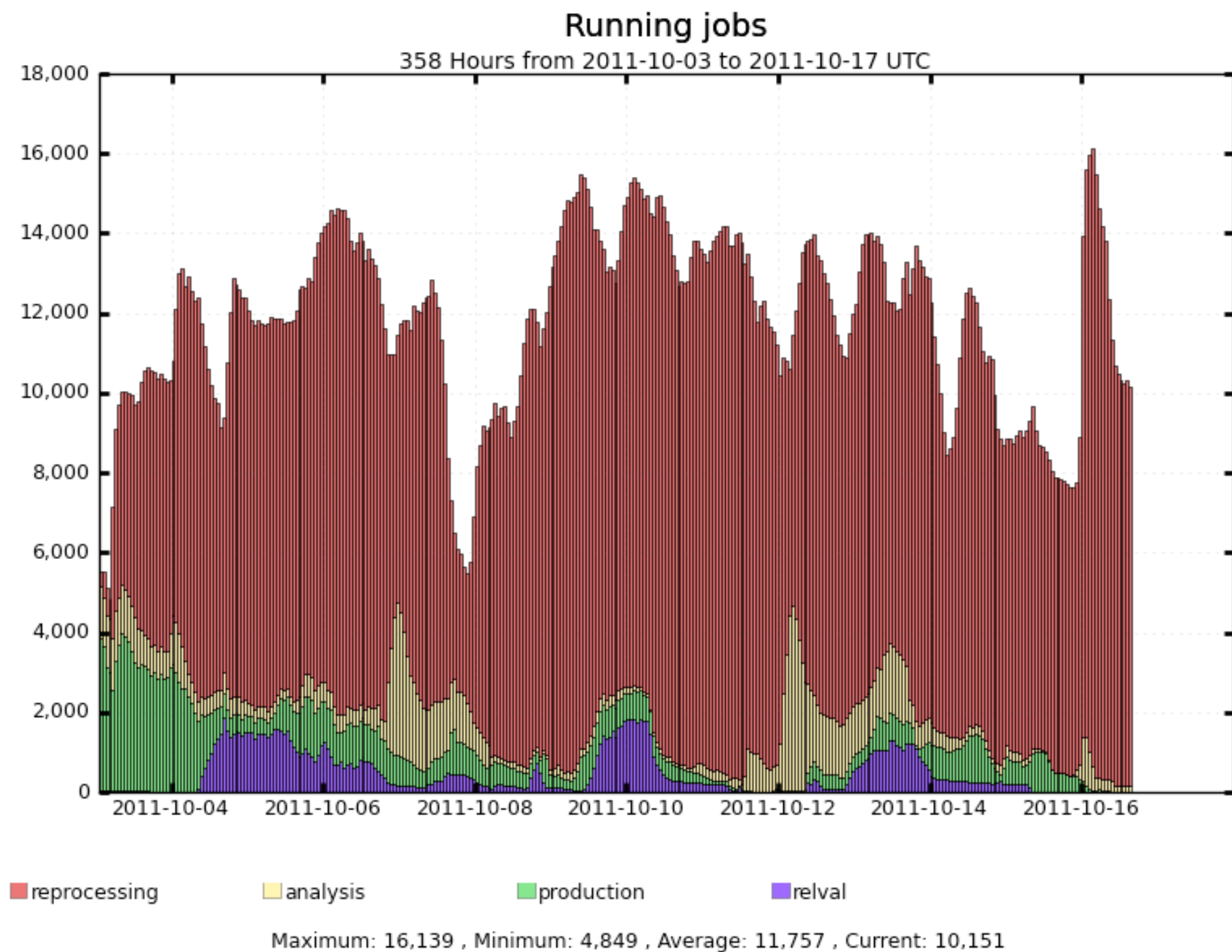
MC Overview: Size [TB]



