CERN–RRB–2018–028
April 2018

Computing Resources Scrutiny Group

C Allton (UK), V Breton (France), G Cancio (CERN), A Connolly (USA), M Delfino (Spain), F Gaede (Germany), J Kleist (Nordic countries), D Lucchesi (Italy, Chairman), H Meinhard (CERN, scientific secretary), P Sinervo (Canada)

1 Introduction

This report discusses resource usage in 2017, pledge fulfillment for 2018 and requests for 2019 for ALICE, ATLAS, CMS and LHCb. The LHC machine performed very well in 2017, providing more data than expected to the experiments. The computing, disk and tape space used is presented and compared to the pledged resources for each experiment, pointing out weaknesses and strengths. Pledge fulfillment and actual deployment is analyzed for 2018. The information from previous years together with the resource requests for 2019 constitutes the inputs examined to arrive at the C-RSG recommendations for 2019 resource procurement, after meeting for clarifications with the computing teams from the experiments. A first, very preliminary, look at 2020 requests is also provided.

2 C-RSG membership

Dugan O'Neil (Canada) and Jeff Templon (The Netherlands) ended their mandates. We thank both Dugan and Jeff for their important contributions over the years of their membership.

Canada has nominated Prof. Pekka Sinervo (University of Toronto) as their representative; Pekka has already actively contributed to the spring 2018 scrutiny. We ask the RRB to officially endorse the nomination.

No new Dutch representative has been nominated yet. We invite the Dutch Funding Agency to nominate one as soon as possible.

The chairperson thanks the C-RSG members for their commitment and the experiments representatives for their collaboration with C-RSG. Thanks are also due to the CERN management for their support and to our scientific secretary, H Meinhard (CERN), for ensuring the smooth running of the group.

3 Interactions with the experiments

The experiments were asked to anticipate the submission of their reports to February 12$^{\text{th}}$, in order to have enough information and interact with LHCC if needed. A second deadline of March 7$^{\text{th}}$ was set in case interaction with LHCC would require major modifications to the originally submitted documents.

The C-RSG thanks the experiments for the timely submission of their detailed documents [1–5]. The group would like to thank the computing representatives of the experiments for their availability, their responses to questions and subsequent requests for further information, and for their helpful discussions with us.

As usual, by agreement with ATLAS and CMS management, a single team of C-RSG referees scrutinized the ATLAS and CMS reports and requests to ensure a consistent approach.

We ask the experiments to submit their documents for the October 2018 RRB by September 3$^{\text{rd}}$ 2018.

4 Overall Resource Usage in 2017

Overall WLCG resources usage is analyzed and compared to previous years. Data are obtained from the EGI portal [6]. The time period goes from April 1st 2017 up to March 31st 2018. Disk and tape space occupancy are not available for the month of March 2018, therefore the tables and plots refer to the accounting numbers as of end of February 2018. Experiment by experiment resources usage will be described in dedicated sections.

The experiments have exploited all available resources and demonstrated their ability to also use opportunistic CPU cycles within WLCG and outside. HPC centers are becoming an important contributor to LHC computing and therefore C-RSG asked the experiments to report on that.

C-RSG reports on the usage following the usual classification of CERN, Tier-1 and Tier-2 even if that starts being superseded.

4.1 CERN, Tier-1 and Tier-2 Usage

The usage relative to the pledged resources for CERN, Tier-1 and Tier-2 is shown in table 1 for the last four years. Values are averaged over the four experiments. CPU usage is calculated as the average of time-integrated CPU power over the RRB year for 2017 and 2016, calendar year is used before. Disk and tape numbers give the occupancy at the end of the RRB or calendar years.

Used/pledged resources

		2017	2016	2015	2014
CPU	CERN	105%	122%	39%	53%
	T1	97%	119%	102%	123%
	T2	143%	151%	111%	152%
Disk	CERN	72%	97%	80%	81%
	T1	88%	72%	82%	95%
	T2	—	—	—	–
Tape	CERN	64%	98%	76%	96%
	T1	53%	67%	69%	89%

Table 1 Usage summary for different Tiers for 2017 and 2016 RRB year and for calendar year 2015 and 2014. Data is from Tier-1 and Tier-2 accounting summaries for WLCG obtained from EGI [6]. CERN percentage is not taken into account properly before 2016 and this explain the difference with the latest years.

As for previous years, CPU usage benefits significantly from beyond pledged resources. Disk space is fully utilized. Tape space usage doesn't saturate the pledges and will be discussed below experiment by experiment.

Figure 1 shows the yearly evolution of the share of CPU usage by experiment at CERN (top left), Tier-1 (top right) and Tier-2 (middle bottom). In each plot, the percentage used by each experiment normalized to the total CPU cycles used is plotted, therefore they sum up to 100% year by year. The major users at CERN are ALICE, ATLAS and CMS, with LHCb being the minor user at the 5% level. Tier-1 CPU usage is dominated by ATLAS, with CMS at 25% and ALICE and LHCb around 15%. Tier-2 sees ATLAS and CMS as dominant users with ALICE and LHCb consuming 10% of the resources.

Figure 2 shows the year by year sharing of disk space by experiment at CERN (left) and Tier-1 (right). The percentage is obtained as the space used by each experiment divided by the total disk space used

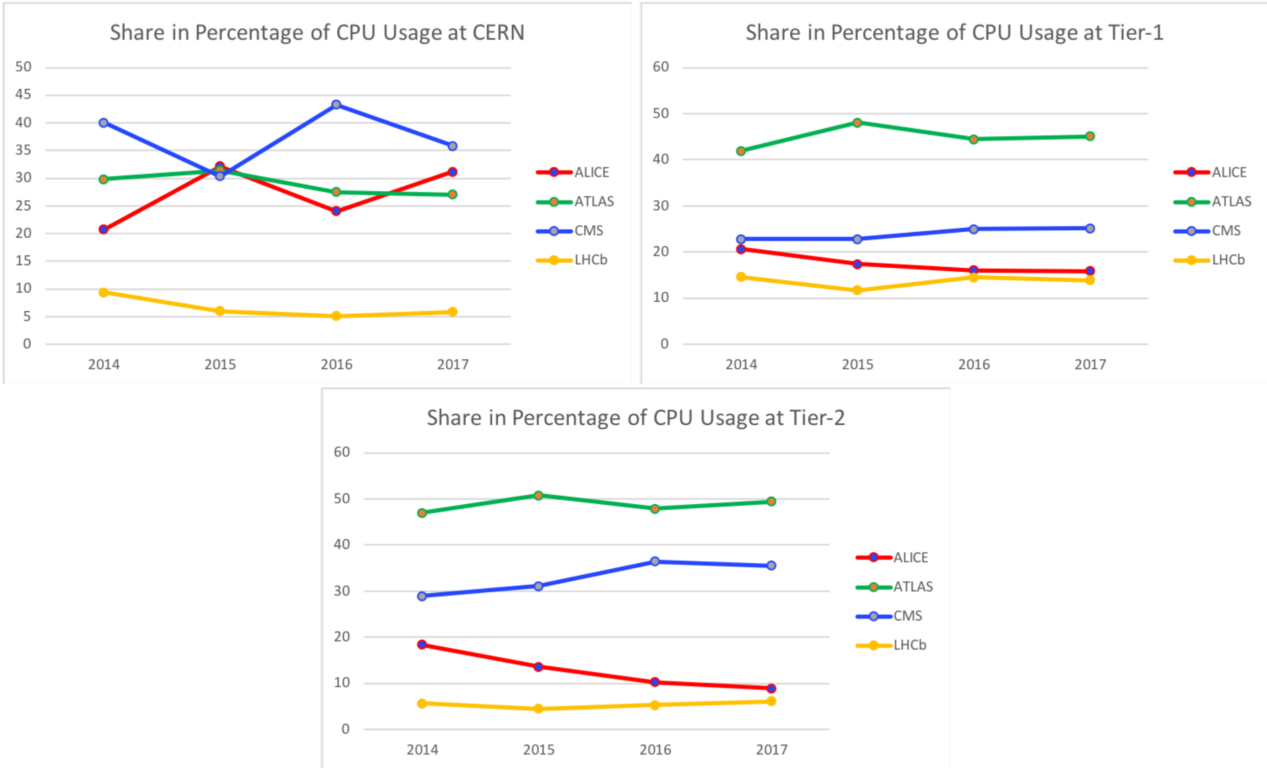

Figure 1 Year by year share of CPU resources by experiment at CERN (top left), Tier-1 (top right) and Tier-2 (bottom). Data were obtained from EGI [6] using the RRB year.

at CERN and Tier-1, therefore they sum up to 100% year by year. Information from Tier-2 is not included since it is not available. At CERN, ALICE, ATLAS and CMS occupy almost the same amount of disk space with LHCb steadily reducing its share in recent years to the current 10%. At Tier-1, ATLAS largely and permanently dominates disk space use with a percentage of around 50%. CMS oscillates between 20% and 30%. ALICE and LHCb are quite stable just above the 10% level.

