October 2 , 2017

ALICE Computing Resources: Usage in 2016 & 2017 and Resource Requirements for 2018-2019

1. Introduction

This report contains the computing resources used in *2016*, update on resources used in 2017 and the requirement for *2018* for the ALICE Collaboration. Compared to our previous report we have reduced our expectations for raw data volume and consequently the requirements for tape based on reduced event size resulting from use of new gas mixture in the TPC and a significant improvement in HLT compression. Since number of recorded events did not change, other requirements remain consistent with previously presented documents.

2. LHC Run Conditions and Assumptions

In Table 1 we list the historic and assumed future LHC run conditions, specifically the live beam times. We have used the latter to determine our resource requirements. We have obtained these from [1] and [2].

Running conditions in 2016

ALICE resumed data taking in stable beam mode end of April 2016 following an intensive commissioning phase of the new TPC readout electronics (RCU2).

ALICE was operating with pp luminosity leveled to 2.6×10^{30} $cm^{-2}s^{-1}$ and limiting the pileup to μ ~ 1%. This mode reduces the track multiplicity due to pileup in the central barrel and consequently minimizes the space charge induced distortions in the TPC that were first observed during 2015 data taking.

During the early phase of data taking, a new cluster finding algorithm running in the HLT was commissioned with uncompressed data with raw event size much larger than the design value. The HLT compression mode was validated and switched on in June 2016.

For the p-Pb at 5.02 TeV running period, ALICE operated at low interaction rate with a trigger scheme optimized to enrich the data set with minimum bias events. For the higher energy at 8 TeV and higher luminosity, the trigger scheme was optimized for rare triggers.

All of the 2016 objectives in terms of statistics have been reached. The data taking efficiency of the experiment was increased to 95%. The total amount of data collected in 2016, stored at T0 and replicated once in the T1s represents 7.5 PB - 6.5 PB of p-p data and 1.0 PB for p-Pb data. (see Fig.3).

Running conditions and resources use in 2017 (April to August)

ALICE started data taking in stable beams mode at the end of May 2017 with the restart of LHC operation with pp collisions at 13 TeV. During the long end of the year end technical stop a few consolidation operation have been performed for most of the detectors. The TPC has been filled with a new gas mixture (Ne-CO_2-N_2) to suppress the strong space charge distortions observed in 2015 and 2016.

ALICE is operating with pp luminosity leveled to 2.6×10^{30} $cm^{-2}s^{-1}$ providing an interaction rate of 150 kHz and limiting the pileup to μ ~ 1%.

The TPC readout rate is set to 430 Hz. The trigger scheme selects minimum bias events, high multiplicity events, muon events, diffractive events, calorimeters and TRD triggered events. The efficiency of the experiment remains above 93%.

In the requirement document presented in April 2017 we based on estimates for data volume in 2017 and 2018 on larger event size of 2.87 MB as observed during the 2016 pp data taking. By changing the TPC gas mixture from Ar-CO_2 to Ne-CO_2-N_2, mainly intended to reduce the large distortions observed during the 2015 and 2016 pp run, the amount of noise and the probability to split space-time clusters were also drastically reduced. In combination with additional improvements in HLT compression, the average event size at nominal data taking conditions is now reduced to 1.7 MB.

The total amount of data collected until end of August 2017, stored at T0 and replicated once at the T1s, represents 1.6 PB. (see Fig.1), for a total of 3.2 PB.

The goal is to reach by the end of the pp run at 13 TeV an integrated luminosity of 14 pb^{-1} to add to the already acquired data amounting to 15 pb^{-1}. The final objective of Run2 is to reach 40 pb^{-1}. A special run of pp collisions at 5.01 TeV is scheduled towards the end of the year. The required statistics can be collected in 6.7 days of stable beams.

Table 1: *Assumptions on lifetime for LHC running from 2015 to 2018, including ALICE data taking efficiency.*

RRB year	High pp /10^6s	HI or low pp /10^6
2015	3.1	0.7
2016	5.2	0.7
2017	6.8	0.6
2018	8.0	1.2

