CMS Computing Resource Request 2013/2014/2015

March 1, 2013

Introduction

The computing facilities for storing, processing, and analyzing the LHC data are the final step in a long series to fully realize the value of the LHC program. The resources available to CMS have been critical to the exciting analysis program demonstrated in CMS during the first run. For the purposes of planning we divide the operations period of the first run into resource years as shown in Table 1.

Year	Start	End	Live Seconds (pp)	Live Seconds (HI)
Resource Year 2013	April 2013	March 2014	0	0
Resource Year 2014	April 2014	March 2015	0	0
Resource Year 2015	April 2015	March 2016	5.2Ms Expected	0.7Ms Expected

Table 1: Contains definitions of LHC Resource Years and the total number of live seconds from the accelerator expected during each period.

The number of live seconds expected during each period for proton-proton (pp) and Heavy Ion (HI) running is given in the final 2 columns. 2013 and 2014 are the two years of Long Shutdown 1 (LS1). The long shutdown is a period with detector work and studies in preparation for higher energy running, but no new data. To make the best use of the shut down period, CMS collected more events than could be promptly reconstructed at the Tier-0 during 2012. The DAQ and High Level Trigger systems are capable of collecting data at higher rates than can be immediately reconstructed even using all the resources at CERN and all the time between fills. The data not immediately reconstructed was repacked into raw data and transferred to Tier-1 centers. This data is commonly referred to as “Parked Data”. This data is currently being reconstructed and is expected to be available to analysis users by the summer of 2013.

In this planning the computing requested in 2014 is essentially only a small increase for tape to archive new samples in preparation for the upgrades and samples to prepare for 2015. The absence of new data and a reasonably prompt analysis of the 2012 samples will allow CMS to do the preparatory work for 2015 with only small changes in the computing capacity.

Looking ahead to 2015 the situation changes. With the discovery of a light Higgs-like boson and interesting searches at the energy frontier, the initial studies from CMS indicate that we will need to substantially increase the trigger rate in order to maintain the physics capability of the experiment. The early studies indicate that somewhere between 800Hz and 1.2kHz will be needed. For the purposes of this document 2015 is just an early indication, but this year it is especially important for the sites to plan the two years together. 2014 contains only modest increases, but there are significant increases on the horizon for 2015. Sites may wish to spread the procurements across 2 years, or hold hardware procurements until the second year to save money.

In time for the October RRB in 2013 there will be a revised estimate for 2015 that will include new estimates taking into account any model or code improvements. The current 2015 estimates are based on some operational improvements motivated by the success of the first run. The document below concentrates a lot on the 2015 year, even though it is just an early indication.

Experience from 2012

In 2012 the machine live time exceeded our planning by 25%. Part of this was the extended run. The instantaneous luminosity at the end of the run and CMS trigger rate were both high in the second half of the year reaching 7E33 and 1 kHz (400 Hz prompt and 600 Hz parked) respectively. The total integrated luminosity was nearly 22fb-1 recorded.

- The Tier-0 center
  - The Tier-0 was capable of using the period between fills more completely than we had planned. A system for spilling Tier-0 jobs into the public queues at CERN was commissioned and heavily used. When the machine was able to refill quickly the Tier-0 would fall behind and we would only catch up during longer refilling periods, machine development, or machine problems. The infrastructure was sufficiently stable to handle this mode of running.
  - The Tier-0 load was higher also due to the need to repack the parked data samples and transfer to Tier-1s. The estimate that the Tier-0 could handle the additional throughput as long as it was not promptly reconstructed proved accurate.
  - The Tier-0 was also used for data reprocessing during periods when the load was not as high. This added to the offline processing capacity.
- The CAF
  - Performed commissioning activities including validation, alignment and calibration activities. The overall average utilization of the CAF remained low, and about half the processing capacity was moved to Tier-0
- The Tier-1 Centers
  - Accepted the data divided into primary datasets from the Tier-0.
  - Served data to requesting Tier-2 facilities.