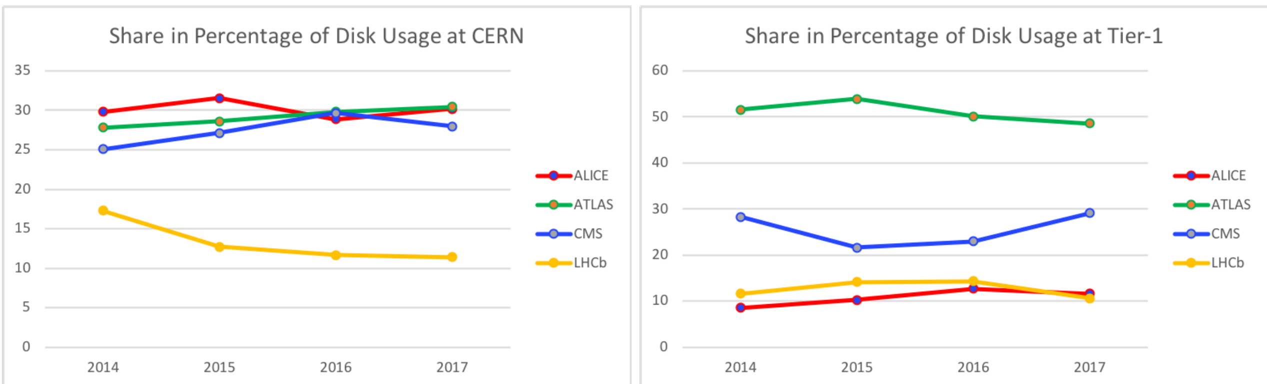

Figure 2 Year by year share of disk space by experiment used at CERN (left) and Tier-1 (right). Data were obtained from EGI [6] using the RRB year.

Usage of tape space is shown in figure 3. The behavior at CERN (left) and at Tier-1 (right) is similar. The major users are ATLAS and CMS with averages around 40%. ALICE and LHCb consume 20% and 15% at CERN, while their share at Tier-1 is around 10%.

Finally, figure 4 explores the year by year evolution of the relative share of resources used at CERN on an experiment by experiment basis. CPU cycles are shown on the left and disk space on the right.

For CPU, an increase in 2016 is common to all experiments due to the fact that CERN provided the requested resources while the same did not happen in all the funding agencies. The CERN contribution for LHCb and ATLAS is around 15%, while for CMS is now less than 25% and for ALICE is slightly more than 40%. The CERN disk space share is taken with respect to CERN and Tier-1, as there is no disk space accounting at Tier-2s. The major disk space consumer at CERN is ALICE, starting from more than 50% and being now at 60%. CMS and LHCb are now slighlty below 40% and ATLAS is below the 30% level. As for the CPU, the CERN disk space share increased for the four experiments at CERN with respect to the outside in 2016 due to the different policy in resources procurement.

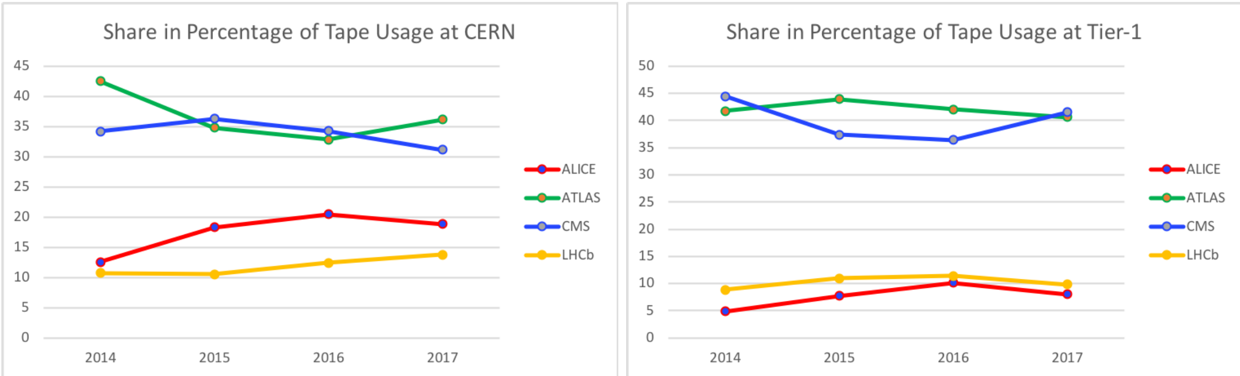

Figure 3 Year by year share of tape space by experiment for CERN (left) and Tier-1 (right). Data were obtained from EGI [6] using the RRB year.

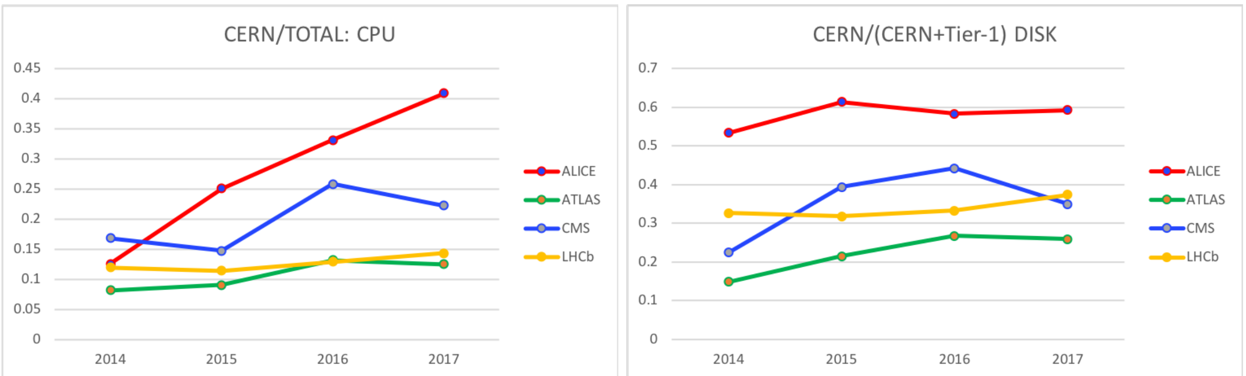

Figure 4 Left: CPU usage at CERN divided by the total, i.e. the sum of CERN, Tier-1 and Tier-2 CPU. Right: Disk space usage at CERN divided by the sum of CERN and Tier-1 usage. Tier-2 disk space consumption is not available in the accounting portal.

4.2 CPU efficiency

The experiments continue to work on improving the CPU efficiency, defined as CPU time divided by wallclock time. In figure 5, on the left, the Tier-1 efficiency is shown for the last RRB year. It can be noticed that CMS is recovering. We noticed that the efficiency is greater than 100% starting from 2018 for LHCb. This has been investigated and understood to be an issue with one site, site managers are working on fixing it. Figure 5, right, the CPU efficiency for the last four years is displayed.