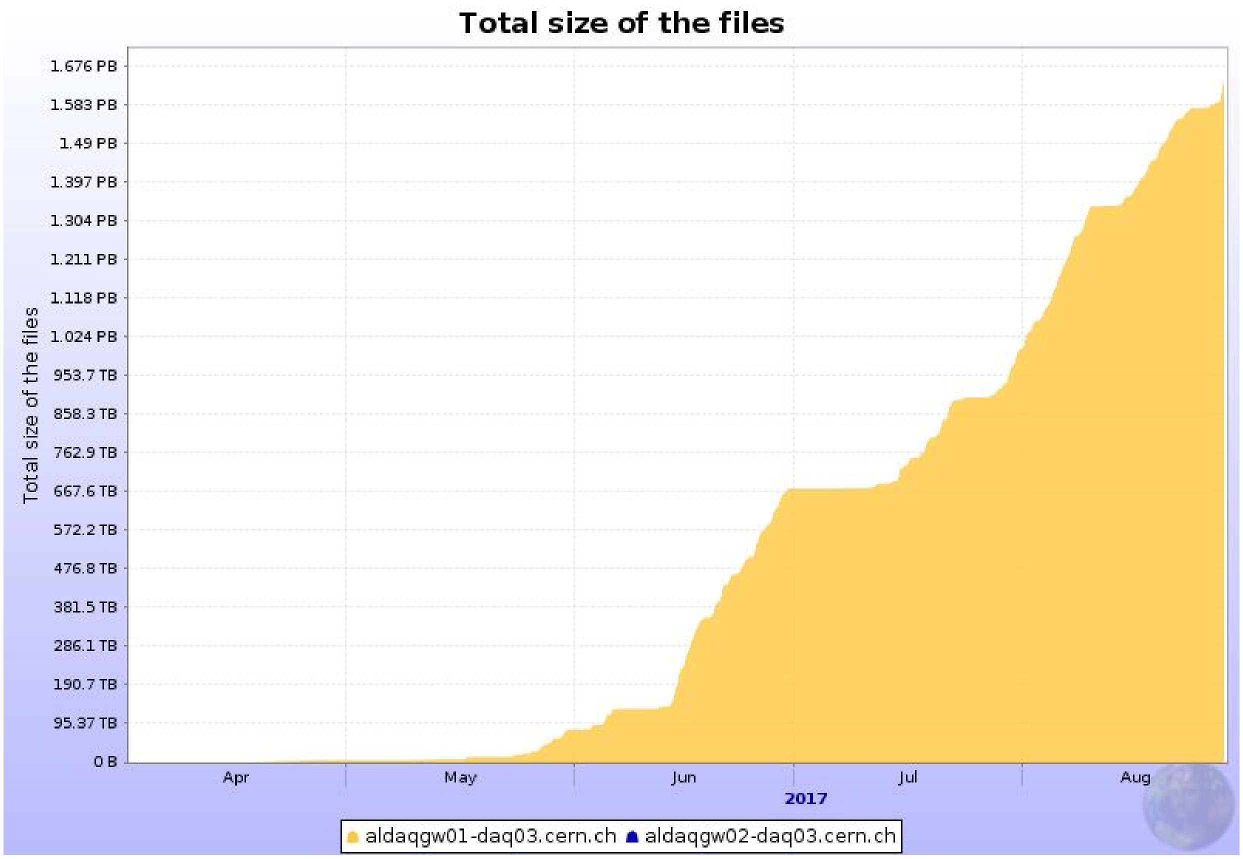


Figure 1: *Raw data collection profile in 2017.*

The CPU usage in Wall kHEPSpec, from 1 April to 30 August 2017, is shown in Table 2. The efficiency at the various tiers remains unchanged with respect to the 2016 levels and is lower at CERN due to a larger proportion of RAW data calibration and reconstruction jobs, which are inherently less CPU efficient than MC. ALICE was able to use opportunistically more than the pledged resources at all tiers. This allowed us to complete the reconstruction backlog of RAW data from 2015 (completed at 80%) and 2016 (completed at 100%) and a good fraction of the associated MC simulation, thus making this data available for physics analysis.

In Table 2 we summarise the current status of disk deployment in 2017, compared to the pledges and the requirements from REBUS [3, 4]. The T2 disk capacity includes 1.74PB (8% of total T2 capacity) of unpledged resources, which are however provided to ALICE. As in the previous reports, we would like to stress that there is a substantial deficit (30%) in the T2 pledges with respect to the required disk capacity. On a positive note, we note that most of the pledges for 2017 are already installed and there is a slight surplus at T0. The latter is due to a small set of old active servers, which will be retired later in the year.

As of end August 2017, the occupied disk space is 53 PB (including the disk buffer for MSS), which amounts to 83% of the installed disk capacity. This relatively comfortable situation is a result of a very aggressive disk cleanup, described below.

Table 2: *Summary of CPU, disk and tape usage in 2017 (April to August) compared to the pledges for the year. Note that we have taken the pledged and requested resource data from REBUS [3, 4], and we have taken our used resources from the EGI accounting portal [5] and MonAlisa accounting portal [6]. For deployed disk, the numbers in brackets denote the buffer in front of the tapes. The disk usage is reported in the text.*

Resource	Site(s)	2017				
		requested	pledged	used	used/pledged	CPU efficiency
CPU (kHS06)	T0	292.0	292.0	403.2	1.38	0.73
	T1	256.0	235.5	294.8	1.25	0.84
	T2	366.0	277.7	290.7	1.05	0.83

				deployed	deployed/ pledged	deployed/ required	used/deployed
Disk (PB)	T0	22.4	22.4	21.0	1.08	1.08	0.63

				(3.3)			
	T1	25.4	21.8	16.7 (3.1)	0.91	0.78	0.97
	T2	31.4	22.8	21.8	0.95	0.69	0.91
Tape (PB)	T0	36.9	36.9				0.71
	T1	43.3	30.6				0.64

In Figure 2 and Table 3 we present the current data popularity numbers. The significant reduction of unpopular data with respect to the 2016 status is a result of a new round of productions cleanup, both MC and RAW data. In total, 3 PB of data was deleted and additional 4 PB of ESD replicas were removed. Furthermore, about 0.5 PB of user data was deleted by introducing strict and lower disk quotas for individual users. The total space gained is 7.5 PB, still insufficient to offset the ever widening 9PB deficit in T2 disk pledges. This action, although very productive, is only delaying the potential disk space crisis at the end of 2018 and beginning of 2019, when the large amount of Pb-Pb data and associated MC will be collected and will need to be processed and analyzed. The remaining 3 PB of data not accessed in the past year (5% of deployed disk) are mostly productions still used to ongoing analyses for future publications and a few reference productions. The majority of these productions have been reduced to a single copy on disk. The status of both classes are continuously evaluated and productions no longer needed on disk will be removed.

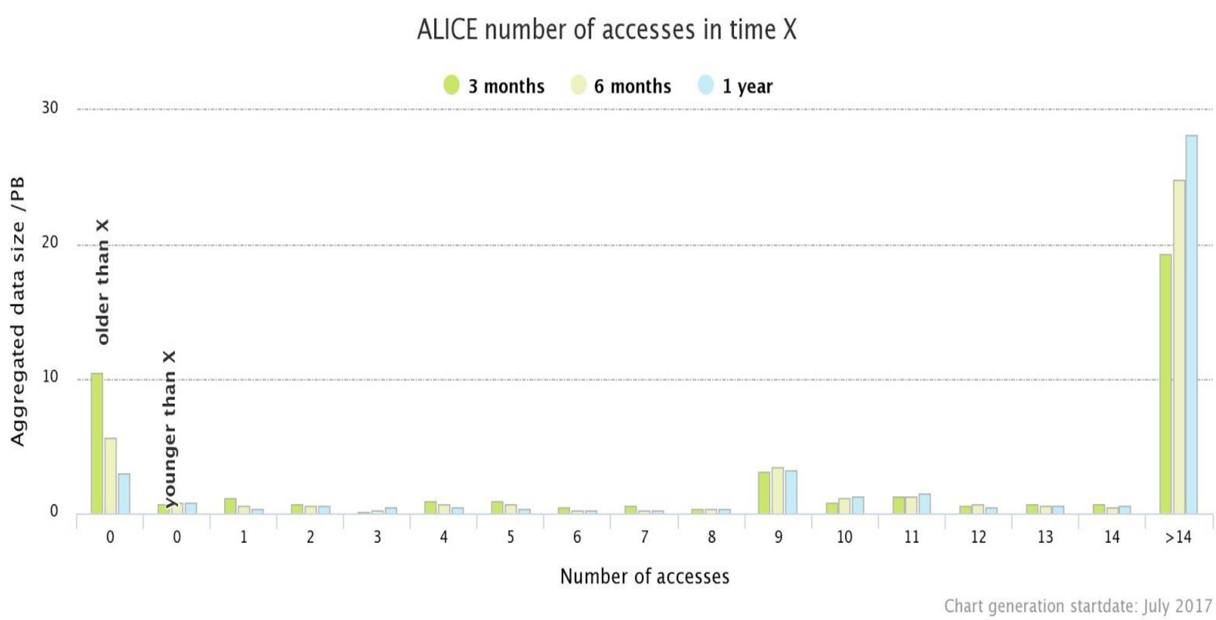


Figure 2: *Volumes of data versus number of accesses in 3-, 6- and 12-month periods - status in July 2017. For each period X, data created in that period but not accessed is in the second bin. The first bin is for data created before the* period *began and not accessed during that period.*