- Reprocessed the entrusted samples with new software versions and updated alignment and calibration constants. During 2012 CMS changed software versions early at the time of ICHEP and expects to stay as the production release for one year. The prompt reconstruction was sufficiently good for conference analyses except for some dedicated channels that used small scale reprocessing samples. The data reprocessing was concentrated on one major reprocessing at the end of the year. Reducing software version changes and extra reprocessing passes allowed CMS to concentrate on simulation, which allowed CMS to complete a large sample in time for the 2013 winter conferences.
- Performed simulated event production and their digitization including addition of PileUp events and reconstruction.
- Accepted simulation data from Tier-2 centers for custodial storage.

- The Tier-2 Centers
  - Provided analysis resources to the physics community.
  - Performed simulated event production.

The pile-up increased in September to a larger value than we anticipated in the planning, which led to a higher event complexity for the final 4 months of running. In computing this manifested itself with longer reconstruction times.

Changes in the operating model for 2014

In 2014 we expect primarily to be reprocessing data samples for legacy archive and preparing simulation samples for upgrade and 2015 preparation work. The higher pile-up expected in 2015 is more difficult to simulate and requires more computing resources. We anticipate changing the setup of the Tier-0 to be similar to that of the other Tier-1s. We will submit more requests through the Tier-0 grid interfaces using pilots, and we will activate more network links to Tier-2s.

We anticipate having access to the CMS high level trigger farm (HLT) for the bulk of the year. This resource augments the Tier-1 data processing capacity by about 40%. This new resource is one of the reasons the CPU increase requested in 2014 is very modest.

Changes in the operation model for 2015

In 2015 there are 3 main factors that drive the need to increase computing resources. The first is that we expect an increase in the number of pile-up events that with the current code would require a factor of 2.5 increase in reconstruction time to process. The second is that the trigger rate expected to grow a factor of 2.5 higher, and the third is that currently the code reconstruction speed depends on out-of-time pile-up in the tracker. Going to 25ns running increases the reconstruction time for the same number of pile-up events by almost a factor of 2. With no changes in the way the experiment works, we would require a factor of 12 (2.5 * 2.5 *2) increase in the processing resources to maintain the current activities. Such a large increase is impossible in a constrained budget, so changes are being planned.

For the purposes of planning 2015, we assume the issue of out-of-time pile-up will be solved. The reconstruction of the tracker was not anticipated to be so strongly dependent on out-of-time pile-up. The other 2 factors are more challenging. We intend to move a substantial fraction of the Tier-0 prompt reconstruction to Tier-1. This is a procedure that we validated in order to promptly complete any 2012 parked samples if necessary. This will reduce the resources available for reprocessing and simulation at the Tier-1s. We have assumed that the organization level of the experiment achieved in 2012 will be maintained in 2015, which will allow fewer reprocessing passes. We also anticipate needing to reduce the fraction of simulated events produced for data events collected. We are working on commissioning the HLT farm for use for offline processing during the annual shutdown. The processing campaign at the end of the year lines up nicely with the availability of the HLT.

Even with these changes we anticipate needing roughly a factor of two increase in processing resources between 2012 and 2015.

Resource Parameters

The most important input parameters by year for the resource calculations are given in Table 2. The expected average trigger rate for prompt reconstruction in CMS is approximately 1000Hz starting in 2015, with an initial overlap between primary datasets of 25%. We do not expect to park any data in 2015, as there is no time to process it before the second long shutdown.

Parameter	**Year**					
	2010	2011	2012	2013	2014	2015
Trigger(Hz)	200Hz-600Hz	300Hz	300Hz	0	0	1000Hz
PDS Overlap Factor	1.3	1.25	1.25	NA	NA	1.25
Parked Data	0	0	600Hz Peak (~400Hz Average)	0	0	0
Tier-0 Recovery	0.75	0.75	0.5	NA	NA	0.50

Table 2: The driving yearly parameters of the CMS Computing

Two pile-up scenarios were used in 2015 for planning: 20 and 35 pile-up interactions. This is the average value expected, which corresponds to a higher peak except in some

leveling scenarios. A ramp up is foreseen, but at both 25ns and 50ns running both pile-up scenarios are possible.