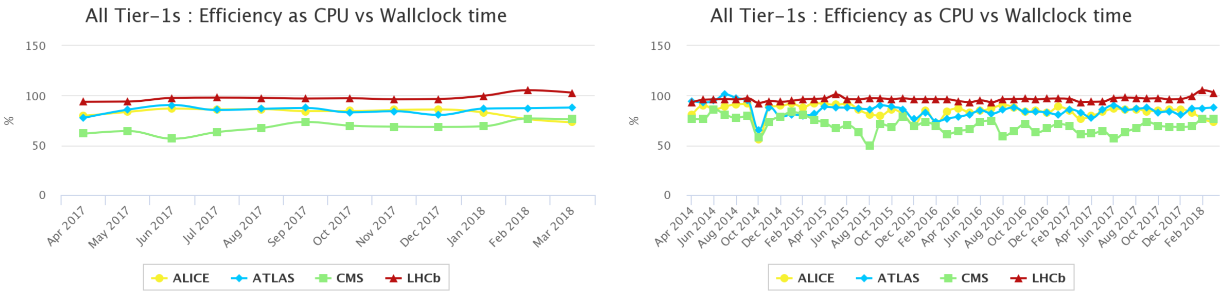

Figure 5 Left: Tier-1 CPU efficiency for the 2017 RRB year. Right: Tier-1 CPU efficiency for the last four years.

5 Resource Usage: ALICE

The CRSG report is based on the resource requests provided by the ALICE experiment [1], written responses to a set of scrutiny questions, and an in-person meeting with the ALICE computing coordinators. The focus of this scrutiny round is the resource usage for 2017, the allocations for 2019, and initial projections of resources required for 2020.

Figure 6 summarizes the computing resource usage by ALICE for the period January 1st to December 31st 2017. ALICE began data taking in May 2017 with a Ne-CO_2- N_2 TPC gas mixture. Beam and trigger conditions were similar to 2016 with LHC delivering 16.6 nb^{-1} of pp at $\sqrt(s) = 13$ TeV integrated luminosity, 170 hours of pp at $\sqrt(s) = 5.02$ TeV, and one LHC fill of Xe-Xe collisions.

Flooding at the CNAF T1 center in November 2017 impacted the CPU and storage available to ALICE (CNAF represents about 22% of the ALICE T1 capacity and holds 6.2PB of RAW data on tape). Additional resources made available by CERN and GridKA offset the computing shortfall. Replication of data at other sites or on tape at CERN is expected to be able to account for possible damage to files but the long term impact (e.g. how much data needs to be regenerated) is still under investigation.

Tape usage at T0 through the end of 2017 amounts to 4.0 PB for a cumulated total of 29.7 PB. Tape usage at T1 is 22.3PB. Combined, this corresponds to 77% of the CRSG approved capacity. The reduction in tape usage, compared to expectation, is due to significant improvements in the compression of the RAW events (reducing the effective size of a RAW event from 2 MB to 1.7 MB). This was achieved by improved HLT compression (increasing from a factor of 5 to 7.2), reduced pileup, and a decrease in split and noise clusters due to the use of Ne-CO_2- N_2. The unused tape and improved compression has been accounted for in the requests for 2018 and 2019.

The CRSG congratulates ALICE on its success in reducing the event size and number of noise clusters and encourages the experiment to work on the reduction of the size of Pb-Pb events which are predicted to increase for the next heavy ion run.

Disk pledged as of Jan 1 2018 was 22.4 PB, 21.8 PB, and 22.7 PB at the T0, T1, and T2 sites respectively. Of this 86%, 83% and 89% have been utilized. There remains a deficit between pledged and requested disk at the T1 and T2s of 14% and 28% respectively. This situation has improved in 2017 compared to previous years. Given the current disk usage and the projections through the end of the scrutiny the under-deployment of disk does not appear to be impacting the experiment significantly.

ALICE used 389 kHS06 at T0 (33% more than the pledged resources), 295.3 kHS06 at T1 (25% more than the pledged resources), and 299 kHS06 at T2 (7% more than the pledged resources). The experiment continues to make good use of opportunistic CPU resources from the Helix Nebula Science Cloud and the Titan supercomputer at ORNL.

ALICE		2017				
		CRSG recomm.	Pledged	Pledged /CRSG	Used	Used /CRSG
CPU	Tier-0	292	292	100%	389	133%
	Tier-1	256	235.5	92%	295	115%
	Tier-2	366	279.6	76%	299	82%
	HLT	n/a	n/a	n/a	26	n/a
	Total	**914**	**807.1**	**88%**	**1010**	**110%**
	Others				*n/a*	*n/a*
Disk	Tier-0	22.4	22.4	100%	19.3	86%
	Tier-1	25.4	21.8	86%	18.245	72%
	Tier-2	31.4	22.7	72%	20.06	64%
	Total	**79.2**	**66.9**	**84%**	**57.6**	**73%**
Tape	Tier-0	36.9	36.9	100%	29.7	80%
	Tier-1	30.9	30.6	99%	22.3	72%
	Total	**67.8**	**67.5**	**100%**	**52**	**77%**

Figure 6 Summary of planned and used resources for ALICE in WLCG year 2017

The HLT provides approximately 3% of the total compute requirements for ALICE and has a utilization rate of more than 40% over the year. The HLT is unlikely to be available to ALICE for all of 2018 as it is scheduled to be dismantled in November 2018.

ALICE continues to demonstrate good compute efficiency at the T1 and T2 sites with CPU-to-wall clock ratios of ~85%. Lower efficiency at T0 is due to the larger proportion of calibration cycles that run at T0 compared to T1s.

Simulations are the dominant component of the computational budget. Given the length of simulation payload jobs (typically 5 hours with Geant3) and the typical backfill time available in HPC systems (<1 hour) ALICE has not had as much success in utilizing opportunistic resources from HPC clusters. Geant4's multithreading and event level parallelism should reduce a simulated event run time to ~1 hour. This should increase the amount of HPC and opportunistic grid time that would be accessible to ALICE. We strongly encourage the development sub-event parallelization of Geant4 as this will enable payload times of substantially less than an hour better matching opportunistic resource availability.

From the data popularity plot is shown in figure 7. ALICE is encouraged to maintain an active disk cleanup policy in order to reduce the volume of infrequently used data to maximum possible extent.

6 Resource Usage: ATLAS

Figure 8 shows an overview of ATLAS' resource usage for 2017. This information is based on the report from ATLAS [2], with pledged resources extracted from REBUS; used resources are obtained from the EGI accounting portal [6].

With regard to the approved 2017 RRB figures, we observe that ATLAS has made a very high usage of pledged CPU and disk resources.

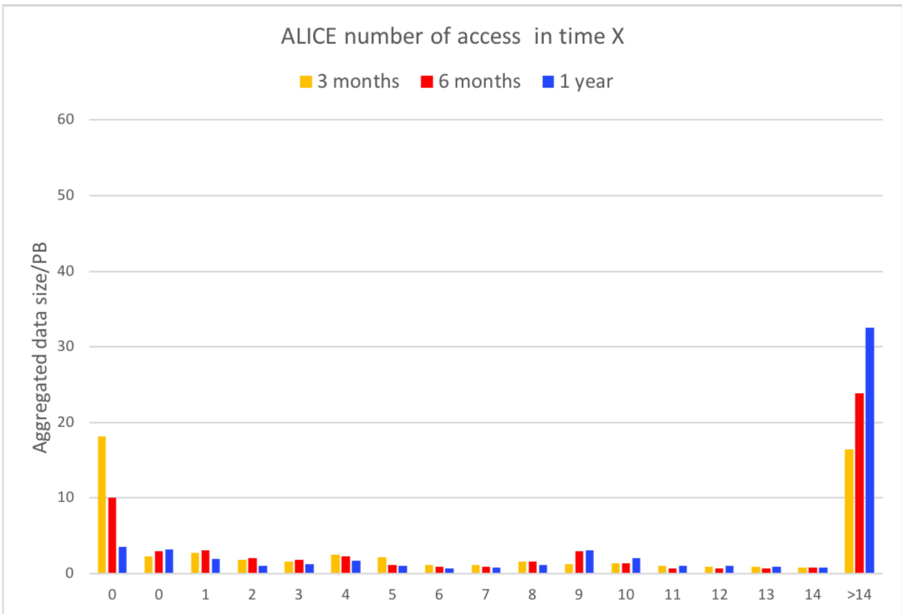

Figure 7 ALICE data popularity plot. Number of accesses to ALICE datasets for periods of 3, 6 and 12 months. The data for unaccessed data is split into two, the leftmost depicts data which are older than 3, 6 and 12 months respectivly and the rightmost data which are younger then 3, 6 and 12 months.