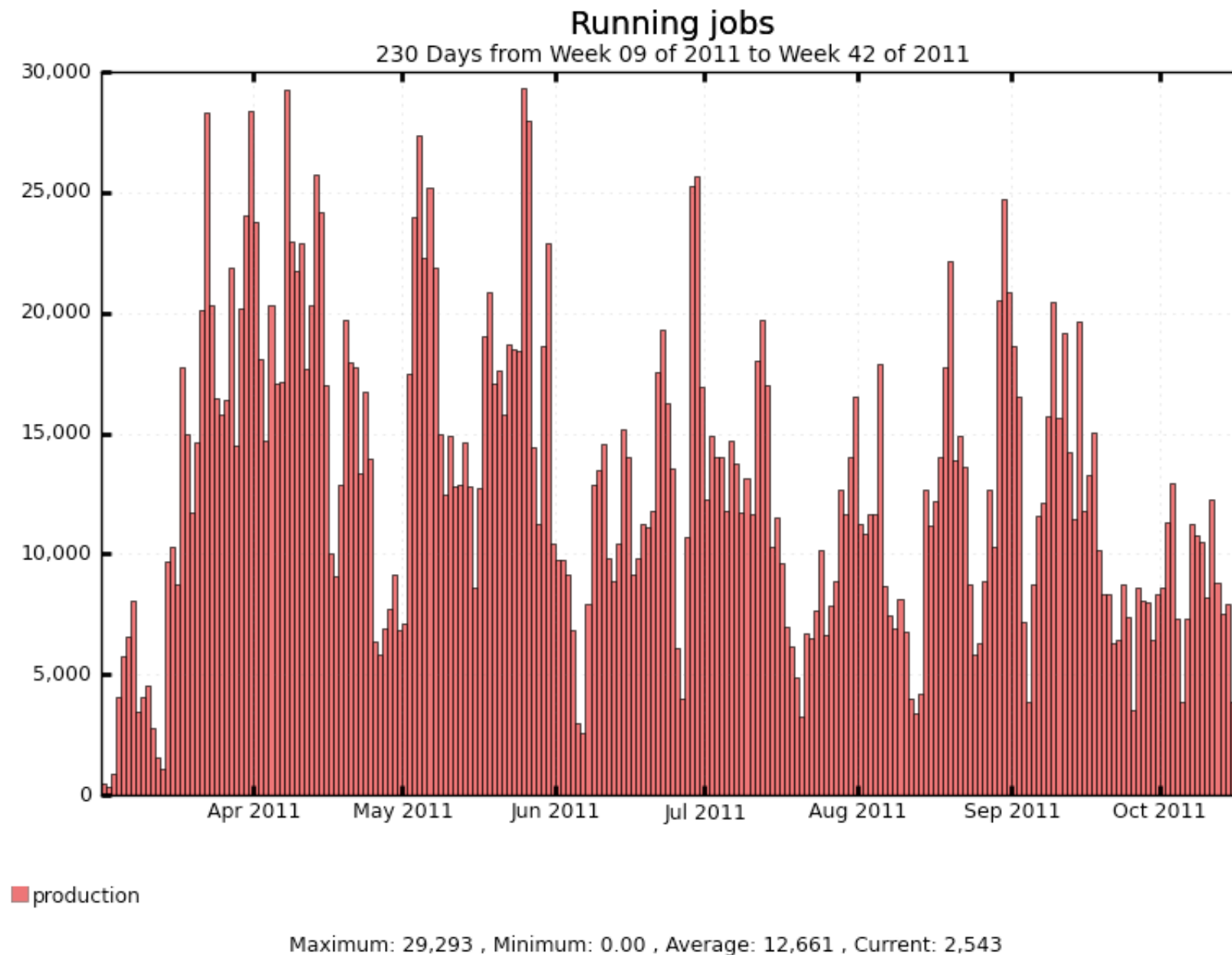
▶ Summer I I digitization/reconstruction keeps up with ongoing simulation

▶ Summer I I FastSim significant in number of events but small in total size => mostly SUSY scan grids

▶ Fall I I already has 2.1 Billion events digitized and 572 Million events reconstructed in CMSSW_4_2_X