Parameter	Expected Pile-Up	
	20	35
RAW Event Size Data (MB)	0.55	0.85
RAW Event Size MC (MB)	1.5	1.5
RECO Event Size Data (MB)	0.7	0.85
RECO Event Size MC (MB)	0.75	0.90
AOD Event Size Data (MB)	0.28	0.38
AOD Event Size MC (MB)	0.33	0.43
Repacker Time (HS06s)	6	7
RECO Time Data (HS06s)	230	500
Gen-Sim Time MC (HS06s)	500	500
Re-digi/Re-RECO Time MC (HS06s)	300	600

Table 3: The event parameters that vary as a function of the number of pile-up interactions per crossing

The processing times for reconstruction and simulation in HepSpec06 units increase with increasing instantaneous pile-up. At very high number of pile-up events the reconstruction time grows non-linearly.

Tier-0 Request

The Tier-0 computing requests for proton-proton running are shown in Table 4.
The processing resource request for CERN does not increase in 2013 and 2014, but the CPU resources of both the Tier-0 and the CAF are assumed to be available for analysis and simulation. These are included in the Tier-0 table and zeroed in the CAF table. This helps to alleviate resource shortages at the Tier-2s by the end of 2013 when the integrated data sample is the largest and in 2014 when there is a need for analysis of old data and simulation and studies for the 2015 run.

Tier-0			
Year	**2013**	**2014**	**2015**
CPU (kHS06)			
Express	0	0	17
Prompt RECO	0	0	210
Repack	0	0	8
Alca Workflow	0	0	6
VOBoxes	12	12	15
Analysis/Simulation	109	109	0
Total	121	121	256
Disk (TB)			
Tier-0 Streamer Pool	0	0	1000
Tier-0 Input Buffer	0	0	500
Tier-0 Export Buffer	0	0	1500

Tier-0			
Year	**2013**	**2014**	**2015**
Tier-0 Production Space	0	0	250
Analysis Disk	7000 (Repurposed CAF disk)	7000 (Repurposed CAF disk)	0
Total T0 Disk (TB)	0	0	3250
Tape (TB)			
Total Tier-0 Volume of RAW pp on tape	7000	7000	13000
Total Tier-0 Volume of RECO pp on tape	12000	12000	17000
Total Tier-0 Volume of AlcaRECO pp on tape	1000	1000	1000
Analysis/Simulation	6000	6000	7000
Total	26000	26000	38000

Table 4: The processing and storage requests for the CERN Tier-0 are shown as a function of year.

CAF Request

The CAF was foreseen at the beginning of the run as a resource for calibration, validation, and prompt analysis. Over the first two years of running several elements of low latency calibration have been added into the automated prompt calibration loop performed during the delay before launching prompt reconstruction. Much of the validation and prompt analysis can be performed at Tier-2 centers. The end result is that the average utilization of the CAF is lower than planned. With the high load expected of Tier-0 prompt reconstruction the CAF resources were reduced in 2012 and reallocated to the Tier-0. This is a significant reduction and missing resources will need to be supplemented with Tier-2 computing.

In 2013 and 2014 all the resources (processing and storage) available at CAF are made available as analysis resources in combination with the bulk of the Tier-0 processing resources. This makes efficient use of these resources during LS1, and offsets some of the needed increases at Tier-2s.

CAF			
Year	**2013**	**2014**	**2015**
CPU (kHS06)			
CAF Processing	0	0	13
CAF Interactive	0	0	1
Total	0	0	14
Disk (TB)			
CAF Express Data Volume (TB)	0	0	2500
CAF Prompt RECO Data Volume (TB)	0	0	2,000
CAF MC Volume (TB)	0	0	1,000
CAF Data AOD			4,000
CAF RelVal Volume (TB)	0	0	300
CAF Tier-2	0	0	1500
Stager Pool	0	0	800

CAF			
Year	**2013**	**2014**	**2015**
Total	0 (Moved to Tier-0 for analysis)	0 (Moved to Tier-0 for analysis)	12100
Tape (TB)			
CAF Tape	0	0	4000

Table 5: The processing and storage requests for the CERN CAF are shown as a function of year.

Tier-1 Request

The Tier-1 request is shown in Table 6 for 2013, 2014, and 2015. The Tier-1 processing resources are driven by the reconstruction time, the total volume of data, and the time allocated to complete a processing pass.