CPU usage has been beyond pledge on Tier-0, Tier-1 and Tier-2 sites by 54, 40, and 518 kHS06, respectively. A significant amount of overpledge CPU resources (98 kHS06) were also obtained from the HLT farm by opportunistic offline usage between data taking periods. As in previous years, ATLAS has also obtained significant amount of non-pledged CPU resources via HPC sites such as NERSC and Oak Ridge (primarily for event generation and Geant4 simulation). ATLAS estimates these to 9.3% of their total number of simulated events. Additional resources were obtained from cloud resources via BOINC/ATLAS@HOME (around 100 kHS06).

CPU usage is dominated by MC simulation, followed by reconstruction of real and simulated data and event generation. Average CPU efficiency levels (calculated as CPU time / wall time) have slightly increased in 2017. The efficiency of the HLT farm has risen from 76% to 90% as it has been mostly used for MC simulation. As reported in the 2017 fall report, ATLAS had initiated work activities aimed to increase efficiency. Two notable improvements for further increasing the efficiency of multi-core jobs are currently being deployed: On one hand, the ATLAS event service which is progressively being rolled out addresses inefficiencies with the multi-process event loop, ensuring that all workers are kept busy until the end of the complete job. On the other hand, the serialized merge step at the end of the job is progressively being replaced by a shared writer mechanism that merges events already during the event loop.

The high LHC efficiency during the last weeks of the 2017 run caused a backlog of 250K jobs on the Tier-0. For 2018, ATLAS needs to ensure sustainability via spill-over of Tier-0 jobs to the grid, given that LHC efficiency and pile-up rates are expected to remain high and 2018 Tier-0 CPU pledges will only be marginally higher than the 2017 ones.

Disk occupancy levels have further risen compared to 2016, and are now reaching the limits (100% on Tier-0 and Tier-2, 96% on Tier-1). Disk space is dominated by 2015/2016 MC and data, with 2017 data progressively increasing over the year. In terms of data classes, DAOD and AOD are the dominating types, with around 1/3 total occupancy each. In order to optimize space fragmentation, ATLAS is further reducing so-called group space that is managed by individual physics groups and has reallocated around 6PB to centrally managed pools. ATLAS is also working on reducing the footprint of DAOD's by scrutinizing their production and optimizing event selection. Moreover, ATLAS is

ATLAS		2017				
		CRSG recomm.	Pledged	Pledged /CRSG	Used	Used /CRSG
CPU	Tier-0	404	367	91%	421	104%
	Tier-1	921	786	85%	826	90%
	Tier-2	1125	917	82%	1505	134%
	HLT	n/a	22	n/a	120	n/a
	Total	**2450**	**2092**	**85%**	**2872**	**117%**
	Others				*230*	*9%**
Disk	Tier-0	25	25	100%	25,0	100%
	Tier-1	68	70	103%	67	99%
	Tier-2	83	78	94%	78	94%
	Total	**176**	**173**	**98%**	**170,0**	**97%**
Tape	Tier-0	77	77	100%	58	75%
	Tier-1	188	175	93%	102	54%
	Total	**265**	**252**	**95%**	**160**	**60%**

Figure 8 Summary of planned and used resources for ATLAS in WLCG year 2017.
*Percentage taken with respect to total CRSG CPU recommendation.

working on improving the compression of AOD's and potentially DAOD's.

There is a significant shortfall between pledged and used tape resources. Tape occupancy is 75% on the Tier-0 and 58% on the Tier-1's, respectively. ATLAS explains this as following: First of all, ATLAS tape requests are 10% larger than the actual needs, in order to allow for efficient usage and flexibility. On the Tier-0, the collaboration was expecting larger datasets for cosmic rays and calibrations. In addition, non-RAW data sets from Run-1 were deleted. On the Tier-1s, the difference is a consequence of a better than expected impact of the lifetime model for old MC datasets which is now applied to all non-raw ATLAS data.

Figure 9 depicts number of accesses to ATLAS datasets for periods of 3, 6 and 12 months. Compared to last years data, we see an increase in the amount of untouched data, at month 6 about 8 PB more. The increase is explained by an increase in the size of the datasets and in quite a large amount of data which are not yet completely ready for analysis.

7 Resource Usage: CMS

Figure 10 shows an overview of CMS' resource usage for 2017. This information is based on the report from CMS [3], with pledged resources extracted from REBUS; used resources are obtained from the EGI accounting portal [6].

During 2017 CMS operated under challenging conditions especially triggered by a major incident at their second biggest Tier-1 and issues with EOS at the Tier-0 which was only resolved in August but also due to less resources being pledged than recommended by C-RSG. The impact was mitigated

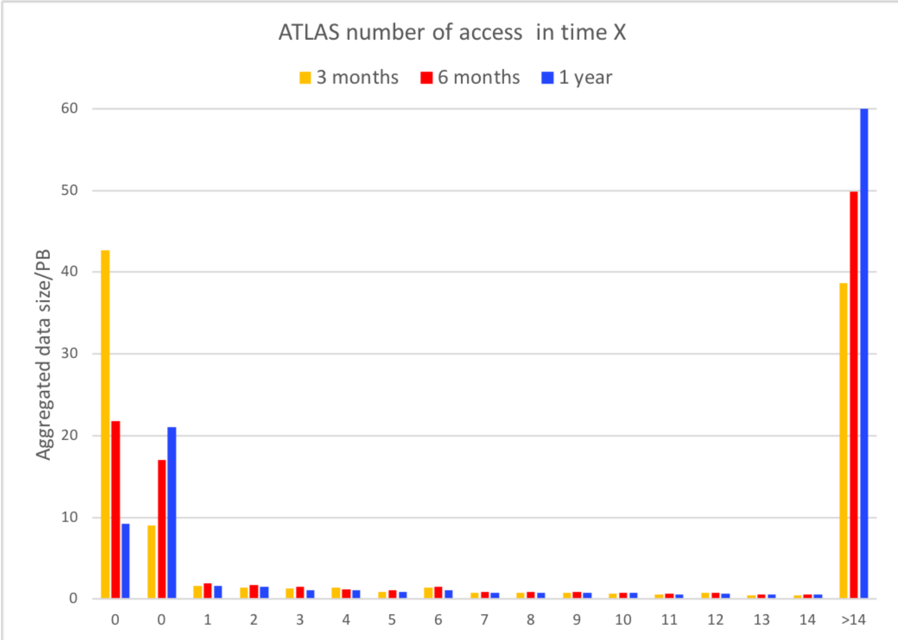

Figure 9 ATLAS data popularity plot. Number of accesses to ATLAS datasets for periods of 3, 6 and 12 months as of January 2018. The data for unaccessed data is split into two, the leftmost depicts data which are older than 3, 6 and 12 months respectively and the rightmost data which are younger then 3, 6 and 12 months.

CMS		2017				
		CRSG recomm.	Pledged	Pledged /CRSG	Used	Used /CRSG
CPU	Tier-0	423	397	94%	326	77%
	Tier-1	600	470	78%	425	71%
	Tier-2	850	772	91%	1133	133%
	HLT	n/a	n/a	n/a	148	n/a
	Total	**1873**	**1639**	**88%**	**2032**	**108%**
	Others				*30*	*2%**
Disk	Tier-0	24,6	25	102%	21,0	85%
	Tier-1	57	45	79%	39	68%
	Tier-2	68	54	79%	48,6	71%
	Total	**149,6**	**124**	**83%**	**108,6**	**73%**
Tape	Tier-0	70,5	70	99%	49	70%
	Tier-1	175	133	76%	111	63%
	Total	**245,5**	**203**	**83%**	**160**	**65%**

Figure 10 Summary of planned and used resources for CMS in WLCG year 2017. Note that CMS manually keeps T2 disk usage at 90% utilization, so the T2 disk utilization does not indicate lack or surplus of T2 disk resources.

