Computing Resources Scrutiny Group Report

Donatella Lucchesi
for the CRSG
CERN, April 25th 2017

Report:
CERN–RRB–2017–056

C-RSG membership

Membership of the CRSG changed for this scrutiny:

C Allton (UK)
V Breton (France)
G Cancio Melia (CERN)
A Connolly (USA)
M Delfino (Spain)
F Gaede (Germany)
J Kleist (Nordic countries)
D Lucchesi (Italy, Chairman)
H Meinhard (CERN, scient. secr.)
D O'Neil (Canada)
J Templon (Netherlands)

German Cancio Melia replaces Tony Cass, thank you Tony for your important contribution over the years.

Thanks to CRSG members for their commitment, to the experiment representatives for their collaboration and to CERN management for the support.

Pledged Resources for 2017

- Given the unexpected large amount of data, 2017 was called as special year. Additional resources were *approved* at the October RRB.
- It was made clear that the agreement among FA, experiments and CERN is that computing for Run 2 stays within *flat budget* → the resources approval was on best effort basis.

$$Pledges\ balance = \frac{total\ offered - experiment\ required}{experiment\ required}$$

	ALICE	ATLAS	CMS	LHCb
CERN CPU	0%	0%	0%	0%
CERN disk	0%	0%	0%	0%
CERN tape	0%	0%	0%	0%
T1 CPU	-8%	-12%	-14%	-4%
T1 disk	-14%	1%	-21%	-6%
T1 tape	-1%	-7%	-24%	-3%
T2 CPU	-24%	-13%	-7%	27%
T2 disk	-28%	-7%	-22%	-30%

This is the starting point.

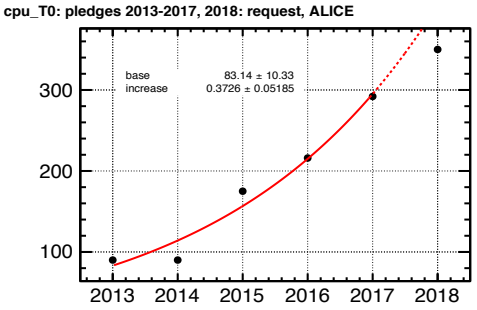
Pledged Resources in the recent years

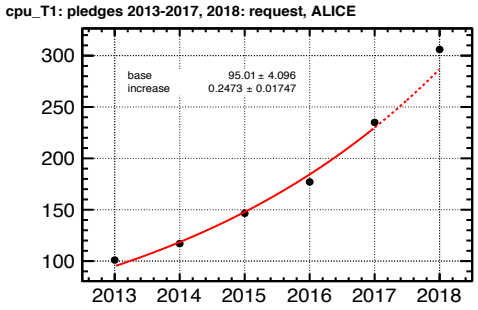
Study of pledged resources

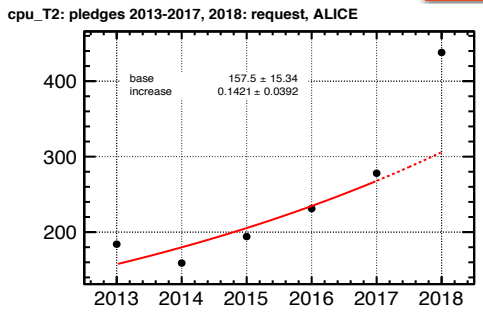
- Plot pledges from 2013 up to 2017 as they are in REBUS as function of the year;
- Fit each plot to measure the actual average increase year per year;
- Display 2018 requests, not used in the fit, and compare them to the fitted value extrapolated to 2018.

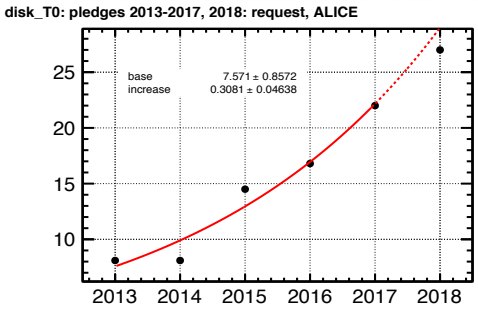
Reminder, flat budget assumptions: 20% increase for CPU, 15% for disk and tape space at constant budget