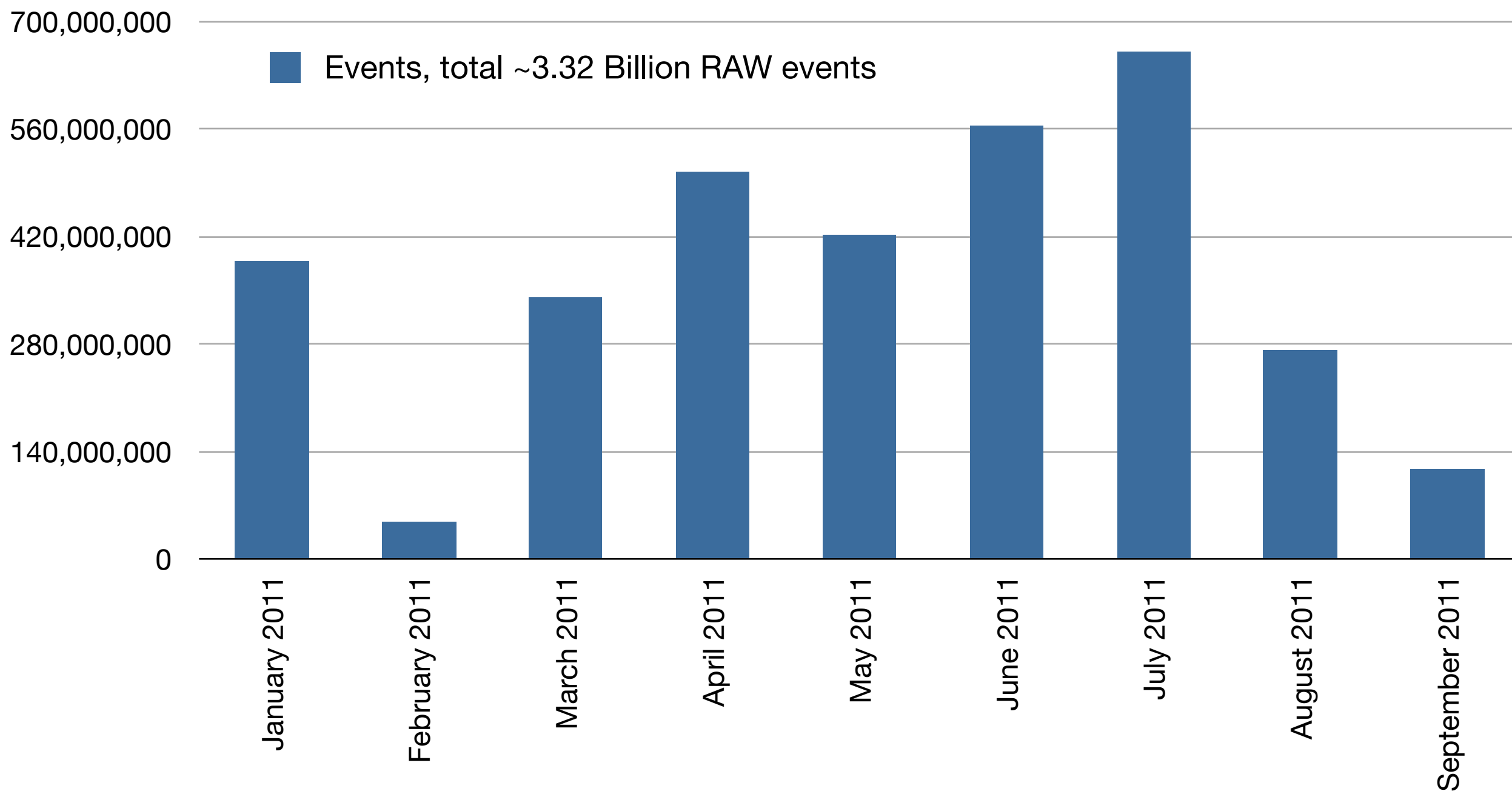


- ▶ Very good resource utilization taking into account that we have about 12k slots pledged on the Tier-1 level



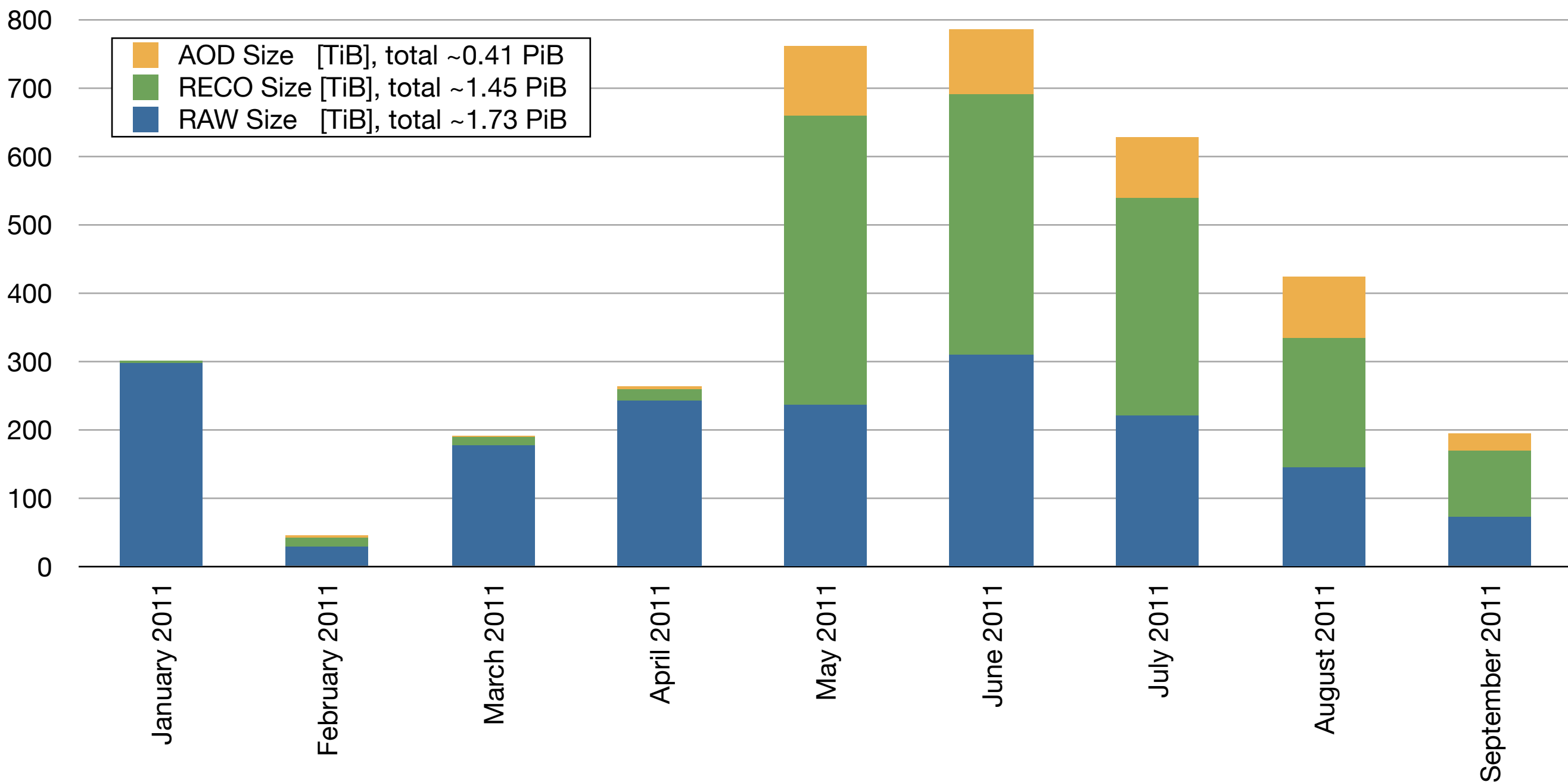
- ▶ Average of “average number of running jobs per 24h”: 12.6k jobs
- ▶ Includes high scale MC tests on the Tier-1 level in August/September
- ▶ Trend to less MC production as Summer I I simulation campaign is in its tails
- ▶ Also switching over to WMAgent based MC production