*Percentage taken with respect to total CRSG CPU recommendation.

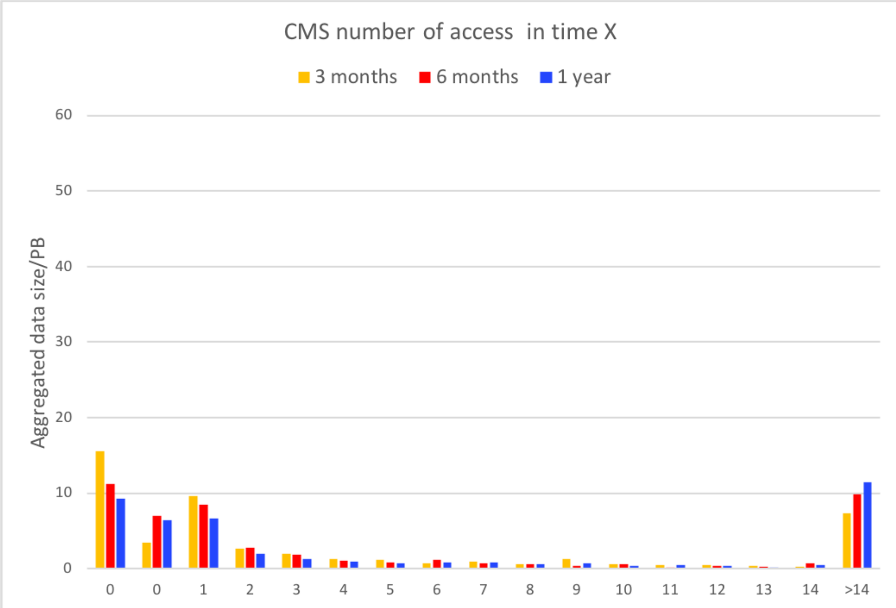

Figure 11 CMS data popularity plot. Number of accesses to CMS datasets for periods of 3, 6 and 12 months as of January 2018. The data for unaccessed data is split into two, the leftmost depicts data which are older than 3, 6 and 12 months respectively and the rightmost data which are younger then 3, 6 and 12 months.

for computationally by additional opportunistic resources provided by Tier-2 sites resulted in an addition 8% compute being consumed compared to the pledged resources. This was made possible by a continued development of the CMS distributed computing model which continue to reduce the functional differences between Tiers, allowing for better flexibility. CMS have also run aggressive deletion campaigns to free tape space at Tier-1s. C-RSG understands that this was only achieved through an additional operational effort.

CMS is making extremely good use of the HLT farm and during 2017 and have also been exploiting on-demand and opportunistic resources. The CMS setup allowing use of the HLT farm even during short bursts of availability makes it contribution to CMS non-negligible.

Overall CMS has made excellent use of the provided resources and put in place a number of improvements to reduce data size and make storage management more dynamic.

In early summer 2017 CMS initiated work on investigating CPU inefficiencies [7]. Generally speaking inefficiencies can be characterized as coming from two main sources; the submission infrastructure and the CMS payload itself. Currently focus is on improving the payload efficiency based on CMS' understanding that remaining inefficiency caused by the submission infrastructure is virtually unavoidable.

Figure 11 depicts number of accesses to CMS datasets for periods of 3, 6 and 12 months. Comparing this plot to the one presented in last years C-RSG report, we see quite an increase in the amount of data accessed 14 or more times. The unaccessed data consists of data which are only kept on disk for 3 months after data taking and then deleted, as well as of data used to test new reconstruction algorithms where the output is only accessed for a fraction of the samples.

8 Resource Usage: LHCb

Figure 12 shows an overview of resource usage by LHCb for WLCG year 2017. This information is based on their report [4], with pledged resources extracted from REBUS and used resources obtained from the EGI accounting portal [6] and LHCb's own accounting if necessary (for example, for Tier-2 disk usage).

LHCb		2017				
		CRSG recomm.	Pledged	Pledged /CRSG	Used	Used /CRSG
CPU	Tier-0	67	67	100%	63	94%
	Tier-1	207	199	96%	245	118%
	Tier-2	116	147	127%	196	169%
	HLT	10	0	0%	208	2080%
	Total	**400**	**413**	**103%**	**712**	**178%**
	Others				*114*	*29%**
Disk	Tier-0	10.9	10.9	100%	6.6	61%
	Tier-1	22.1	20.9	95%	16.9	76%
	Tier-2	4.7	3.3	70%	3.6	77%
	Total	**37.7**	**35.1**	**93%**	**27.1**	**72%**
Tape	Tier-0	25.2	25.2	100%	22.5	89%
	Tier-1	43.3	42.0	97%	31.3	72%
	Total	**68.5**	**67.19**	**98%**	**53.8**	**79%**

*Percentage taken with respect to Total CRSG CPU recommendation

Figure 12 Summary of planned and used resources for LHCb in WLCG year 2017

The pledged and used Tier-1 resources reflect the impact of the unscheduled downtime due to the accident at CNAF diluted over the whole year. It should be noted that the actual impact was substantial and concentrated in time, with about 18% of LHCb's data becoming inaccessible for many weeks. The situation is largely recovered thanks to the efforts of CNAF coordinated with the LHCb data processing team. Nevertheless, some stripping activities which should have been completed in 2017 are still ongoing, processing the data stored at CNAF.

Pledged disk storage has not been fully utilized on average (61% at Tier-0 and about 75% at Tier-1 and Tier-2), although peak utilization has used resources to the maximum available. These figures illustrate LHCb's very active disk management and economies in disk brought by improvements in data formats and computing model.

LHCb has used a higher percentage of the pledged tape than in previous years. This reflects the structural adjustment made last year to bring requests better in line with the computing model.

The LHCb production system is capable of identifying and using additional CPU capacity at WLCG Tier centers, the HLT and other resource providers. This additional capacity is used to provide additional simulations which improve or enable various analyses. For 2017, an additional 18% and 69% of CPU has been provided at Tier-1s and Tier-2s, respectively. Use of the HLT CPU has been extremely effective, yielding over 20 times more CPU than planned. CPU from resources outside WLCG is substantial, at the 30% level. These extra resources are used for non-critical tasks.

Regular monitoring of data popularity continues, as can be seen in Figure 13. The results are used to optimize the use of disk. This allowed some 3.8 PB of data to be removed from disk during 2017, which corresponds to 11% of the total deployed disk space.

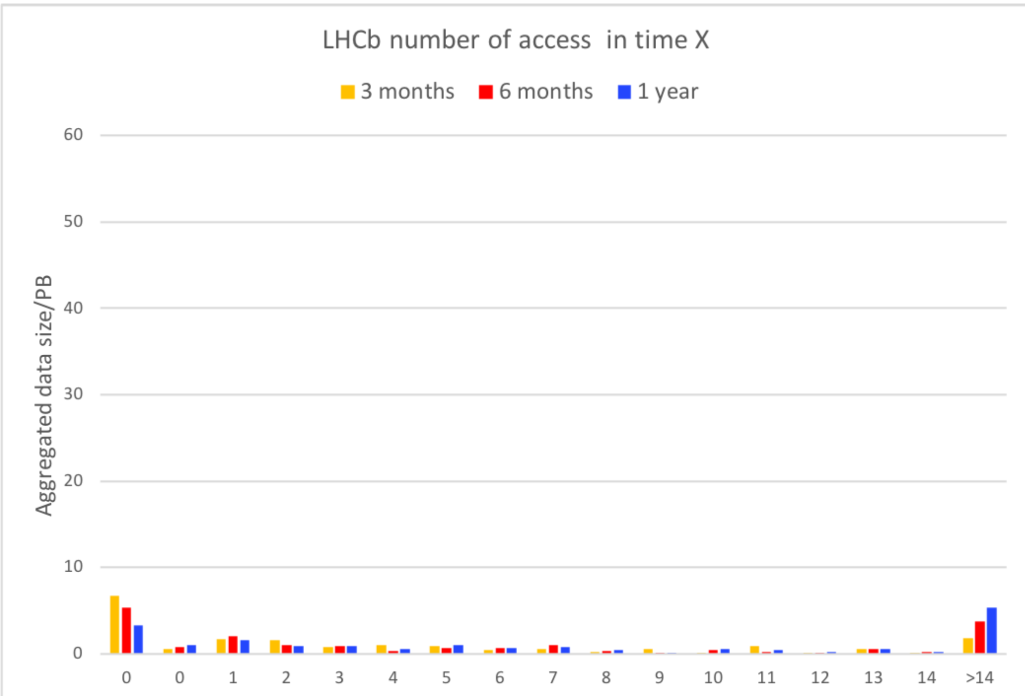

Figure 13 LHCb data popularity plot. Number of accesses to LHCb datasets for periods of 3, 6 and 12 months as of January 2018. The data for unaccessed data is split into two, the leftmost depicts data which are older than 3, 6 and 12 months respectivly and the rightmost data which are younger then 3, 6 and 12 months.