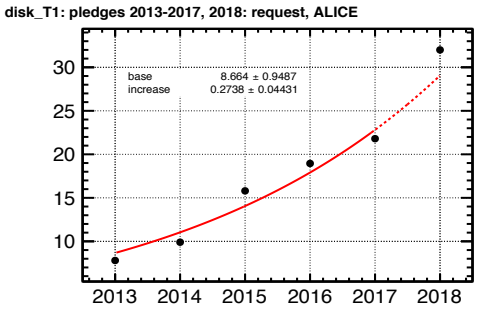
Alice History

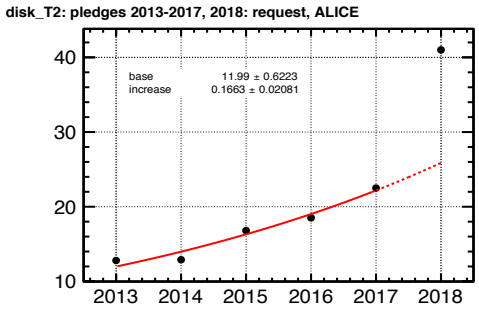




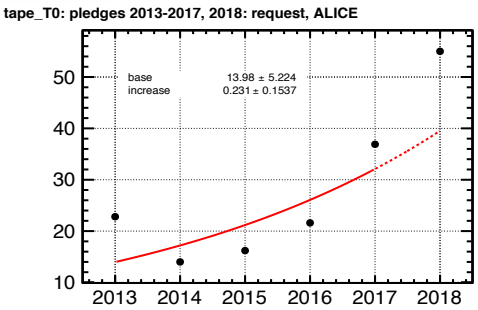


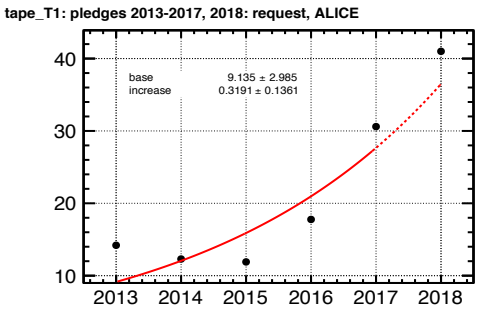




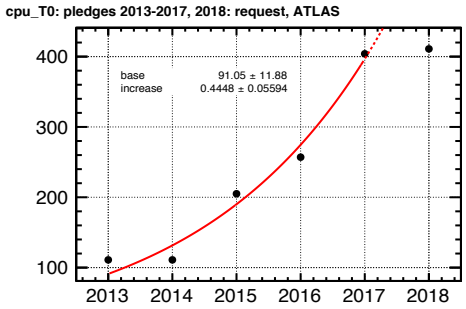


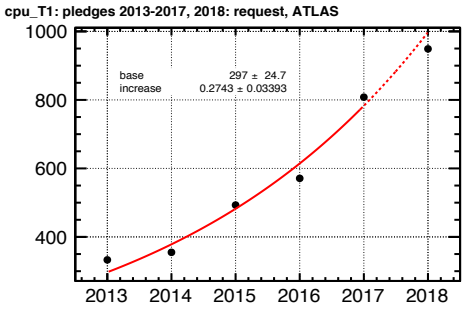
Fit result

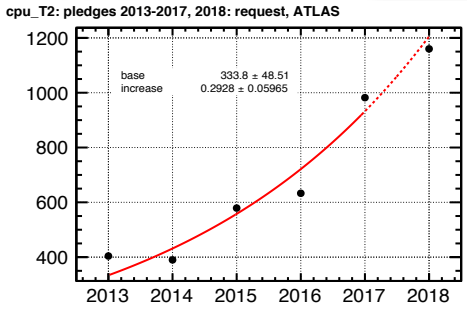


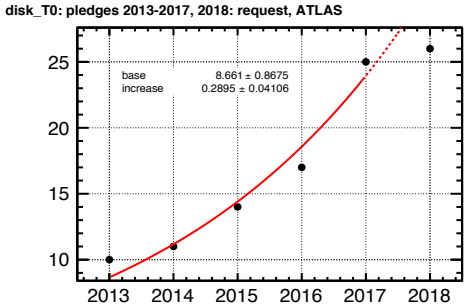


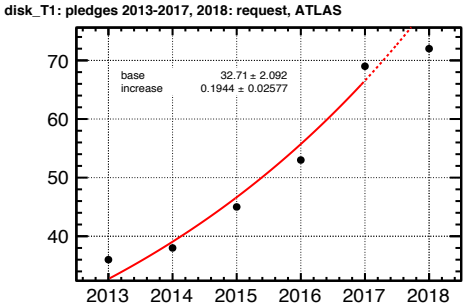
ATLAS History

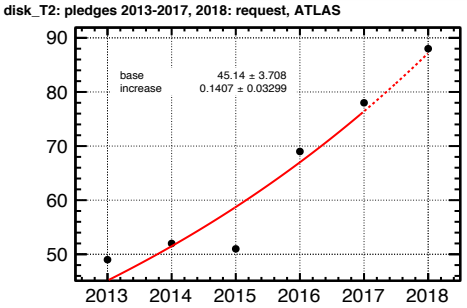




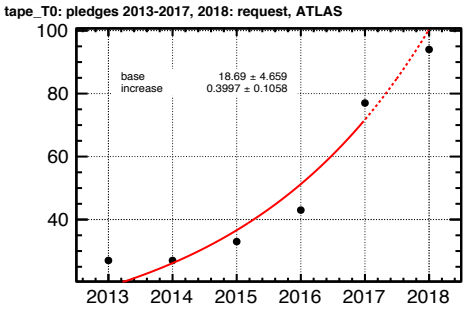


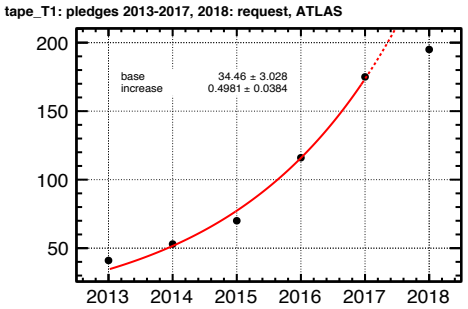