Tier-1			
Year	**2013**	**2014**	**2015**
CPU (kHS06)			
Processing	165	165	325
Disk (TB)			
Disk Space RAW Data	3500	3500	3500
Disk Space RECO Data	2000	2000	2000
Disk Space AOD Data	7000	7000	9000
Disk Space RAW MC	1500	1500	1500
Disk Space RECO MC	1500	1500	1500
Disk Space AOD MC	5000	5000	6000

Tier-1			
Year	**2013**	**2014**	**2015**
Disk Space Skimming	3000	3000	5000
Tier-1 Temp Disk	2500	2500	2500
Total	26000	26000	30000
Tape (TB)			
Total RAW Data	5000	5000	9000
Total RAW Simulation	14000	16000	18000
Total RECO Data	8000	8000	13000
Total Skimmed Data	2000	2000	3000
Total RECO MC	8000	9000	12000
Total AOD Data	7000	8000	12000
Total AOD MC	6000	7500	12500
Tape Total	50000	55500	79500

Table 6: The processing and storage requests for the total Tier-1 centers are shown as a function of year.

The CMS computing model assumes that one copy of the current version of the reconstructed data and the current year's raw data is kept on disk, with 10% of 2 copies of preceding versions of the RECO. Additionally 10% of all simulated events are kept on disk. When simulated events are needed for reprocessing or transfer to Tier-2 centers they are generally expected to be staged from tape. If comparisons of large samples of old RECO events are needed, they too will need to be staged in an organized way. CMS can trigger data to be staged either through the SRM interface or through local scripts. CMS is currently engaged with the Tier-1s to further separate active disk and archival tape. This is expected to improve the management of the disk resources.

In addition to separating disk and tape, CMS has implemented a data federation based on Xrootd. This allows sites to fall back to wide area distribution if a file is not found

locally, and it allows the bulk of the experiment data to be read by remote applications for limited use cases.