Table 3: *Volumes of data (in PB) versus number of accesses in 3-, 6- and 12-month periods - status in July 2017. For each period X, data created in that period but not accessed is in the second bin. The first bin is for data created before the period began and not accessed during that period.*

Acc	0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	>14
3 m	10.5	0.68	1.16	0.69	0.15	0.94	0.86	0.41	0.54	0.30	3.06	0.85	1.29	0.59	0.68	0.66	19.3
6 m	5.7	0.83	0.56	0.58	0.26	0.72	0.66	0.24	0.27	0.40	3.45	1.12	1.31	0.70	0.57	0.47	24.8
1 y	3.0	0.86	0.31	0.63	0.43	0.51	0.31	0.27	0.19	0.39	3.16	1.30	1.51	0.47	0.52	0.56	28.2

3. Computing Usage in 2016

Comparison of Requested, Pledged and Usage

In Table 4 we summarize the computing resource usage by the ALICE experiment for the period April 1st 2016 to March 31st 2017. Note that we quote the pledged and requested resource data from REBUS [3, 4], and we have taken our used resources from the EGI accounting portal [5] and/or from the MonALISA portal [6]. Sources [5] and [6] agree to a large extent and the information taken from any of these is complementary.

Tape Usage

After successful pp and Pb-Pb data taking in 2015, ALICE started the 2016 data taking in April 2016 using upgraded TPC readout. The initial data taking was done without compression, in order to commission the new readout and to allow for tuning and validation of the HLT cluster finding algorithms. The HLT compression was switched on in June.

The tape usage is dominated by RAW data recording and is displayed on Figure 3 for the entire Run2 period until the end of 2016. The data collected at T0 in 2016 amounts to 7.5 PB for a cumulative total of 25.5 PB exceeding by 18% the 2016 pledged resources. The accumulated data at the T1s storages amounts to 18.5 PB exceeding by 4% the 2016 pledged resources. The used tape storage corresponds to 97% and 91% of the required and CRSG approved storage in T0 and T1s, respectively.

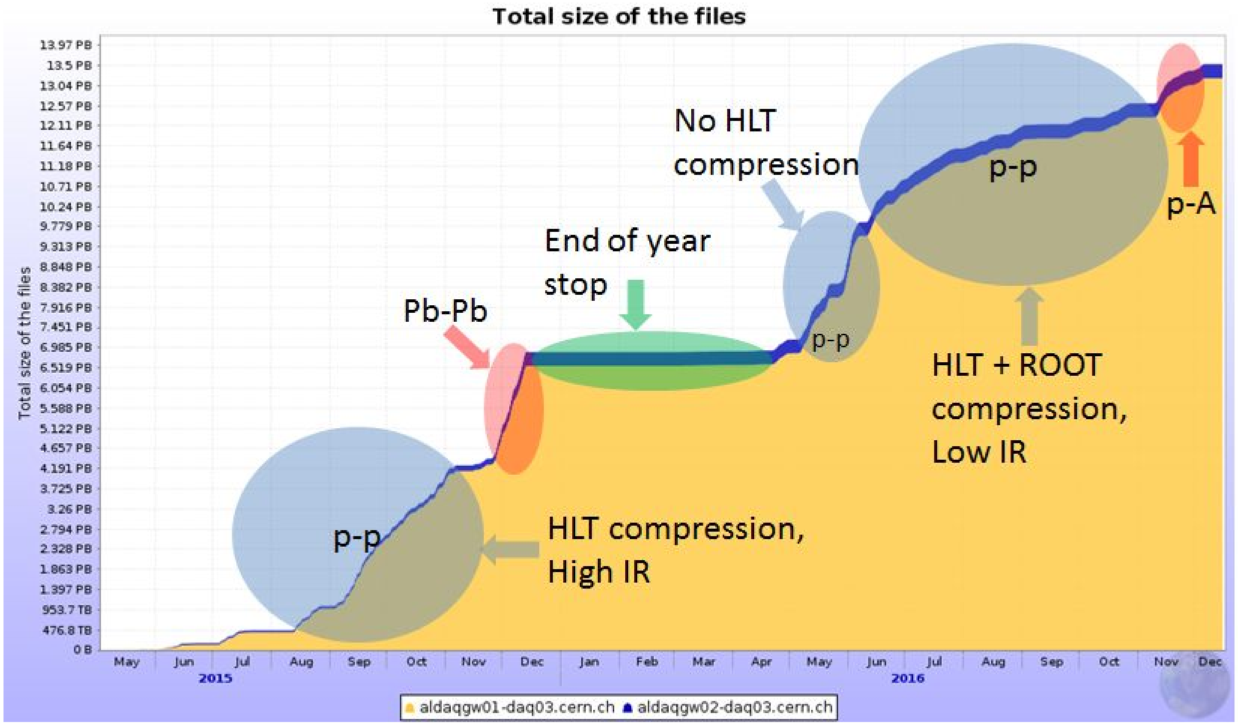


Figure 3: Raw data accumulated during Run 2.

Disk Usage

The disk usage in 2016 is summarized in Table 4. As agreed with the sites, the installation of new disk resources follows closely the usage throughout the year and is moderated by the continuous removal of unpopular datasets and reduction of replicas for the same whenever possible.

The disk storage resources are distributed as follows: 23% in T0, 33% in T1s and 44% in T2s, similar to the values reported in the previous accounting period. In total 92% of the 2016 pledged resources have been used. Note that 66% of 2017 pledged disk resources are already used and this is further exacerbated by the insufficient pledges at the T2s, where the occupancy of the storage is getting close to 80%.

Table 4 also summarizes the delivered and pledged disk resources in the T0, T1s and T2s. In general, all pledges for 2016 have been delivered to ALICE. The remaining deficit at some of the T2s is being discussed with the affected centres. The deficit is worrisome, as there is a continuous trend of under-pledging of storage, despite the ALICE resources being scrutinized and fully approved by all relevant committees.

CPU Usage

The CPU usage in 2016 is given in Table 4 in Wall HEPSpec06 units. All available CPU slots have been used throughout the year and the opportunistic usage of CPU resources has been systematically exploited. The contribution from non WLCG sites amounts to 4% of the pledged resources.

Averaged over the year, 563 kHEP-SPEC06 were used in total, peaking at 660 kHEP-SPEC06. shared among Tier categories as 28% at T0, 36% at T1s and 36% at T2s.

The additional CPU required for the new TPC distortion calibration (5% of the total ALICE CPU) has been offset by improvements in the reconstruction software requirements. The relative share among tasks is given in Table 5.

