Report of the Computing Resources Scrutiny Group

CERN–RRB–2014–049

Jonathan Flynn

for the CRSG

29 April 2014

Contents

Report on use of resources in 2013, requests for 2015

Contents

Assumptions for LHC Run 2

RRB year	pp/Ms	HI/Ms	$\mathcal{L}/10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$	pp pileup
2015	3	0.7	1	25
2016	5	0.7	1.5	40
2017	7	0.7	1.5	40

- pp running at 13 TeV CM energy
- LHC live time grows through the run
- 25 ns bunch spacing (except for short initial period)
- pp pileup important for ATLAS and CMS
- Revised efficiency assumptions

CPU organised	CPU analysis	disk	tape
85%	70%	100%	85%

- We take trigger rates after LHCC scrutiny

Contents

Overall assessment

- WLCG resources intensively used
- Computing models evolving to optimise use of resources
  - tier hierarchy dissolving (helped by good networking)
  - fewer reprocessings
  - reduced number of copies of data
  - fewer data types saved
- Offline use of HLT farms demonstrated by ATLAS, CMS and LHCb. All experiments plan to use them in Run 2
- Benefit from resources outside WLCG

Overall assessment

Software development

- faster algorithms, faster libraries → reduced CPU use per event
- reduced memory consumption
- adaptation to changing architectures

CRSG strongly supports these efforts which have lasting benefits for future resource use

CRSG asked all experiments for data-popularity information for disk use

- look beyond occupancy
- minimise storage of data which is never or seldom read
- pursue further in future scrutinies

Meeting flat-budget growth?

- Requests up to 2017 in the computing model update
- Assume performance increases at constant currency

  20% CPU
  15% disk and tape
  per annum

- Assume software improvements, problems like increased pileup solved. Much progress during LS1; more to be done.
- Depends on
  - when you start
  - what you start from (requests, pledges, installed)
  - performance of the LHC and the experiments

Meeting flat budgets?

Overview of sum of all requests from 2014 start

- CPU and disk at T0 jump above FB in 2015 but subsequent growth within FB
- Other resources, apart from T2 CPU, grow above FB earlier or later in Run 2
- For 2013 start: 2015 jump at T0, growth of tape

Full exploitation of physics potential of LHC and experiments from 2015 will require significant increase in resources.

- Meeting FB growth with FB spending depends on past funding, hardware replacement cycles, other costs (eg people, electricity)
- Might need increased budget in short term even to meet fixed-cost hardware performance increase

Contents

Overall used/pledged Jan–Dec 2013

		average	end of year
CPU	CERN	66%	—
	T1	114%	—
	T2	158%	—
Disk	CERN	116%	119%
	T1	140%	143%
	T2	—	—
Tape	CERN	106%	109%
	T1	82%	87%

- Similar to 2012: more use of CPU at CERN; use of pledged tape continues to rise
- Significant beyond-pledge use

From WLCG accounting (T2 disk info not available); averages are time-integrated; end of year uses capacity.

2013 fulfilment of pledges: installed/pledged

CPU	
CERN	100%
T1	107%
T2	134%†

Disk	
CERN	100%
T1	115%
T2	—

Tape	
CERN	94%
T1	99%

- Situation at end of Dec 2013, from WLCG accounting
- † T2 CPU percentage is delivered/pledged for Dec 2013 from WLCG T2 reports

Resource use at CERN plus T1s

End of 2013

	CPU	Disk	Tape	% CPU at CERN
ALICE	16%	12%	8%	44%
ATLAS	53%	47%	41%	13%
CMS	19%	29%	41%	40%
LHCb	11%	11%	10%	21%

- First three columns show division between experiments
- Last column is percentage of total CPU consumption by each experiment which was at CERN (column need not sum to 100%)
- Pattern similar to 2012

CPU is time-integrated over the year; storage is capacity in use at year-end. Data from EGI accounting.

T2 CPU usage

Distribution of time-integrated T2 CPU consumption by experiment

	2013	2012	2011
ALICE	10%	7%	9%
ATLAS	56%	53%	53%
CMS	27%	35%	30%
LHCb	7%	5%	7%

Data from EGI accounting. Calendar years 2013, 2012 and 2011.