MC in 2011: Simulated Events per Month

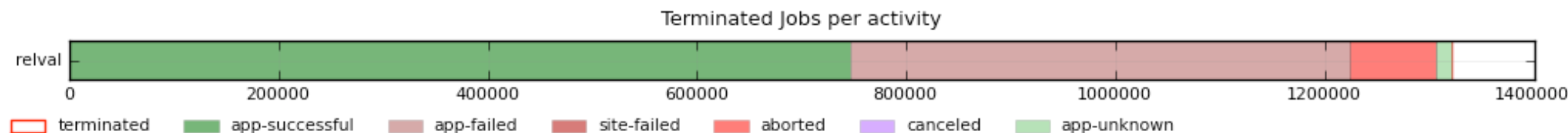


- ▶ Record month: July
- ▶ January/February: 8 TeV MC
- ▶ Summer I I simulation campaign: 2.7 Billion events

MC in 2011: Size in TiB per Month



► Summer I I GEN-SIM samples: 1.43 PB total



- ▶ Produced and processed since March 2011:
 - ▶ Number of events: 632 Million events
 - ▶ Total size: 143 TB

- ▶ Number of RelVal jobs:
 - ▶ Successful: 747 Thousand jobs
 - ▶ Application failed: 478 Thousand jobs
 - ▶ Aborted (usually memory problem): 97 Thousand jobs

- ▶ Processing plans for the last months of 2011
- ▶ Assuming 75% processing efficiency and more resources than pledged (experience value)
- ▶ Assumptions for processing times listed below

	Time per Event [s]	
MC: 4_2 reco	16	takes twice as long as 4_4
MC: 4_4 reco	8	plain average from Fall11 rereco tests in 4_4_0_pre8 from Fall11 test rereco
MC: 4_2 digi	6	plain average from Fall11 GEN-RAW tests for TOP: TOP-Fall11_R1-00024 (from Fall11)
Data: 4_4 reco of Run2011B	12	from prompt reco analysis from 111003
Data: 4_4 reco of Run2011A	6	twice as fast as Run2011B because of PileUp

	Total days
MC: 4_2 reco of 1 Billion events	19
MC: 4_4 reco of 1 Billion events	10
MC: 4_4 reco of 1.5 Billion events	15
Data: 4_4 reco of Run2011B	13
Data: 4_4 reco of Run2011A	6

Cutting it close to finish
by 20th of December

- ▶ Processing on Tier-0, Tier-1 and Tier-2 level very successful
- ▶ A big thanks to all especially all operators and coordinators and the developers
- ▶ We have planned a full queue for the remaining months of 2011
- ▶ Very important is that all the requests are submitted as soon as possible so that we avoid running out of work
- ▶ We're cutting it close with all the work planned.