The HLT farm of ALICE has been successfully integrated into the Grid and is providing about 5% of the total CPU resources. This capacity has been already subtracted from our requirements in 2016 and 2017.

Table 4: *Summary of resource usage in 2016 (April 2016 − March 2017), with the pledges made for 2016 and the requests made in 2015 for 2016. Note that we have taken the pledged and requested resource data from REBUS [3, 4], and we have taken our used resources from the EGI accounting portal [5] and MonAlisa accounting portal [6].*

Resource	Site(s)	2016				
		requested	pledged	used	used/pledged	CPU efficiency
CPU (kHS06)	T0	224.0	215.0	247.3	1.15	0.72
	T1	145.0	177.0	245.1	1.38	0.80
	T2	235.0	231.2	257.1	1.11	0.82
Disk (PB)	T0	15.9	16.8	16.3	0.97	
	T1	20.3	19.0	19.5	1.03	
	T2	24.7	18.5	14.0	0.76	
Tape (PB)	T0	26.3	21.6	25.3	1.17	
	T1	20.3	17.8	17.1	0.96	

Breakdown of CPU and DISK Usage by Job Type

In Table 5 we list the percentage of the (global) disk and cpu resources used for simulation, reconstruction and data analysis in *2016*. The average share of CPU utilisation between the three main ALICE workflows is as follows: simulation (69%), reconstruction (10%), analysis (21%). This ratio is essentially unchanged from the previous accounting period.

Table 5: *Percentage of the (global) CPU and disk storage resources used for simulation, reconstruction and data analysis in 2016 (i.e. January 2016 − 1 to December 2016).*

Resource	Simulation	Reconstruction	Data Analysis
CPU	69%	10%	21%
Disk	61%	32%	6%

Analysis of Access-Frequency Data

The ALICE data popularity plot (Figure 4) shows the number of accesses in a given time period. For the recently produced data, the volume of data accessed zero times is small and also there is a small percentage of data accessed less than 14 times. This trend is consistent with the data popularity in the previous accounting periods. The high popularity of the current data is attributed to the preponderance of organized analysis, which tends to access complete datasets while minimizing the bandwidth requirements to the storage by grouping together many user tasks.

What appears to be a large volume of data not used for more than one year, is a consequence of the ALICE physics coordination policy that requires that any data used for publication remains available for analysis. If we re-calculate the first bin by extending the period to the last 2 years, the non-accessed data volume is reduced to 2.25PB (0.55PB of RAW data ESDs and AODs and 1.7PB of MC ESDs and AODs). This is 5% of the 44PB total disk available (excluding tape buffers). Consequently, in our resources requirements we have not foreseen tape storage for derived data.

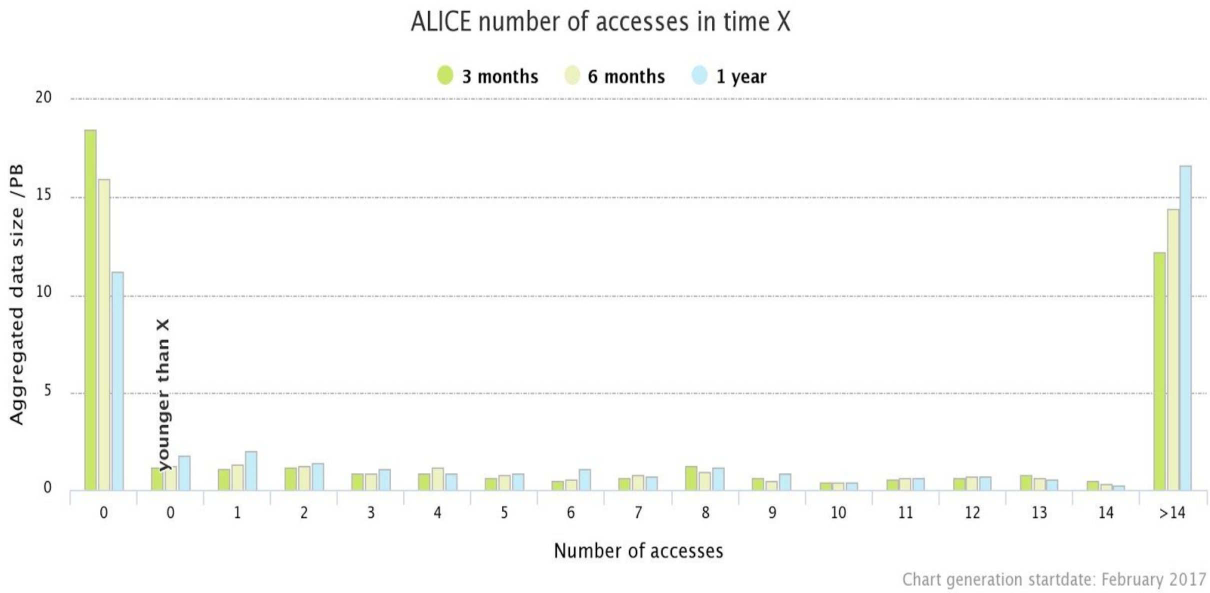


Figure 4: *Volumes of data versus number of accesses in 3-, 6- and 12-month periods. For each period X, data created in that period but not accessed is in the second bin. The first bin is for data created before the period began and not accessed during that period.*

4. Resource Requirements for 2018

Running scenario

The running baseline scenario for the rest of Run 2 is planned to fulfill the goals in terms of statistics for which the ALICE scientific program was originally approved.

For pp running, considering the higher LHC efficiency, ALICE can collect the event statistics as defined in the physics programme by running at a reduced interaction rate..

For the PbPb data taking at the end of 2018, with an anticipated HLT compression of a factor of up to 7.2 and including data from other detectors, we project a total readout rate of 10 GB/s.

The running scenario is summarized in Table 1, and presented as data accumulation curve versus time in Figure 5. It should be noted that the Pb-Pb data which will be collected at the end of 2018 has a large impact on resources requirements in 2019 and beyond. This is discussed later in the document. The data accumulation scenario has been modified with respect to the originally expected in the beginning of 2017. The total amount of tape resources needed in 2017 and 2018 is lowered by 7 PB due to two main factors:

a) The change of gas mixture in the TPC using Ne instead of Ar and resulting in reduction of spurious clusters

b) Improved HLT compression of TPC data from 5.2 to 7.2

These factors have maximum impact and result in lower RAW data volume in p-p running. Note that the use of the Ne gas mixture in the TPC was not yet approved at the time of compiling the 2017 requirements in the beginning of the year, as well as the improved HLT compression algorithm was not yet validated.