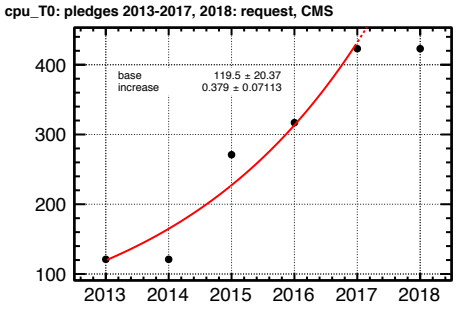


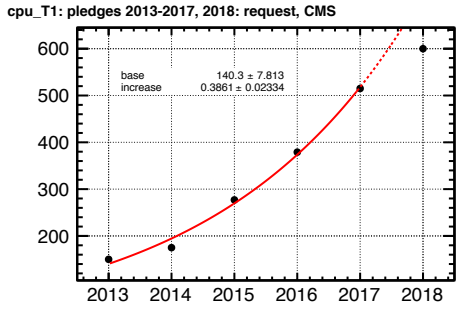
Fit result

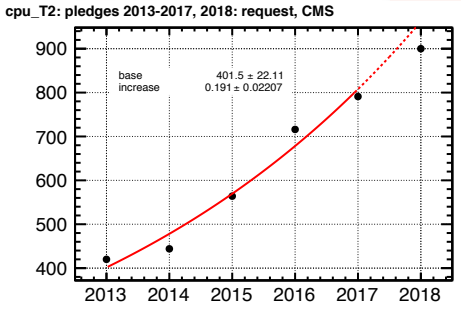


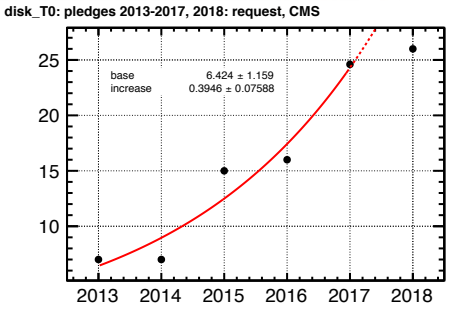


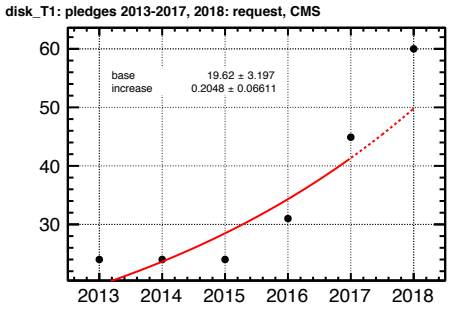
CMS History

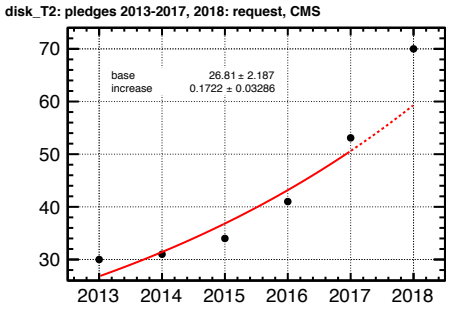




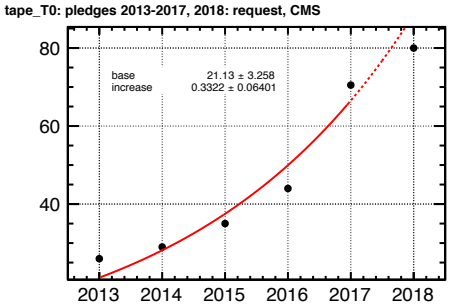


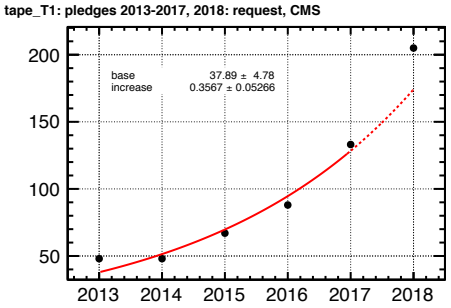




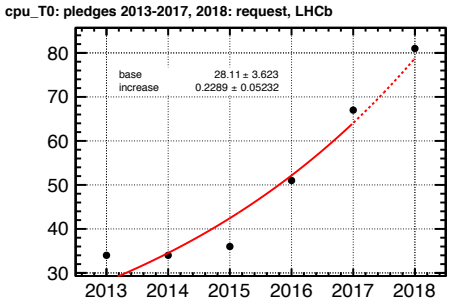


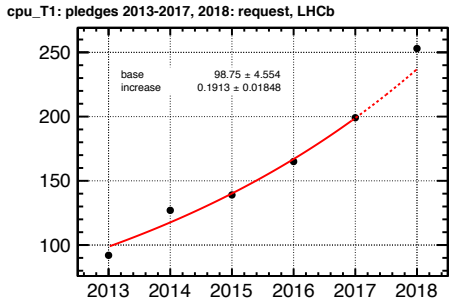
Fit result

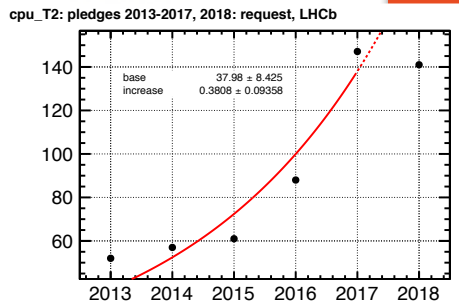


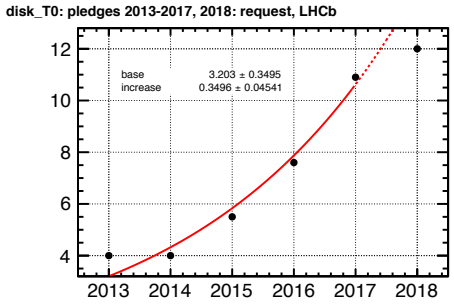