LHCb reports good progress on the refinement of their TURBO data format, which greatly optimizes the use of storage while maintaining a high functionality for analysis. This results in an absolute level of accessed data volume from disk which is quite moderate when compared to the total in WLCG, as can also be seen in Figure 13 (the vertical scale in all data popularity plots in this report is the same).

RRB year	pp/10^6 s	HI/10^6 s	pp pileup
2015	3	0.7	25
2016	5	0.7	35
2017	7.8	-	35
2018	7.8	1.2	35
2019	-	-	-

Table 2 Assumptions on live time for LHC running in Run 2, 2015 to 2019. The final column gives the anticipated average pileup for ATLAS and CMS during pp running for each year.

Figure 17 summarizes the situation with pledged resources for WLCG year 2018, which is mostly satisfactory. The overpledged situation for tape at Tier-1s is most likely due to capacity provisioned prior to the downward adjustment of requests by LHCb in Fall 2017.

Disk pledged at Tier-2s falls short of the C-RSG recommendations. On an absolute scale, however, the shortfall is small and can be compensated by using disk deployed at Tier-1s if late pledges from Tier-2s do not arrive.

9 Background for 2019 Resource Requirements

The 2019 will be a shutdown year for LHC [8]. Therefore, requests are dominated by the needs for the processing of data collected until the end of 2018, Monte Carlo production and studies for the upgrade phase.

Table 2 summarizes the status of the machine as has been considered by the experiment and the C-RSG group to evaluated the needs. ATLAS and CMS are not requesting any resource increase at Tier-0 for 2019, while LHCb asks for a modest increase and ALICE asks for a substantial increase. Tier-0 management expects [9] that CERN resources pledged in REBUS will be identical in 2019 to 2018; the basic reasoning is that CERN intends to re-profile a maximum of funding to the beginning of Run 3, when resource needs are expected to be as high as about twice the 2018 level. Re-profiling of funding by other Funding Agencies may also impact their ability to fully pledge resources requested and recommended for 2019 (and 2020). The situation will have to be closely monitored in future rounds of review.

The analysis and discussion of the individual experiment needs is given in the next sections.

10 Resource Requests: ALICE

ALICE resource requests for 2019 are shown in figure 14. The requests have changed substantially since the October 2017 RRB with a 25% decrease in requests for CPU, and a 12% decrease in disk. Over the 2018 agreed resources, these requests represent an increase of 21% for CPU, 24% for disk and a 9% decrease in tape. We note that, when comparing to the 2018 request for resources, the CRSG was not able to make a recommendation for T2 CPU and disk given the information presented to the CRSG at the time. In agreement with the ALICE experiment we have, therefore, considered the 2018 pledged resources as the baseline for future requests.

The resource requests for disk (24% for Tier-1 and 17% for Tier-2) and CPU (19% for Tier-1 and 20% for Tier-2) at the Tier-1 and Tier-2 sites are consistent with the nominal growth limits requested by the funding agencies.

ALICE		2017			2018		2019		
		CRSG recomm.	Pledged	Used	CRSG recomm.	Pledged	Request	2019 req. /2018 CRSG	C-RSG recomm.
CPU	Tier-0	292	292	389	350	350	430	123%	430
	Tier-1	256	235.5	295	307	279.5	365	119%	365
	Tier-2	366	279.6	299	312.9	312.9	376	120%	376
	HLT	n/a	n/a	26	n/a	n/a	n/a	n/a	0
	Total	**914**	**807.1**	**1010**	**969.9**	**942.4**	**1171**	**121%**	**1171**
	Others			*39*					
Disk	Tier-0	22.4	22.4	19.3	26.2	26.2	34.3	131%	34.3
	Tier-1	25.4	21.8	18.245	30.5	30.4	37.9	124%	37.9
	Tier-2	31.4	22.7	20.06	29	29	33.9	117%	33.9
	Total	**79.2**	**66.9**	**57.6**	**85.7**	**85.6**	**106.1**	**124%**	**106.1**
Tape	Tier-0	36.9	36.9	29.7	49.1	49.1	44.2	90%	44.2
	Tier-1	30.9	30.6	22.3	40.9	42.2	37.7	92%	37.7
	Total	**67.8**	**67.5**	**52**	**90**	**91.3**	**81.9**	**91%**	**81.9**

Figure 14 ALICE resources request and C-RSG recommendations.

As described in the ALICE report, the request for a 31% increase in Tier-0 disk and 23% in Tier-0 CPU is mainly driven by the processing and analysis of Pb-Pb data expected in 2018. For 2019, ALICE will undertake an initial pass of the Pb-Pb data sample and complete two passes of the pp data (including MC production). The event size for Pb-Pb reconstruction size is expected to increase compared to the 2015 run with the ESD+AOD increasing from 1.4 MB to 4.2 MB per event and the Monte-Carlo increasing from 2.7 MB to 11MB per event. This increase is due to a different trigger mix during Pb-Pb data taking between 2015 and 2018. In 2015 the trigger mix was dominated by minimum bias events. In 2018 the readout rate will double with a factor of three increase in central events compared to minimum bias events. We strongly encourage ALICE to work to reduce the size of these events and associated MCs to mitigate the requested increase in 2019 resources.

C-RSG requested a mitigation strategy from ALICE to address a hypothetical increase of 20% as a consequence of changes in the running conditions of the experiment (e.g. increased pile-up or luminosity). Under such circumstances, ALICE made the following considerations: they will keep data on tape only with its reconstruction, MC, and analysis delayed until sufficient amount of CPU and disk resources are available. The required resources are expected to scale linearly with the amount of data.

We welcome ALICE's work on Geant4 multithreading. We strongly encourage the development of sub-event parallelization of Geant4 as this will enable payload times of substantially less than an hour better matching HPC and opportunistic resource availability.

ALICE has made an initial estimate of their 2020 needs with no increase in CPU or tape at any of the sites, but an increase in disk of 22%, 20%, and 15% at Tier-0, Tier-1, and Tier-2 respectively. The expectation for 2020 is that ALICE with complete the third pass for the pp data and the final two passes for the Pb-Pb data collected in 2018 (plus associated MC production). 20% of CPU resources in 2020 will be spent on a validation of Geant4 with the remaining 80% devoted to the processing of the pp, Pb-Pb data and simulations.

In summary, the C-RSG endorses the ALICE Tier-0, Tier-1 and Tier-2 resource requests for 2019.

ATLAS		2017			2018		2019		
		CRSG recomm.	Pledged	Used	CRSG recomm.	Pledged	Request	2019 req. /2018 CRSG	C-RSG recomm.
CPU	Tier-0	404	367	421	411	411	411	100%	411
	Tier-1	921	786	826	949	969	1057	111%	1057
	Tier-2	1125	917	1505	1160	1136	1292	111%	1292
	HLT	n/a	22	120	n/a	0	0	n/a	n/a
	Total	**2450**	**2092**	**2872**	**2520**	**2516**	**2760**	**110%**	**2760**
	Others			*230*					
Disk	Tier-0	25	25	25,0	26	27,0	26	100%	26,0
	Tier-1	68	70	67	72	80,0	88	122%	88,0
	Tier-2	83	78	78	88	86	108	123%	108,0
	Total	**176**	**173**	**170,0**	**186**	**193**	**222**	**119%**	**222,0**
Tape	Tier-0	77	77	58	94	105,0	94,0	100%	94,0
	Tier-1	188	175	102	195	196,0	221	113%	221,0
	Total	**265**	**252**	**160**	**289**	**301**	**315**	**109%**	**315,0**

Figure 15 ATLAS resources request and CRSG recommendations.