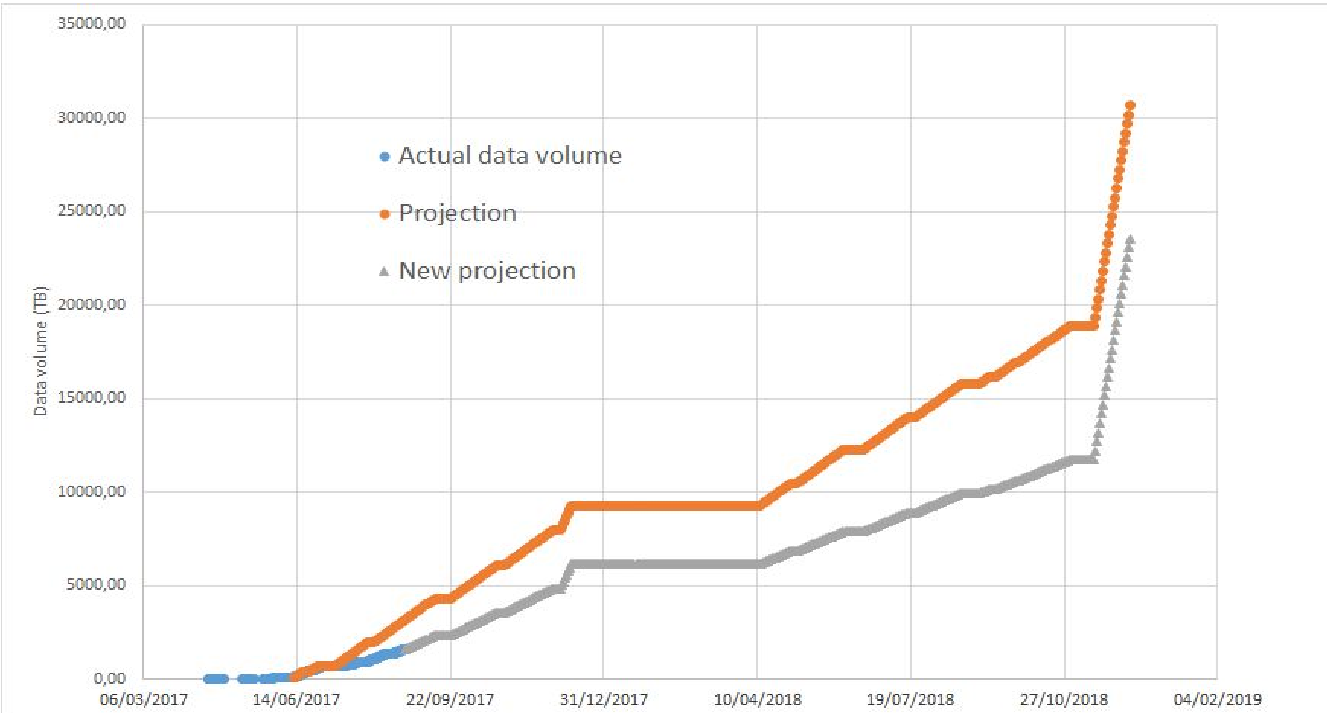


***Figure 5**: A graphical representation of 2017-2018 data taking scenario. The 'New projection' series represents the updated scenario taking into account the smaller event size in p-p.*

Computing model parameters

Within the running scenario discussed previously and with the breakthrough achieved for HLT compression, the average Pb-Pb raw event size is substantially reduced, keeping the tape usage within the projected limit. The size of the derived data, reconstructed and Monte-Carlo, and the processing time have been updated to reflect the sizes achieved during the 2016 processing.

The various event sizes are summarized in Table 6. The RAW event sizes take into account the planned trigger and event modes and projected pileup as well as the currently achieved HLT and ROOT compression.

The updated computing powers needed to process (calibration and reconstruction) one event have been updated taking into account the actual processing time achieved in the

2016 reconstruction, analysis and Monte-Carlo productions and are reported in Table 7. We do not expect to dramatically improve those numbers in the next 3 years.

In Table 8 we list our requests for *2018* together with the pledge made for *2017*.

Table 6: *Event size used to calculate the resources required in 2018*.

system	RAW (MB)	ESD+AOD (MB)	Monte-Carlo (MB)
pp	1.7	0.3	0.5
Pb-Pb	5.0	1.4	2.7

Table 7: *Computing power used to calculate the resources required in 2017-2018*.

system	Reconstruction (kHSO6xs)	Analysis Train (kHSO6xs)	Monte-Carlo (kHSO6xs)
pp	0.19	0.20	0.90
p-Pb	0.23	0.70	0.70
Pb-Pb	0.91	3.70	3.70

Data Volume

For pp data taking, the raw data volume is calculated as the raw data size (given in Table 6) times the effective running time (given in Table 1) times the readout rate (see text). For Pb-Pb the data taking will saturate the DAQ bandwidth of 10 GB/s.

The total data volume to be collected and stored on tape at T0 in 2018 amounts to 17.65 PB: 6.15 PB of pp data and 11.5 PB of Pb-Pb data. These projections have to be taken with caution because of relatively short heavy ion data taking period that yields large amount of data during. All validated raw data will be replicated at T1s. If necessary and in order to alleviate the 2018 tape storage request for T1s, we have considered to postpone the replication of the pp data taken in 2018. The preferred scenario would however be to have all the tape storage in T1s available for pp data already in 2018. This would be beneficial for the prompt data processing of these unique data in view of the Quark Matter conference scheduled in autumn 2019.

The accumulated data volume anticipated at the end of Run2 and reported in Table 8 is obtained by adding to the used resources in 2016 (see Table 4) the 2017 data volume estimated to be about 6.1 PB and the 2018 data volume, estimated at 17.65 PB

CPU Resources

The CPU resources are estimated based on the following processing profile:

- Reconstruction pass 1 is performed in T0 and T1s during the 6 months following end of data taking;
- Reconstruction pass 2 is performed in T0 and T1s during the year following end of pass 1;
- Reconstruction pass 3 is performed in T0 and T1s during the year following end of pass 2;
- Monte-Carlo production is continuously performed over the year in all Tier categories; one pp MC event is produced per raw event and 0.18 AA event per raw event.
- Analysis is continuously performed over the year in all Tier categories; the fraction of analysis is taken as 20% of the total CPU resources (see Table 3).

The total amount of needed CPU resources is averaged over the year and distributed in Tiers to flatten their CPU usage.

An overall increase of 51% of the CPU resources is required in 2019. This steeper increase in CPU requirements is driven by two factors: 1) the reconstruction passes and accompanying MC productions for the large PbPb data sample that will be taken at the end of 2018 and 2) MC productions matching the 2017/2018 data, since the projected resources increase for 2017 and 2018 imply that we cannot perform all the needed MC simulations in those years.