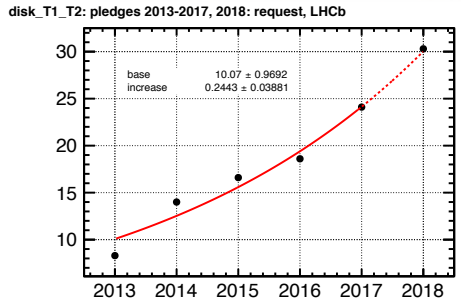
LHCb History



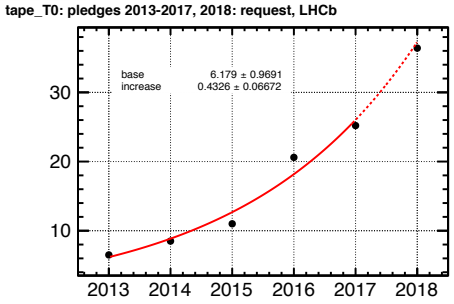


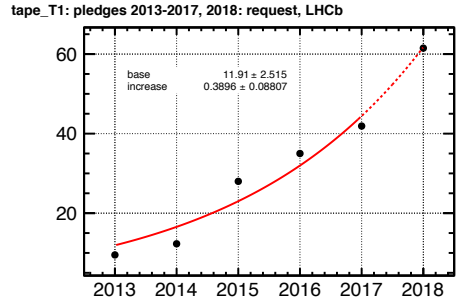






Fit result



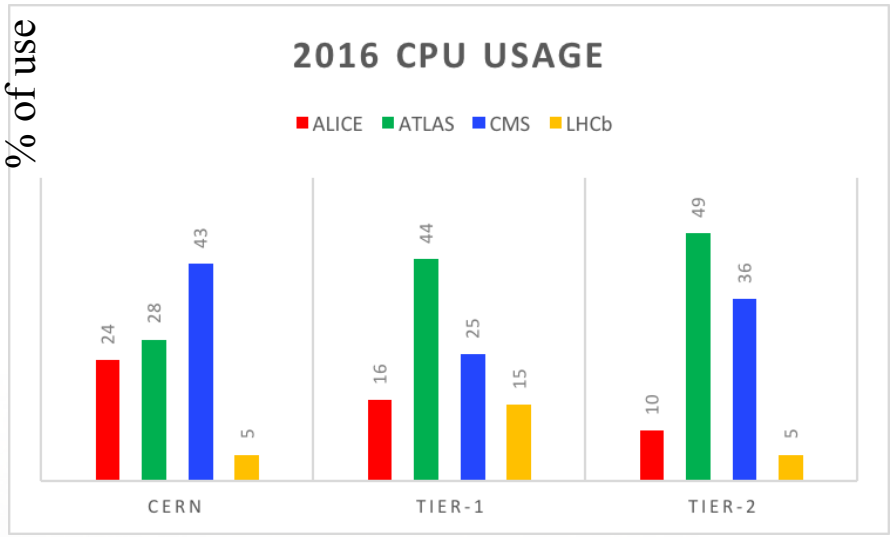


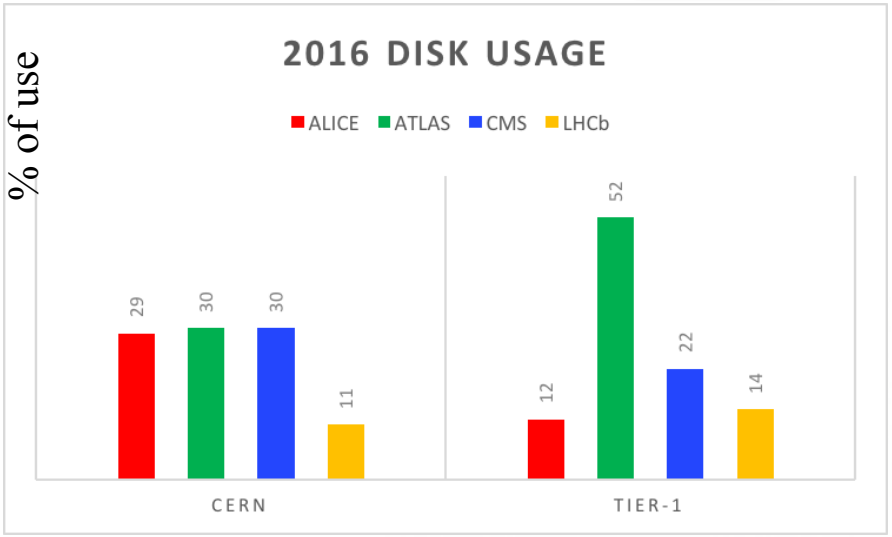
Summary of average increase

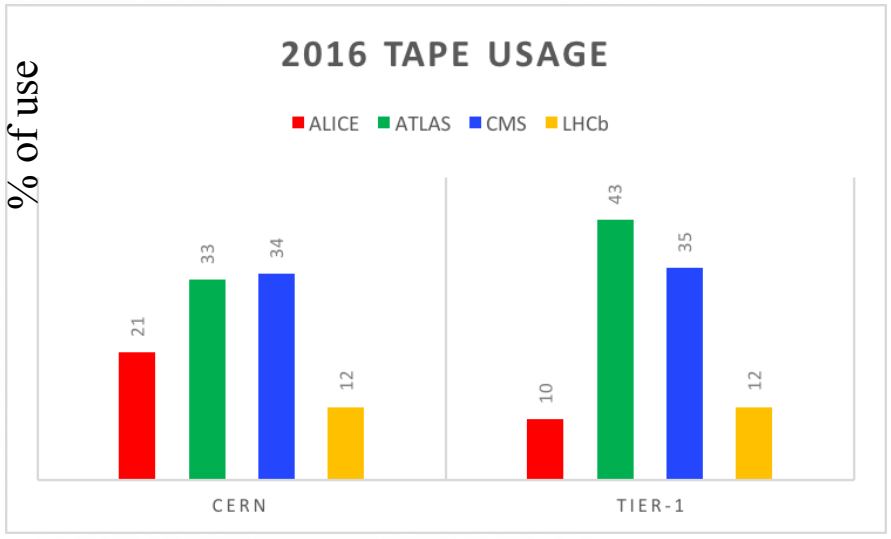
Average increase

	ALICE	ATLAS	CMS	LHCb
CERN CPU	37%	45%	38%	23%
CERN disk	31%	29%	39%	35%
CERN tape	23%	40%	33%	43%
T1 CPU	25%	27%	39%	19%
T1 disk	27%	19%	20%	24%
T1 tape	32%	50%	36%	39%
T2 CPU	14%	29%	19%	38%
T2 disk	17%	14%	17%	24%