11 Resource Requests: ATLAS

ATLAS resource requests for 2019 are shown in figure 15. The requests have not been changed since the October 2017 RRB. Over 2018 agreed resources, they represent an increase of 10% for CPU, 19% for disk and 9% for tape. In addition and as in previous years, ATLAS expects beyond-pledge CPU resources to be substantial (around 700 kHS06), with WLCG Tier-2 sites providing the largest share (around 100% of additional capacity over pledged resources). This continued reliance on non-committed, non-guaranteed resources remains a risk for the experiment.

As mentioned in the CRSG Fall 2017 report, disk requests for Tier-1 and Tier-2 (23% each) are above nominal growth limits as requested by the funding agencies. With the exceptional LHC performance during 2017 and its probable continuation during 2018, disk resources continue to be critical for ATLAS data and MC sample generation. It is yet too soon to quantify the potential impact of efforts invested by ATLAS in reducing DAOD occupancy and compression improvements.

As described in the usage section, current ATLAS tape occupancy is significantly below pledged resources both on Tier-0 and Tier-1 sites. In addition, 2019 requests for Tier-1 tape represent an increase of 13% over 2018 RRB agreed resources. While these numbers were provided before the implementation of the new MC dataset lifetime model, ATLAS still expects tape usage to catch up with pledges over 2018, 2019 and 2020: On the Tier-0, ATLAS expects additional 22PB on 2018 that include also Heavy-Ion data. Another 5PB are used as contingency in the case of a 20% extra LHC luminosity. This results in 84PB for 2018, on which another 10% are added on top as efficiency/flexibility margin, resulting in a final 92PB (2PB below the 2017 spring RRB request). On Tier-1 sites, ATLAS expects usage to reach 161PB in 2018 (vs. 195PB pledged) and 221PB (as pledged) in 2019. This increase is due to the prolonged need for datasets produced and simulated during this period for full data analysis of Run-2.

C-RSG requested ATLAS a mitigation strategy to address a hypothetical increase of 20% as a consequence of changes in the running conditions of the experiment (e.g. increased pile-up or luminosity). Under such circumstances, ATLAS makes the following considerations:

For Tier-0 processing, a 20% increase would translate in 20% more CPU resources that would have to be offloaded to Tier-1's via spill-over operations. While the additional CPU resources themselves would be minimal (around 2% of the T1/T2 pledges), additional manpower would be required. ATLAS estimates that an additional operational effort of 0.5 FTE would be needed for ensuring smooth spillover operations. Higher data volumes would also require disk usage increases in the order of 8PB for AOD/DAOD events; here the mitigation would consist in increased usage of tape-based analysis workflows. This implies moving AOD datasets to tape and process them from there, thus reducing pressure on disk space. ATLAS is progressively increasing the role of tape in serveral T1 workflows using a "data carousel" approach, staging data to disk in advance to serve running jobs.

An additional validated mitigation option would be to reduce the trigger rate but ATLAS does not favour it as it impacts physics data analysis by using different trigger efficiencies.

2020 Outlook ATLAS has not been able to quantify their needs for 2020 as their computing needs for LS2 still need to be understood; these will depend on several factors such as the new simulation and reconstruction software, generation of MC samples and the strategy to follow for completing Run-2 analysis. ATLAS plans to report 2020 and beyond resource needs at the 2018 Fall RRB meeting. We strongly suggest ATLAS to submit their estimates earlier in the process in order to allow C-RSG to review them and provide guidance at the spring RRB meeting.

Recommendation Overall and despite the previously mentioned disk space increases, C-RSG endorses the ATLAS resource requests for 2019. We must point out that ATLAS has significantly more disk resources than CMS on the Tier-1 and Tier-2 sites. The difference continues to grow in the 2019 requests, with ATLAS' requests representing 29% more Tier-1 disk and 38% more Tier-2 disk than CMS. ATLAS explains this by differences in the computing model such as the data format and event sizes. CMS has made significant progress with the development and deployment of small-sized data formats such as miniAOD, and is investigating even more compact formats such as nanoAOD. C-RSG encourages ATLAS to consider similar goals and to increase efforts in investigating smaller data formats, higher compression rates and/or virtual data for fast simulation streams.

12 Resource Requests: CMS

CMS resource requests for 2019 are shown in table 16. The requests remain mainly unchanged from October 2017 RRB, with a reduction of 10 PB of tape at T1. The increase from the from the October RRB range from 0% at the CERN Tier-0 to between 8 and 17% for T1 and T2 resources.

As pointed out in previous C-RSG reports, the CMS continues to see fewer resources provided compared to the recommendations though the shortfall is now reduced to a much less alarming level.

CMS expect that the main activities for 2019 will be legacy reprocessing of the full Run-2 data and Monte Carlo samples, continuation of Phase 2 upgrade studies and analysis activties required to finalize Run-2 analyses and preparation of Run-3 and 4.

As with the other experiments C-RSG enquired what mitigiation strategy CMS would apply to address a hypothetical increase of 20% as a consequence of changes in the running conditions of the experiment (e.g. increased pile-up or luminosity). The CMS mitigation strategy would be

CMS		2017			2018		2019		
		CRSG recomm.	Pledged	Used	CRSG recomm.	Pledged	Request	2019 req. /2018 CRSG	C-RSG recomm.
CPU	Tier-0	423	397	326	423	423	423	100%	423
	Tier-1	600	470	425	600	562	650	108%	650
	Tier-2	850	772	1133	900	940	1000	111%	1000
	HLT	n/a	n/a	148	n/a	n/a	n/a	n/a	n/a
	Total	**1873**	**1639**	**2032**	**1923**	**1925**	**2073**	**108%**	**2073**
	Others			*30*					
Disk	Tier-0	24.6	25	21.0	26	26.1	26.1	100%	26.1
	Tier-1	57	45	39	60	55.4	68	113%	68.0
	Tier-2	68	54	48.6	70	66.7	78	111%	78.0
	Total	**149.6**	**124**	**108.6**	**156**	**148.2**	**172.1**	**110%**	**172.1**
Tape	Tier-0	70.5	70	49	99	97.0	99.0	100%	99.0
	Tier-1	175	133	111	188	166.0	220	117%	220.0
	Total	**245.5**	**203**	**160**	**287**	**263**	**319**	**111%**	**319.0**

Figure 16 CMS resources request and the CRSG recommendations.

- For T0, CMS could use T1s or CERN_T2 for a fraction of prompt reprocessing. This was already used last October. With the new T0 infrastructure based on HTCondor, the distinction T0-T2 fades, and the extension to T2 resources is transparent and based on the pool settings making this mitigation easy to apply;
- Concerning T0 disk buffers and transfers to T1s, CMS are continuing the decommissioning of the RECO data tier, which is by far the largest data tier Tier-0 has to keep in its buffers and send to Tier-1s;
- Regarding T1 and T2 disks, in case of troubles CMS can further reduce the number of AOD(SIM) copies (which is already below 1.
- For T1 and T2s, CPU can be mitigated if the same 20‰ increase is not requested / needed for Monte Carlo, and disk areas can be limited to (even more) recent data. The price to pay is a higher level of stage in/out from/to tape, and slower analysis operations in case of datasets on tape only. It is still unclear the impact for major reprocessing, where increased delay is expected due to the lower availability of prestaged samples.

2020 Outlook At the moment CMS do not have a full model for 2020, as it depends on the 2021 LHC machine parameters which are till mostly unknown. And the lookout below is to be considered highly speculative and subject to change.