Disk Resources

The disk resources requirements follow closely the requirements for CPU resources. Reconstructed data are stored where they are produced. We keep on disk one copy and one replica of reconstructed and MC data. Only one full pass of reconstructed data and its associated MC events are kept on disk. Earlier passes are kept with one copy only as newer data become available. The required disk resources include 3 PB of disk buffer for the taping system at Tier0 and 3 PB at Tier1s.

Table 8: Resources requested for *2018* together with the pledge made for *2017*.

Resource	Site	2017 Pledge	2018	Growth
CPU (kHS06)	T0	292.0	350.0	0.20
	T1	235.5	307.0	0.30
	T2	277.7	389.0	0.40
Disk (PB)	T0	22.4	26.2	0.17
	T1	21.8	30.5	0.40
	T2	22.8	35.2	0.54
Tape (PB)	T0	36.9	49.1	0.33
	T1	30.6	40.9	0.34

4. Resource Requirements for 2019

During the Long Shutdown 2, the ALICE computing resources requirements are calculated to complete the Run2 data processing and associated Monte-Carlo following the strategy outlined in the previous sections. This scenario assumes that Run2 data will be processed at least once by the end of 2019. The required resources are listed in Table 9.

Table 9: Resources requested for 2018-*2019*.

Resource	Site	2018 request	2019	Growth
CPU (kHS06)	T0	350.0	496.0	0.42
	T1	307.0	465.0	0.51
	T2	389.0	589.0	0.51
Disk (PB)	T0	26.2	30.7	0.17
	T1	30.5	35.8	0.17
	T2	35.2	42.0	0.19
Tape (PB)	T0	49.1	49.1	0.0
	T1	40.9	40.9	0.0

References

[1] I Bird, P Buncic, F Carminati, M Cattaneo, P Clarke, I Fisk, J Harvey, B Kersevan, P Mato, R Mount and B Panzer-Steindel (Eds), Update of the Computing Models of the WLCG and the LHC Experiments, v.2.8
[2] Longer term LHC schedule

[3] REBUS Pledge Summary

[4] REBUS Pledge Fulfillment

[5] EGI Accounting Portal

[6] MonALISA accounting portal

CRSG Q&A, 21 September 2017

1. page 2: here you state that the luminosity in 2016 was leveled to 2.6 e30 cm^-1s^-1, whereas in the April usage report you wrote it was 5.0 e30 cm^-1s^-1

a) which is the correct number - or has there been a change in the cause of the year ?

A: **The running conditions have been changed during the year. For pp running the interaction rate was set to 200 kHz (5.4×10^{30} Hz/cm^{-2}) until week 24, to 100kHz (1.7×10^{30} Hz/cm^{-2}) from week 24 to 35 and 38-43, and to 300-550 kHz from week 35 to 26. It averages out to 2.6×10^{30} Hz/cm^{-2} and our tape usage is not proportional to the IR, since the trigger readout rate is levelled.**

2. page 2: what is the definition of the parameter μ=1% for the pileup ?

A: μ is the average number of pp interactions per bunch crossing . The average number of pile-up events (PE) is the zero-suppressed (no interaction) mean of a Poisson distribution P(n) with mean = μ: Average PE = μ/(1-P(0)) = μ/(1-exp(-μ)). For μ=0.01, the average PE = 1.005

3. page 2: you write that you had based your requirement for 2017/2018 on the larger event size of 2.87 MB/evt, whereas it is now only 1.7MB/evt due to the new TPC gas mixture and the HLT compression - in the April usage report you wrote:

The compression mode has been effective from June 2016, the raw data size design value (1.5 MB/event) has been reached and stabilized.

A: This statement is correct for 2016 data taking period from June to September and given trigger configuration taking into account the total number of events, which include barrel detectors triggers and muon only triggers. The interaction rate was 100 kHz in 2016.

a) why is the compressed raw event size now only 1.7 MB rather than 1.5MB ?

A: This is because the interaction rate in 2017 was increased from 100 kHz to 150 kHz, increasing therefore the pile up and consequently the raw event size.

b) in particular when the TPC gas change should have contributed to the reduction and was not mentioned in the April report ?

A: The possible change of the TPC gas has been discussed during the winter shut down. In April we had no empirical indication of what the reduction in the raw data size would be and we were ready to go back to $\mathrm{Ar}\text{-}\mathrm{CO_2}$ in case if beam tests with $\mathrm{Ne}\text{-}\mathrm{CO_2}\text{-}\mathrm{N_2}$ TPC gas mixture to do to result in reduced distortions and do not show that detector can be operated at sufficiently high interaction rate.

c) do I understand this correctly that the tape requirements for 2017 and 2018 that we scrutinized in previous rounds of the CRSG were based on an estimated raw data size that was 70% higher than it actually is ?

A: Yes this is correct for pp data taking. The data size we had taken previously was the empirical data size observed during the 2016 data taking period. At the time of preparing the previous report we had no indication on what the data size would be with the new TPC gas and the the new HLT compression algorithm. However, in our request we included the mitigation scenario where pp data taken in 2018 would not be immediately replicated in order to save tapes at T1s.

d) what is the change in the event size for PbPb raw events due to the compression and new gas mixture ?

A: No change has been considered in the present requirements as compared to the empirical data size observed during the previous Pb-Pb data taking period. The data taking in PbPb central collisions will be limited by the available DAQ bandwidth of 10 GB/s which we expect to saturate.

4. page 3: up to Aug. you collected 1.6 PB of raw data that leads to a total of 3.2 PB of raw data

a) how much more do you expect this year ?

b) is it roughly the same again (14pb^-1 vs. 15 pb^-1 already collected) ?

A: The numbers can be read from fig. 5. Until the run of the pp run we plan to collect a total 5 PB. The low energy pp run towards the end of the year will add on top 1.2 PB.

5.page 4: in the text you write that you have occupied 53PB of disk amounting to 83% of installed

a) can you add the corresponding numbers to table 2 (i.e. add a used column) ?

A: Done. The numbers are those observed at the time of writing this answer.

b) can you add the corresponding numbers for tape to table 2, i.e. what is pledged/installed and used up to now ?

A: Done. The used tape storage at the time of writing this answer amounts to 26.21 PB at T0 and 19.62 PB at T1s.

c) page 4: what is the definition of the numbers for CPU in table 2 ? Is the used 'average CPU power' over the observed time ?

A: Correct.

6. page 4: says 80% only of the 2015 data reconstructed, but 100% of 2016 reconstructed. Is there a reason that the most recent data was reconstructed before the older data?

A: The priority was mostly technical - part of the 2016 data was already staged on disk from tape and it is favourable for the calibration process to be continuous, period after period. To process the 2015 data, the tapes had to be re-spinned anyway.

7. page 4, Table 2: why is there a "deployed" column in the Disk part rather than a "Used" column?