Resource Use per Tier in 2016

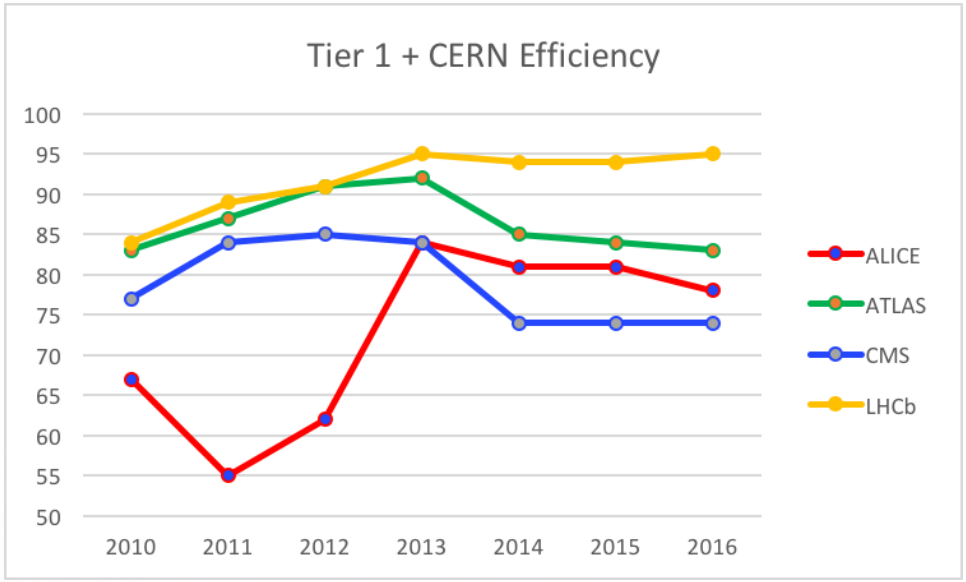


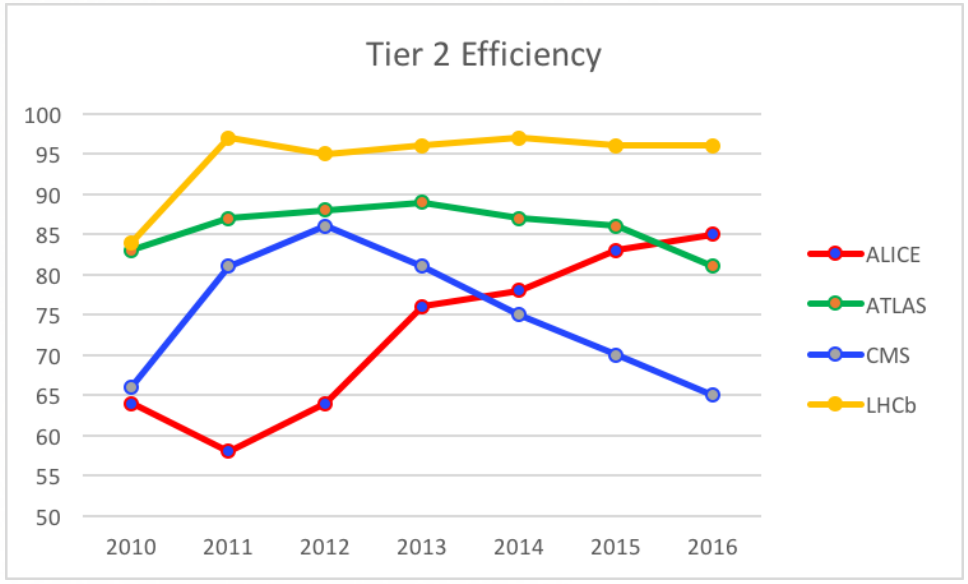




Similar behavior in previous year

CPU Efficiency History

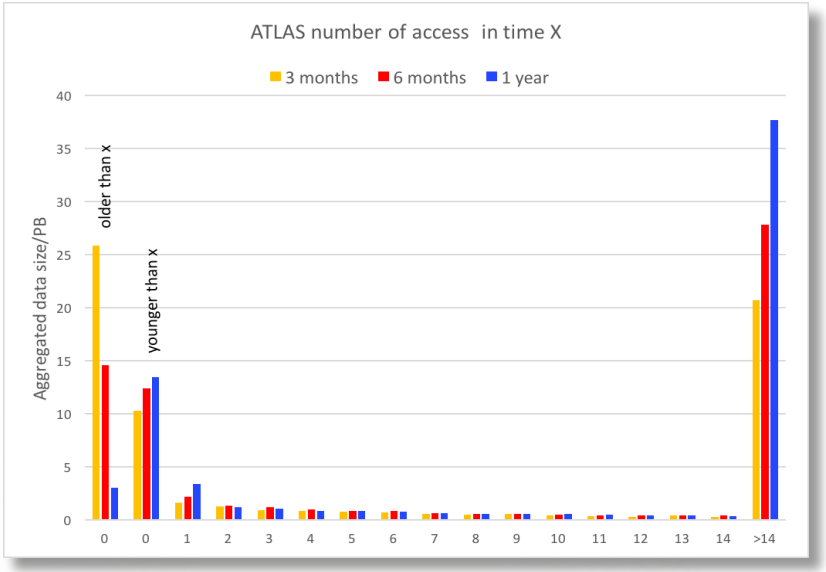


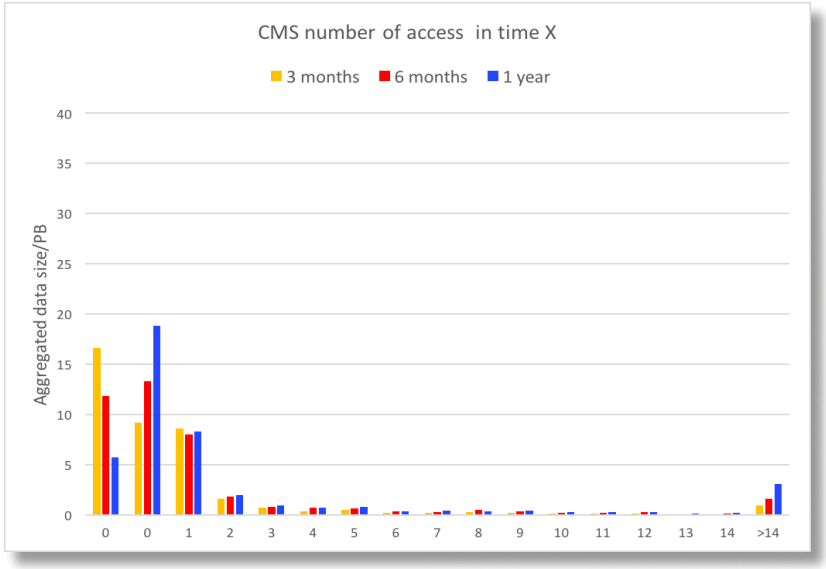


Efficiency: normalized CPU time over normalized wall clock time • nº processors