For 2020 the main CMS activlity will be preparation for Run-III, via a dress rehearsal of the full computing system. Assuming a higher average PU of 55 for 2021 compared to 45 for 2018 and other

LHCb		2017			2018		2019		
		CRSG recomm.	Pledged	Used	CRSG recomm.	Pledged	Request	2019 req. /2018 CRSG	C-RSG recomm.
CPU	Tier-0	67	67	63	88	88	86	98%	86
	Tier-1	207	199	245	253	250	271	107%	271
	Tier-2	116	147	196	141	164	152	108%	152
	HLT	10	0	208	10	0	10	100%	10
	Total	**400**	**413**	**712**	**492**	**502**	**519**	**105%**	**519**
	Others			*114*					
Disk	Tier-0	10.9	10.9	6.6	11.4	11.4	14.1	124%	14.1
	Tier-1	22.1	20.9	16.9	24.5	26.3	27.9	114%	27.9
	Tier-2	4.7	3.3	3.6	5.7	3.7	6.8	119%	6.8
	Total	**37.7**	**35.1**	**27.1**	**41.6**	**41.4**	**48.8**	**117%**	**48.8**
Tape	Tier-0	25.2	25.2	22.5	33.6	33.6	35.0	104%	35.0
	Tier-1	43.3	42.0	31.3	45.6	56.9	50.9	112%	50.9
	Total	**68.5**	**67.19**	**53.8**	**79.2**	**90.5**	**85.9**	**108%**	**85.9**

Figure 17 Trend of resources for LHCb and requests and CRSG recommendations for 2019.

parameters unchanged, CMS's prelminiary models indicate that for 2020 no increase would be needed, provided completion of legacy LS2 reprocessing before spring 2020 followed by storage cleaning.

Looking further into 2021, the current CMS estimates indicate an increase with respect to 2019 requests of

- CPU: +30%
- Disk: +30%
- Tape: +22%

Recommendation Overall, C-RSG endorses the CMS resource requests for 2019. We note that the 2018 pledge is now less than 5% below for disk and 8% for tape compared to C-RSG recommendations, which lowers previous C-RSG concerns about the lack for pledged resources to CMS.

We applaude CMS's work on using reduced size data-formats such as miniAOD and their investigations into even more compact nanoAOD. We also appriciate the work done to investigate the background for the relative low CPU efficiency and urge CMS to continue to seek improvements to CPU efficency.

13 Resource Requests: LHCb

Figure 17 shows the request submitted by LHCb [5] in the context of previous years. Growth in requested tape resources is small, reflecting the shutdown of LHC. Moderate growth in CPU and disk resources is requested in order to accommodate a legacy re-stripping of all Run2 data and to complete all corresponding simulations, in order to fully enable completion of Run2 data analysis. Simulation needs in support of physics analysis dominate the use of CPU and have a significant impact on disk and tape storage.

LHC running conditions may of course vary with respect to the assumptions. LHCb's contingency plan remains the same as in previous years. If more data volume than foreseen is accumulated during 2018, LHCb will use their powerful online selection capabilities to park on tape specific datasets for which analysis will be postponed. This has a multiplying effect, as the corresponding need for simulation will also be postponed.

LHCb is putting effort into developing data and analysis preservation procedures (DP/DA) and is participating in the CERN Open Data (COD) project. These activities are concentrated at CERN and are considered part of a pilot project of the Laboratory, and therefore are not included in Figure 17. For completeness, we mention that LHCb is requesting for 2019 at CERN 30 kHS06-years of CPU and 0.5 PB of disk for DP/DA, and 0.5 PB of storage in the COD project for their first release of Open Data.

2020 Outlook

LHCb is well advanced in estimating their resource needs for 2020. In spite of the lack of new data from LHC, there will be a need for 15-20% growth in CPU and disk resources, whilst tape growth is expected to be negligible. These resources will be needed to overlap the consolidation of all processed data from Run2 and the corresponding physics analysis and simulations with a substantial simulation effort for Run3, which includes simulation support for determining parameters for the upgraded detector including those related to data acquisition.

Funding agencies are advised to take this forward outlook for 2020 into consideration in their plans.

LHCb recommendations

C-RSG congratulates LHCb on the very successful management of their computing model and the links to their data processing operations, including usage monitoring and future resource prediction. It is also very satisfactory to see the good coordination with LHCb's physics analysis coordination in order to manage the use of additional resources as they become available.

C-RSG has thoroughly reviewed the LHCb 2019 request. The requested WLCG resources are deemed to be correct to allow the proper processing of the Run2 data, making it available to collaboration members for physics analysis. The resources also allow for a basic set of simulations needed by the main analyses, as prioritized by the collaboration's physics coordination structure. LHCb plans to continue to identify and use additional resources when available, in order to perform simulations to improve existing analyses or enable additional ones. C-RSG also considers that LHCb has an appropriate contingency plan in case more data is accumulated in 2018 than planned.

C-RSG takes note of the needs reported by LHCb for analysis and data preservation as well as for participation in CERN's Open Data project. C-RSG encourages CERN to continue making these resources available to LHCb,

C-RSG recommends the granting of all requested resources for LHCb in 2019. C-RSG advises to take into consideration the 2020 outlook provided by LHCb, which will be updated and fully reviewed in one year.

14 Comments and recommendations

- Data popularity plot monitors only part of the disk space and the EGI accounting numbers are available only for Tier-0 and Tier-1, C-RSG would require to monitor also the Tier-2 disk space usage starting from the next scrutiny.
- CERN is expecting that Tier-0 resources pledged will not increase until Run 3 when the needed computing power and storage space will be provided. The purchasing profile can be different for different Funding Agencies. C-RSG would request that the experiments for the October scrutiny provide requests assuming:
    - the purchasing will be delayed until the begin of Run 3;

- the resources are provided year by year during the shutdown.

- The CRSG requests that CERNâĂŹs expectations for the computing resource model for Run 3 (e.g. the available resources at CERN as a function of time) be documented and communicated to the experiments prior to the start of Run 3.

- The C-RSG commends the experiments for having started the activity on long term data preservation and their involvement on making data available for education through Open Data Portal. C-RSG expects these activities continue and become part of the collaborations computing projects. C-RSG recommends that these projects are discussed within WLCG and the experiments in order to have a structured projects with the appropriate funding.

- The C-RSG appreciates the continued work by the experiments on increasing the computational efficiency of their work-flows and simulations, and on reducing the CPU and disk resources required to addressing the increase in the luminosity of the LHC. The C-RSG hope that all the experiments reach the same level of involvement in improving the computational efficiency and reducing the data size format.

- The C-RSG would request that as part of future resources assessments the experiments provide a proposed mitigation strategy to address changes in the assumed running conditions for the experiment (e.g. pile-up, luminosity or other effects) at the level of a 20% increase.

- As in the past, some experiments have been particularly successful in securing non-WLCG CPU resources, C-RSG encourage all experiments to pursue this. Experiments are encouraged to gather accounting information on the use of non-WLCG resources in order to be able to report them in future to the C-RSG. Furthermore we welcome the fact that every experiment has made use of their HLT farms to augment their cpu resources.

References

[1] ALICE Collaboration. ALICE Computing Resources Usage in 2017-2018 and Resource Requirements for 2018, 2019 and 2020.

[2] ATLAS Collaboration. ATLAS Computing Status and Plans. Report to the C-RSG, April 2018. 10-Feb-2018.

[3] CMS Collaboration. CMS Software/Computing Resource Utilization & Request 2017-2019, Spring 2017. February 2017.

[4] LHCb Collaboration. LHCb Computing Resource usage in 2017. Technical Report LHCb-PUB-2018-002. CERN-LHCb-PUB-2018-002, CERN, Geneva, Mar 2018.

[5] LHCb Collaboration. LHCb Computing Resources: 2019 and 2020 requests. Technical Report LHCb-PUB-2018-001. CERN-LHCb-PUB-2018-001, CERN, Geneva, Mar 2018.

[6] EGI. EGI Accounting Portal.

[7] CMS Collaboration. Rolling report on CMS CPU efficiency, Spring 2017. February 2017.

[8] CERN. LHC Schedule.

[9] I. Bird B. Panzer-Steindel. Private communication, C-RSG of March, 7 2018 and following discussions.