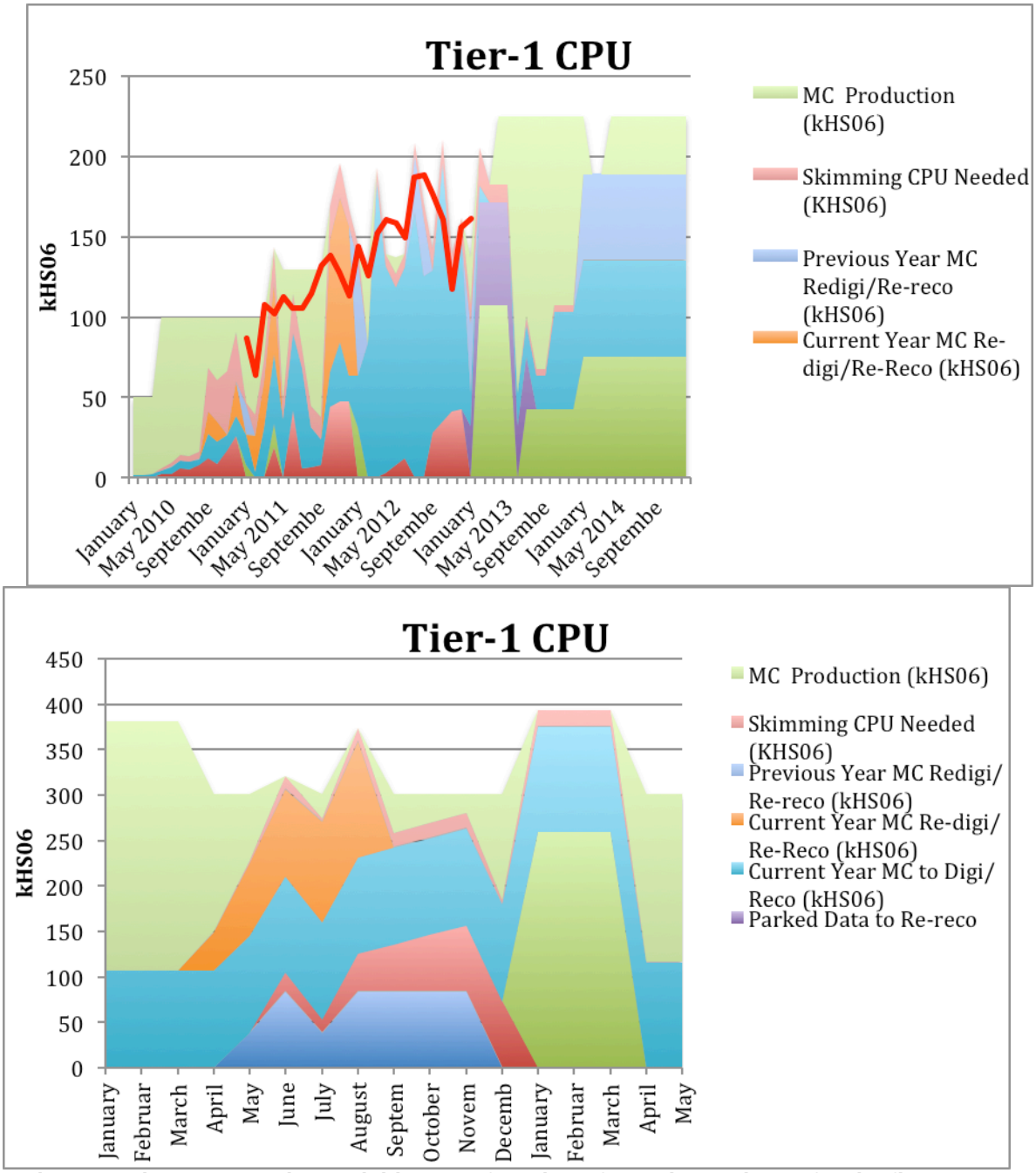

Figure 1: Tier-1 processing activities as a function of month are shown for the first run to 2014, and 2015. The peaks are the main processing passes of the simulation and data concentrated before major conferences. The orange between is simulated event production capacity.

Tier-2 Request

The Tier-2 resource request can be seen in table 7. The resource needs for Tier-2 do not grow as substantially in 2015 as the Tier-0 or Tier-1 resources primarily due to the presence of a single year of data. In the planning we assume the majority of the analysis on the first run have been completed by 2015 and there is not a need to host or analysis as many of the datasets from the first 3 years. We calculate the resources for the Tier-2s increase more in 2016 with 2 years of data to analyze.

Tier-2			
Year	**2013**	**2014**	**2015**
CPU (kHS06)			
Analysis	250	300	400
MC Production	100	100	100
Total	350	400	500
Disk (TB)			
RECO Data on Disk	1000	1000	1000
AOD Data on Disk	8000	8000	8000
RECO MC on Disk	2000	2000	2000
AOD MC on Disk	8000	8000	10000
User Space On Tier-2s	4800	4800	8400
Production Space on Tier-2s	2000	2000	2000
Total	25800	25800	31400

Table 7: The processing and storage requests for the total Tier-2 centers are shown as a function of year.

The total amount of processing and storage resources needed for analysis scale strongly with the transition from RECO to AOD for analysis. In 2010 CMS did not write the AOD until well into the run. We are assuming that in 2015 even with a new energy that the majority of users analyze from the AOD. The total Tier-2 analysis needs also

scale with the number of events collected. Figure 2 shows the storage evolution for the Tier-2s.

Explicitly in 2013 and 2014, CMS will use the Tier-0 and CAF processing and disk resources for analysis work when they are not needed for data collection. The simulation and analysis processing and storage resources at the Tier-2s have been calculated assuming the resources freed at CERN are available.