A: As requested above, both numbers are in the table now.

8. page 6: How did you fit the data on tape if it was larger than the pledges?

A: Some of the computing centres allocated tapes above the pledged capacity, for example at CERN, CNAF, CCIN2P3 and KISTI.

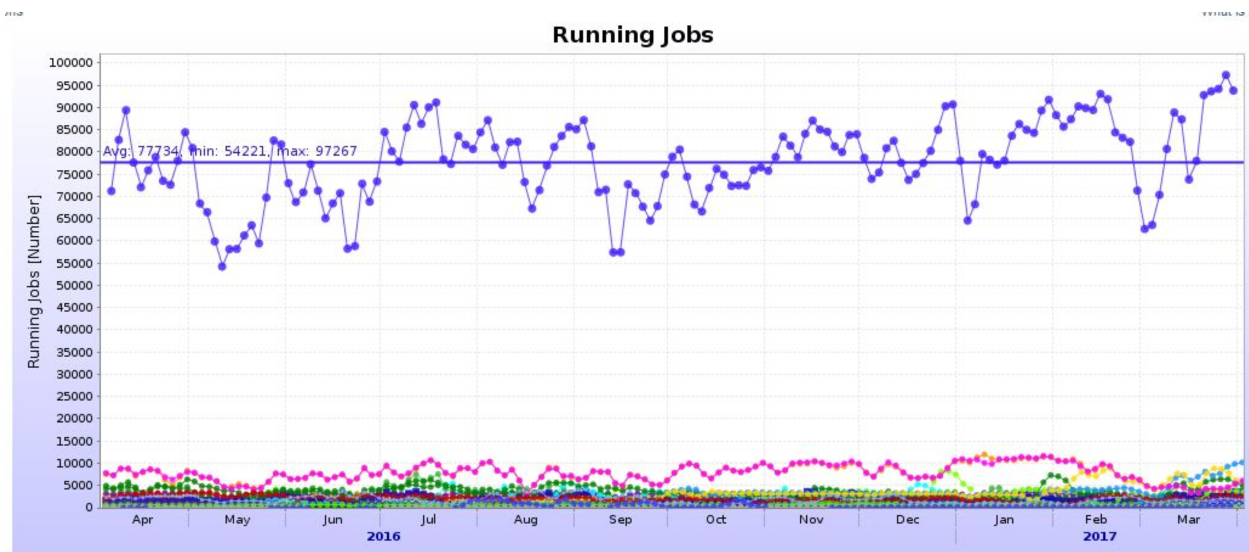
9. page 7, « disk usage », second paragraph: "66% of 2017 pledged disk resources already used « . I don't understand the point

A: The disk used at the time of writing the report represented 66% of the disk pledged for 2017.

10. page 8: you write that 563 kHepSpec06 have been used on average, peaking at 660 kHepSpec06

- why is this not reflected in the 'used' column in table 4 ?
  (see previous question on table 2 (5c))

A: We report average values only, as these are usually compared to the pledges. In the period April 2016 to March 2017, the distribution around the mean is not gaussian, here is a plot illustrating the typical job behavior, proportional to the power, from 1 April 2016 to 31 March 2017



11. page 8, table 4: are disk buffers included as in table 2? What about table 8 and table 9 on page 14? It would help to have a consistent presentation of the numbers across the document.

A: Disk buffers in front of the taping system at T0 and T1s are consistently included in the disk resources numbers, including tables 8 and 9.

12. page 9 . Predrag's covering email says that the ALICE forecast for data volume ON TAPE has been reduced (due to better TPC performance and HLT compression), but "other requirements don't change". But surely the disk requirements will also reduce since disk is also used to (temporarily) store events? p9, table 5 says reconstruction takes 32% of disk.

A: If by "temporarily store events" you refer to the disk buffer in front of the tape system, the compression factor plays smaller role - higher compression factor facilitates the data reconstruction and reduces the tape re-spin (favourable). The buffers sizes are kept at a minimum for optimal tape performance and in several cases are 'indivisible' - single disk server.
If you refer to the output of reconstruction - there is no reduction. The $Ne\text{-}CO_2\text{-}N_2$ gas reduces the number of fake clusters, which are filtered out by the HLT compression algorithm or during the reconstruction phase. These clusters are not written out. The HLT compression reduced the RAW data volume and not the reconstruction output. If anything, due to higher interaction rate the reconstruction output and time is 20% bigger than last year but we choose not to increase our requirements and try to reduce information content of the output in order to compensate for the increase.

13. page 10: you write that you running scenario is based on the goals in terms of statistics for which the ALICE scientific program was originally approved
a) what are these exact numbers for the statistics ?
b) is there a document that shows these approved goals/numbers ?

A: This has been discussed at several occasions with the LHCC, has been documented and summarized in "Feed-back from ALICE on potential mitigation measures for computing resources shortage", which was sent to CRSG with the April report. Quoting this document:

The objectives are as follows:

- **For Pb-Pb collisions:**
    - **Reach the target of 1 nb^{-1} integrated luminosity in Pb-Pb for rare triggers.**
    - **Increase the statistics of the unbiased data sample, including minimum bias and centrality triggered events.**
- **For pp collisions:**
    - **Collect a reference rare triggers sample with an integrated luminosity of 40 pb^{-1}, which is equivalent to the 1 nb^{-1} sample in Pb-Pb collisions.**
    - **Enlarge the statistics of the unbiased data sample, including minimum bias collisions at top energy.**
    - **Collect a reference sample of 10^9 events at the reference energy of 5.02 TeV**
- **For p-Pb collisions:**
    - **Enlarge the existing data sample, in particular the unbiased events sample at 5.02 TeV.**

14. page 12, table 6: what is "ESD+AOD" ?

A: ESD (Event Summary Data) is output from reconstruction, containing still sufficient detector and tracking information to perform post-processing, for example to fine tune the calibration of a detector. It also contains information allowing for all types of ALICE physics analysis. AOD (Analysis Object Data) contain only filtered physics information, on which the majority of physics analysis can be done.

15. page 13: “one pp MC event is produced per raw event” and in Table 6 the event size of the MC event is two or three times smaller than the raw events. So why are the disk requirements for Simulation (i.e. MC?) in Table 5 61% of the total?

A: Actually, for the calculation of the required disk, we use the ESD+AOD size (see table 6). The MC ESD+AOD is 60% (for p-p) and 52% (for Pb-Pb) larger than the RAW data ESD+AOD. We do not keep RAW data on disk, with the exception of the disk buffers in front of tapes and we do not keep the simulated RAW data. In addition to the general-purpose MC, we run set of physics-specific MC simulations including those with different MC generators which add up to the 61% of the total occupied disk space.