Data from EGI account portal

Analysis of Access-Frequency Data ATLAS & CMS





Volume of data vs. nº of access in 3-, 6-, 12-months.

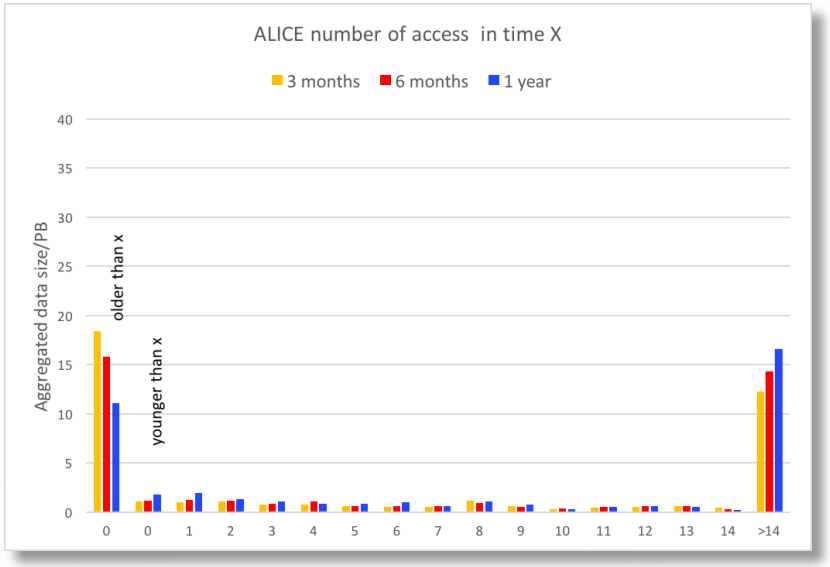
1st bin: data created before the period and not accessed in the period.

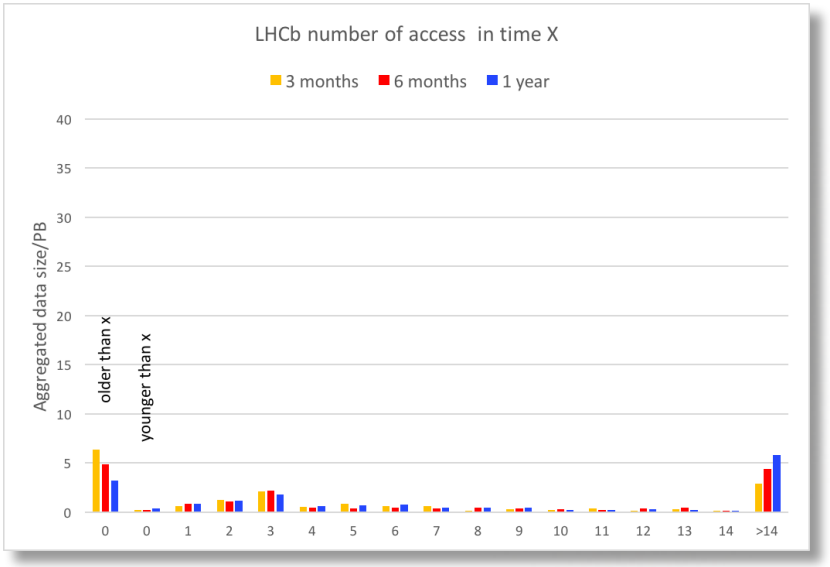
2nd bin: data created in that period and not accessed in the period.