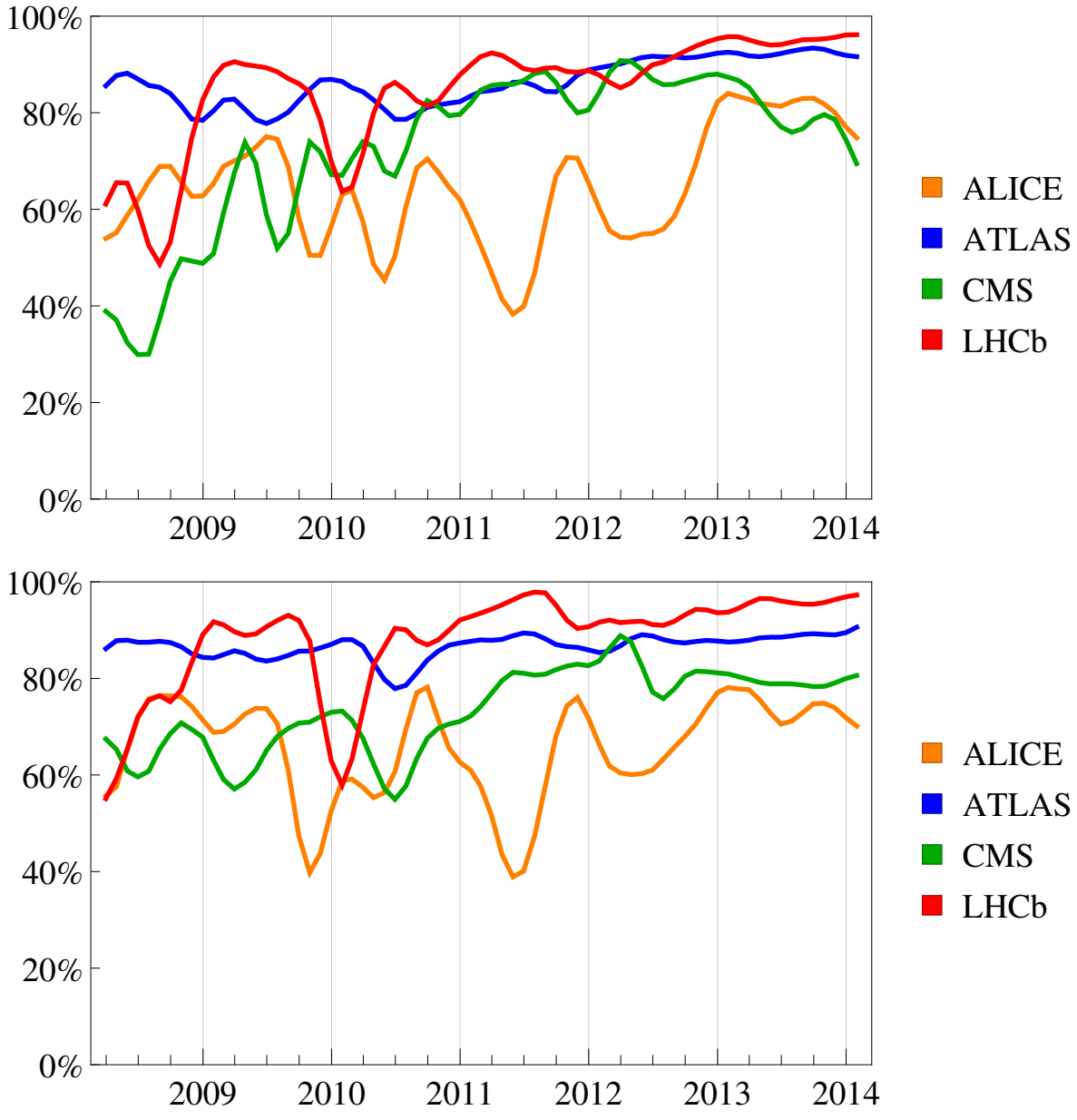
CPU efficiency

CPU time/wall time for calendar years 2013, 2012, 2011

	CERN plus T1			T2		
	2013	2012	2011	2013	2012	2011
ALICE	82%	64%	57%	76%	64%	60%
ATLAS	93%	92%	87%	89%	88%	88%
CMS	81%	88%	84%	80%	83%	82%
LHCb	96%	92%	90%	96%	95%	97%