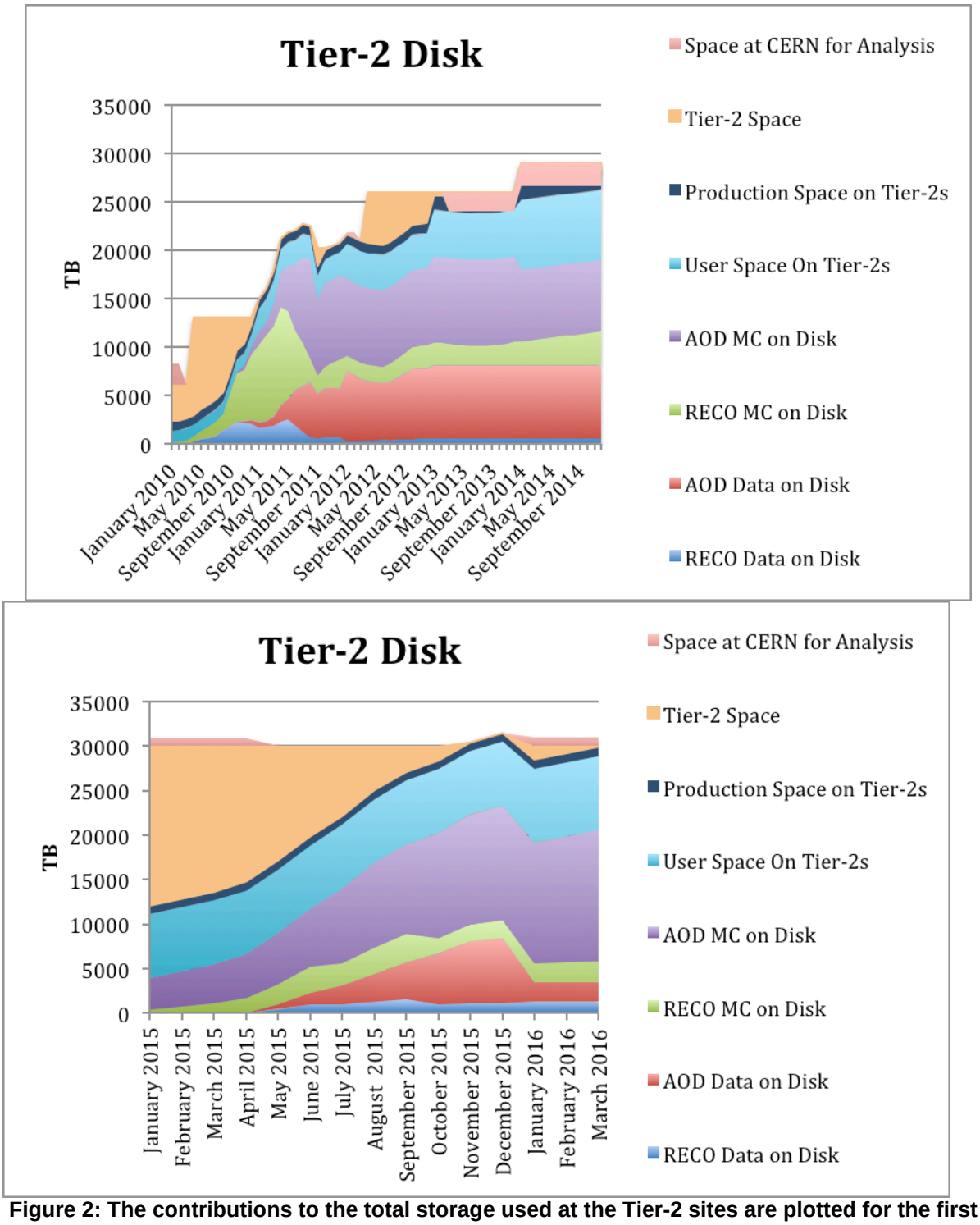

Figure 2: The contributions to the total storage used at the Tier-2 sites are plotted for the first run up to 2014 and 2015. The green bars are the contribution from RECO events, the proportion of which increase at the beginning when most of the analysis is on RECO files and decrease as the experiment transitions to AOD.