Disk space is efficiently used thanks to the lifetime associated to each dataset.

CMS: sizeable amount of data in 2nd bin due to massive MC production for winter conferences.

Analysis of Access-Frequency Data ALICE & LHCb





LHCb routinely monitors datasets usage pattern and purge the unused one, in 2016, 3.5 PB of disk space were recovered.

ALICE will work to improve disk space usage.

ALICE Resource Use

Resource	Site(s)	2016 request	2016 pledged	2016 used	2016 used/pledged	2016 CPU efficiency
CPU (kHS06)	T0 + CAF	215	216	218	101 %	73%
	T1	157	177	253	143 %	80%
	T2	237	231	255	110 %	82%
Disk (PB)	T0	16.80	16.80	13.3	79 %	
	T1	21.0	18.95	17.4	92 %	
	T2	26.1	18.51	14.0	76 %	
Tape (PB)	T0	21.6	21.6	25.5	118 %	
	T1	15.6	17.77	18.5	104 %	

- Fulfilled data taking program; 40% of 2016 p-p to be reconstructed.
- Good CPU usage, dominated by the simulation. HLT is 5% of total CPU.
- Disk storage equally distributed among the tiers, occupied by simulation and reconstruction data, analysis data ~15%. Disk space: used 70% of the requested 90% of the installed.
- From popularity plot, 11 PB of data are not accessed in the last year, we invite the collaboration to investigate and take actions.
- Tape space filled mainly by RAW data, used more than pledged.

ATLAS Resource Use

Resource	Site	Pledged	Used	Used/Pledged	Average CPU efficiency
CPU (kHS06)	T0+CAF	235	241	103%	87%
	T1	538	642	129%	82%
	T2	610	1235	202%	80%
	HLT	22	56	255%	76%
Disk (PB)	T0+CAF	17	14	82%	
	T1	53	48	91%	
	T2	69	72	104%	
Tape (PB)	T0	42	41	98%	
	T1	119	67	56%	

- ATLAS used ~ twice the pledged CPU, mainly for MC data simulation and production.
- Tape space at T1 is filled at 56% due to the delay in MC reconstruction.
- High priority was in reducing the number of DAODs copies and Run-1 data and MC removal from disk to save space.

CMS Resource Use

Resource	Site	Pledged	Used	Used/Pledged	Average CPU efficiency
CPU (kHS06)	T0+CAF	306	260	85%	81%
	T1	348	358	103%	68%
	T2	677	872	129%	65%
	HLT	-	51	-	80%
Disk (PB)	T0+CAF	16	15	90%	
	T1	30	30	100%	
	T2	41	37	90%	
Tape (PB)	T0+CAF	44	40	91%	
	T1	88	73	83%	

- CPU efficiency at T1 and T2 is low due to scheduling overheads, switching from single-core to multi-core. Further analysis is underway.
- Due to the missing pledged tape space at T1, a massive tape deletions was performed to free up 30 PB in addition to an aggressive disk clean-up. All this was very costly in human resources.
- ECoM-17 Evolution of the Computing Model" committee was form to study and optimize requests before Run-3

LHCb Resource Use

Resource	Site(s)	2016 pledge	2016 used	Used/pledged	Average CPU efficiency
CPU (kHS06)	T0	51	31.4	62%	91%
	T1	165	201.4	122%	95%
	T2	88.6	115.6	130%	96%
Disk (PB)	T0	7.6	4.59	60%	
	T1	15.9	13.0	82%	
	T2	2.7	2.8	103%	
Tape (PB)	T0	20.6	16.6	81%	
	T1	35.0	22.8	65%	

- As in the past very high CPU efficiency, HLT is included in T1 resources.
- Stripping output 120MB/s instead 100MB/s and the TURBO stream 50kB/event instead 10kB/event → parking data and reduction of stored copies were implemented to mitigate the requests.
- Aggressive data access monitoring helped to clean up disk space.

Data taking assumptions

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	0.7	35

Peak luminosity 1.7×10^{34} cm^{-2}s^{-1}. It may be pessimistic…

Resource requests for 2018

- During the October RRB FA made clear that 2018 could not be a(nother) special year.
- Experiments were asked by LHCC/WLCG to develop a strategy to mitigate resource requests without jeopardize physics.
- Final LHCC document reports: "*The LHCC notes that the margins to reduce the resource usage in the short term without impact on physics have been exhausted*"
- CRSG decided to evaluate 2018 requests respect to 2017 pledges

- CMS has a deficit of tape at T1
- CMS and ALICE have deficit of disk at T1 and T2
- CPU deficit is not crucial