16. page14: in table 8 your tape requests for 2018 are only very slightly smaller than in the April CRSG request (49.1 vs 55 and 40.9 vs 41)

a) why is the reduction not of the order of the change in raw event size (-40%) from TPC gas and compression ?

A: As written in the text, the reduction in RAW data size is only for p-p data. In the April report, we proposed that the tapes at T1s are purchased in 2019 instead of 2018 as a cost saving measure and our request for 2019 was flat at T0 and increased at T1s. For simplicity, we now ask for the necessary tapes in 2018. If there is need to postpone the purchase of tapes, we can again lower the request in 2018 at T1s and add the needed capacity in 2019. This means that pp data taken in 2018 would be replicated to T1s only in 2019.

b) as the numbers for 2017 were also estimated with a much larger event size, there should be unused tape capacities that can be used for the 2018 running (and additionally reduce the increase for 2018)

A: This is correct and is taken into account in the tape requirements for 2018 (table 8).

17. table 8, page 14: we can of course go and check but could you document precisely the impact of the change of gas mixture in the TPC and the improved compression of TPC data on 2018 and 2019 requests for CPU, disk and storage?

A: As explained in the answer to Q.12, the gas mixture and the HLT compression only reduces the raw event size and therefore only impacts the required tape resources but

not the disk or CPU resources. The tape resources needed in 2018 have been adjusted taking into account the reduced raw event size.

18. And finally: the requests for 2018 and 2019 are very large. Is there a plan B ?

A: There is no compromise possible on tape resources if ALICE wants to fulfill it scientific program in terms of statistics. A lack of CPU and disk resources can be mitigated by extending the data processing (reconstruction & MC) into 2020. This of course would severely impact the timely scientific production and competitiveness of the ALICE Collaboration and will affect negatively the work on the upgrade project. Detailed explanations are given in the "Feed-back from ALICE on potential mitigation measures for computing resources shortage" document, which was sent to CRSG with the April report.

19: p2 ALICE quote the data collected in terms of PB would be possible to have it in fb-1?

A: The integrated luminosity collected during 2016 for the various trigger conditions and the 3 collision systems are as follows:

1. **pp @ 13 TeV (May to October):**
    a. **Di-muon and single muon high p_T: 9.7 pb^{-1}**
    b. **Single muon low p_T: 0.56 pb^{-1}**
    c. **High multiplicity: 4.44 pb^{-1}**
    d. **PHOS photon: 3.2 pb^{-1}**
    e. **Double gap (diffractive physics): 0.47 pb^{-1}**
    f. **Minimum Bias: 0.012 pb^{-1}**
2. **p-Pb @ 5 TeV (November, 10 to December, 6)**
    a. **Di-muon: 3.38 nb^{-1}**
    b. **Minimum Bias FAST: 0.36 nb^{-1}**
    c. **Minimum Bias CENT: 0.18 nb^{-1}**
3. **p-Pb @ 8 TeV (November 18 to November 25)**
    a. **Muon: 8.68 nb^{-1}**
    b. **PHOS photon: 4.52 nb^{-1}**
    c. **EMCAL jet and photon: 3.74 nb^{-1}**
    d. **UPC central: 2.00 nb^{-1}**
    e. **TRD quarkonia and nuclei: 0.45 nb^{-1}**
    f. **High multiplicity: 1.92 nb^{-1}**
    g. **Minimum Bias: 0.03 nb^{-1}**
4. **Pb-p @ 8 TeV (November 27 to December 4)**
    a. **Muon: 12.77 nb^{-1}**
    b. **PHOS photon: 7.47 nb^{-1}**
    c. **EMCAL jet and photon: 7.17 nb^{-1}**
    d. **UPC central: 0.74 nb^{-1}**
    e. **TRD quarkonia and nuclei: 0.36 nb^{-1}**

f. **High multiplicity: 1.39 nb^{-1}**
g. **Minimum Bias: 0.03 nb^{-1}**

20: p5 When data was deleted exactly? Deleted from where, T1, T2 in which proportion?

A: The removal and replica reduction ran from beginning of June to end July 2017. In proportion from the total volume, the breakdown is 20% from T0; 35% from T1s, 45% from T2s. The clean-up logic was to remove data from the storage elements with highest occupancy first, whenever possible.

21: p8 the 5% of the total cpu required for the TPC distortion calibration is now recovered? can it be subtracted from the total need?

A: While the distortions in TPC are indeed much smaller with new gas mixture and comparable to those observed in Run 1, they still exist and remain time dependent therefore the same calibration procedure developed for correcting 2016 data continues to be used.

22: p8 in table 4 used disk @T2 in 2016 is quite less than the pledged (not requested, pledged) can you comment with respect to the 2017 and 2018 requests?

A: This is exactly the issue - the deployed resources at T2s are below the pledges, and we could use only what is actually deployed. This year, the pledges for T2 storage are installed at 95% level, we do not know if the remaining 5% will be installed by the end of the year.

23: p9 we would like to understand why the disk space is used for 61% for simulation. Can we have the breakdown on the disk usage for simulated data, detector data?

A: The output of simulation are MC ESDs and MC AODs, which contain the equivalent information as ESDs and AODs (which are results of RAW data reconstruction) and in addition the original MC information. The data formats used in ALICE are described in Table 1 of *Feed-back from ALICE on potential mitigation measures for computing resources shortage* document . Please refer also to the answer of Q.15

24: p12 ESD and AOD are for data and MC? MC data are much smaller than data, but they occupy 61% of the disk. Can you comment?

A: Yes, the ESDs and AODs have the same format for simulated and RAW data reconstruction. The ESD+AOD from RAW data reconstruction are smaller than these from MC (table 6). This question is quite similar to Q.15 above. Please refer also to the answer therein.

25: p12 table 7 the MC CPU time is for generation+simulation+reconstruction?

A: Ýes, in the simulation workflow, the processes are combined together - generation-transport-digitization-reconstruction.

26: p13 Why the processing needs three passes?

A: In our computing model the first pass follows immediately after data taking and is suitable for detailed QA and basic physics analysis; second pass is done few months later with improved calibration and reconstruction code, more precise tracking and track matching in all barrel detectors, including the final TOF calibration; the third and final pass is usually a year later with final calibration and reconstruction code, including final particle ID tuning and tracking, allowing for most precise physics analysis.

27: p13 Can we have the breakdown in MC-production / MC reconstruction / data reconstruction / data analysis of the CPU usage?

A: MC production is 69% of the total CPU use, of which 75% is simulation and 25% reconstruction. The time spent in simulation is mostly in the transport (55%), detector response calculation (35%) while the remaining 10% is spent in I/O, QA and MC event generators. The RAW data reconstruction uses 10% of the total CPU, analysis (sum of MC and RAW data analysis) uses 21% of the total CPU.