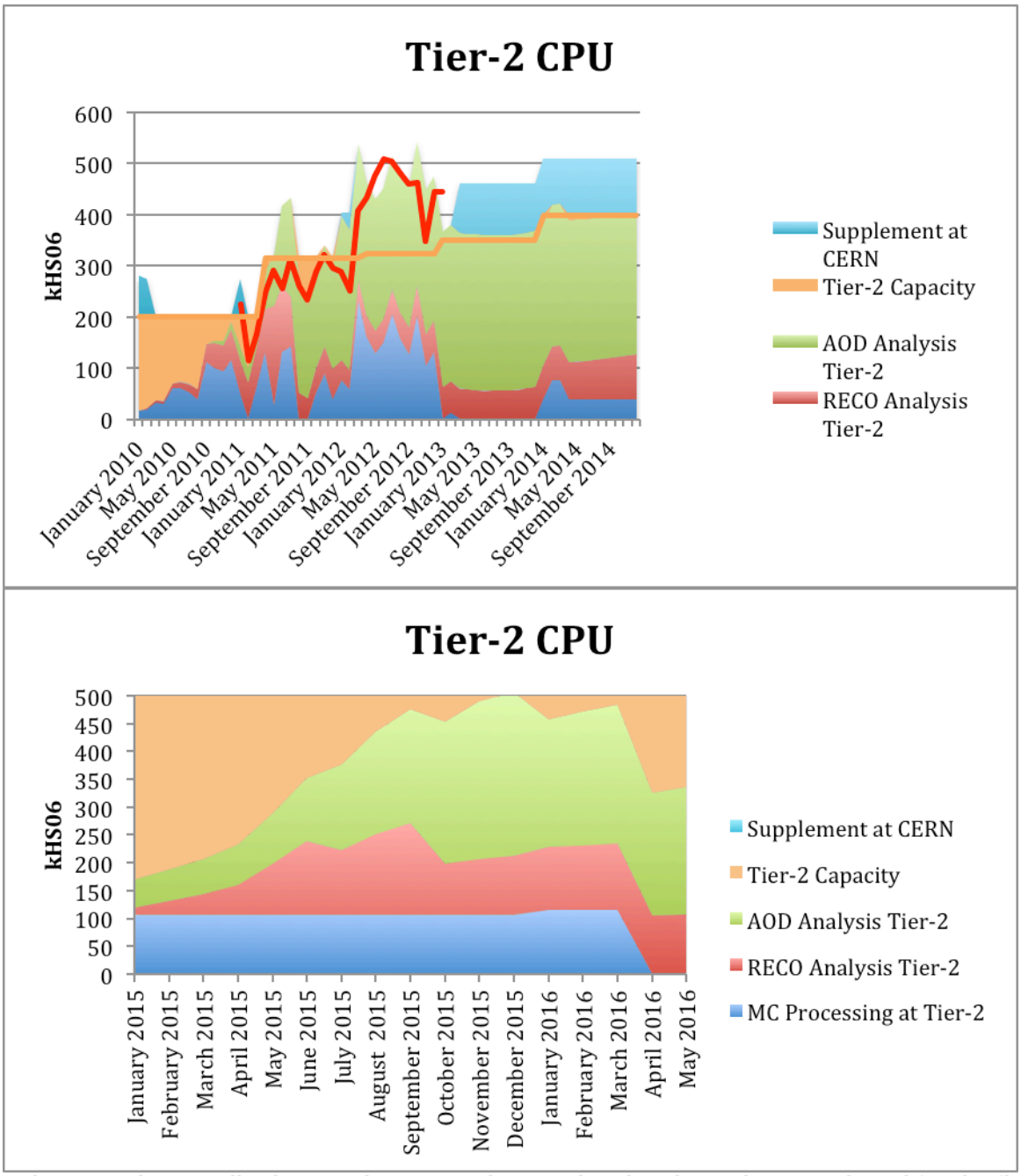

Figure 3: The contributions to the processing used at the Tier-2 sites are plotted for the first run up to 2014 and 2015. The green bars are the contribution from RECO events, the proportion of which increase at the beginning when most of the analysis is on RECO files and decrease as the experiment transitions to AOD.

Summary

The total processing, disk, and tape request for 2013, 2014 and 2015 resource years are shown in Table 8. The resource request maintains the fundamentals of the computing model and combines with our best understanding from the operational experience we have with collision data. The request in 2014 is very small, but CMS will request a substantial increase with the next run in 2015.

	2013	Increase from	2014	Increase from	2015	Increase from
Tier-0 CPU (kHS06)	121	0%	121	0%	256	111%
Tier-0 Disk (TB)	7000	0%	7000	0%	3250	-53%
Tier-0 Tape (TB)	26000	0%	26000	0%	38000	46%
CAF CPU (kHS06)	0	0%	0	0%	12	
CAF Disk (TB)	0	0%	0	0%	12100	
T1 CPU (kHS06)	165	12%	165	0%	325	96%
T1 Disk (TB)	26000	0%	26000	0%	30000	15%
T1 Tape (TB)	50000	11%	55500	11%	79500	43%
T2 CPU (kHS06)	350	8%	400	14%		25%
T2 Disk (TB)	26000	0%	27000	4%	31400	16%

Table 8: Shows the processing, disk, and tape resources requested by CMS for all centrally controlled computing tiers.