	ALICE	ATLAS	CMS	LHCb
CERN CPU	0%	0%	0%	0%
CERN disk	0%	0%	0%	0%
CERN tape	0%	0%	0%	0%
T1 CPU	-8%	-12%	-14%	-4%
T1 disk	-14%	1%	-21%	-6%
T1 tape	-1%	-7%	-24%	-3%
T2 CPU	-24%	-13%	-7%	27%
T2 disk	-28%	-7%	-22%	-30%

Alice Scrutiny

Resource	Site	2017 pledge	2018 ALICE	Growth	2018 CRSG	Growth
CPU (kHS06)	T0+CAF	292	350	20%	350	20%
	T1	235	306	30%	306	30%
	T2	278	438	58%	-	-
Disk (PB)	T0+CAF	22	27	23%	27	23%
	T1	21.8	32	47%	-	-
	T2	22.5	41	82%	-	-
Tape (PB)	T0+CAF	36.9	55	49%	55	49%
	T1	30.6	41	34%	41	34%

- ALICE has deficit in disk at T1 and T2, this was happening since few years generating the 82% (it is 31% respect to the requests)
- ALICE requests are beyond flat budget and higher than the average increase 2013-2017. CRSG had several discussion with the experiment, however the ALICE computing coordinator made it very clear that the increased requests are needed to achieve the physics program endorsed by LHCC.

Alice Scrutiny cont'd

Resource	Site	2017 pledge	2018 ALICE	Growth	2018 CRSG	Growth
CPU (kHS06)	T0+CAF	292	350	20%	350	20%
	T1	235	306	30%	306	30%
	T2	278	438	58%	-	-
Disk (PB)	T0+CAF	22	27	23%	27	23%
	T1	21.8	32	47%	-	-
	T2	22.5	41	82%	-	-
Tape (PB)	T0+CAF	36.9	55	49%	55	49%
	T1	30.6	41	34%	41	34%

- Tape space is crucial for raw data storage and CRSG supports the request.
- CPU at T0 and T1 as well as disk space at T0 are supported, needed to process data.
- For CPU at T2, disk space at T1 and T2 CRSG is not in a position to accept these requests. We asks experiment, LHCC and FA to take the actions necessary to reconcile the deficit, which imply to increase the pledges and reduce the resource needs. We propose to scrutinize these requests in October.

ATLAS Scrutiny

Resource	Site	2017 Pledge	2018 ATLAS	Growth	2018 CRSG	Growth
CPU (kHS06)	T0+CAF	404	411	2%	411	2%
	T1	808	949	17%	949	17%
	T2	982	1160	18%	1160	18%
Disk (PB)	T0+CAF	25	26	4%	26	4%
	T1	69	72	4%	72	4%
	T2	78	88	13%	88	13%
Tape (PB)	T0+CAF	77	94	22%	94	22%
	T1	174	195	12%	195	12%

- ATLAS requests are within the expected flat budget increase and below the average 2013-2017 increase.
- CRSG recommends the requests.
- Beyond pledge resources is about 30% of the pledges, ATLAS expect to continue to receive a sizeable amount of over pledge CPU, which remain a risk for the experiment.

CMS Scrutiny

Resource	Site	2017 Pledge	2018 CMS	Growth	2018 CRSG	Growth
CPU (kHS06)	T0+CAF	423	423	0%	423	0%
	T1	515	600	17%	600	17%
	T2	791	900	14%	900	14%
Disk (PB)	T0+CAF	25	26	6%	26	6%
	T1	45	60	34%	60	34%
	T2	53	70	32%	70	32%
Tape (PB)	T0+CAF	71	80	13%	97	36%
	T1	133	205	54%	188	41%

- CMS has deficit in pledges for tape and disk space.
- Disk space requests for 2018 at T1 and T2 are above the general funding profile, we recommend them and since the CPU requests are below the flat budget expectations we ask the FA to help on disk.
- CMS is suffering from a deficit of tape space at T1 since few years. To mitigate the request at T1, that unlikely will be satisfied this year, we ask CERN to provide more tape space. Requests for CERN are CPU = 0 and only 6% increase in disk space.

LHCb Scrutiny

Resource	Sites(s)	2017 Pledge	2018 LHCb	Growth	2018 CRSG
CPU(kHS06)	T0	67	81	21%	81
	T1	199	253	27%	253
	T2	147	141	-4%	141
	HLT + Yandex	20	20	-	20
Disk (PB)	T0	10.9	12	12%	12
	T1	20.8	24.5	20%	24.5
	T2	3.3	5.8	-	5.8
Tape (PB)	T0	25.2	36.4	44%	36.4
	T1	41.9	61.5	47%	61.5

- LHCb requests for CPU and disk space are slightly above the flat budget expectation but within/below the average 2013-2017 increase.
- Tape requests are higher in order to “park” data. This was chosen to mitigate requests of CPU and disk space. The increase in tape space in 2019 will be below 10%.
- CRGS recommends the requests.

Comments and Recommendations

- C-RSG commends the WLCG and the experiments for the work done to mitigate the resource requests without jeopardize physics.
- C-RSG encourages all experiments to pursue use of non-WLCG CPU resources. To help monitor this, we recommend that all experiments quantify more fully the non-WLCG resources in their future reports.
- CPU efficiency and reduction of data stored on disk are almost at the limit. It is not clear that there is substantially more efficiency that can be gain without extensive reworking of the computing model.
- C-RSG strongly support software engineering development and recommended that sufficient effort is funded to support this activity in the collaborations.
- The assumption of a flat budget is not consistent with the historical pledge resources and we recommend a reevaluation of the assumptions of what a flat budget entails.