Data from EGI accounting portal

CPU history: efficiency



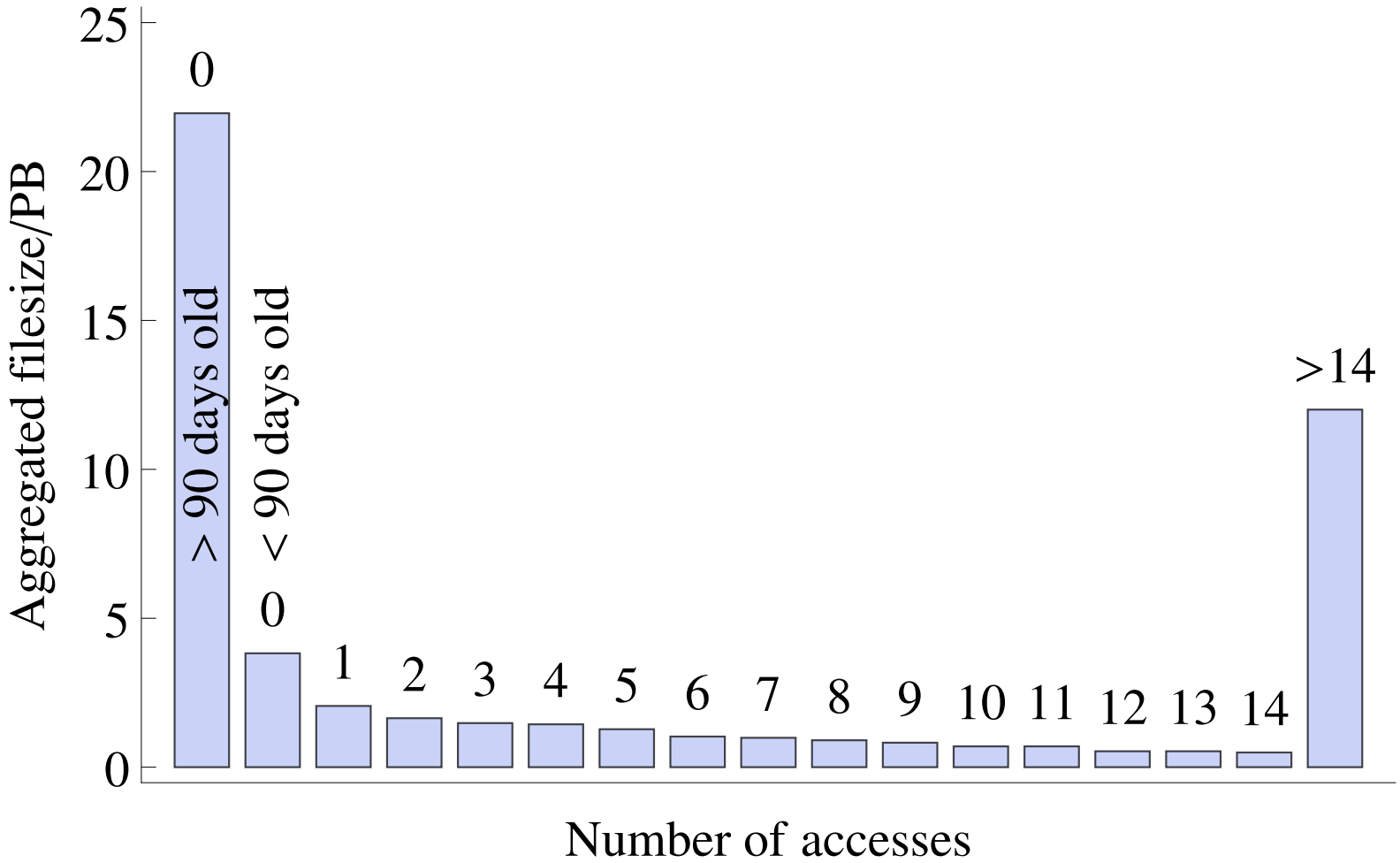
Top: CERN plus T1
Bottom: T2

Normalised CPU time (HS06·hrs) over normalised elapsed time.

Data from EGI accounting

Gaussian-smoothed monthly values

Data popularity



Volume of data versus number of accesses in ATLAS DATADISK at T1s and T2s for 90 days to 14 March 2014

Data popularity

- CRSG asked all experiments for data-popularity information for disk use
- Minimise storage of data which is never or seldom read
- We show ATLAS plot because we found it most useful
- Pursue this with all experiments in next and future scrutinies
- We hope that revealing and monitoring this information will lead to more efficient use of disk space

ALICE 2013

		Request	Pledge	Used	$\frac{\text{Used}}{\text{Pledge}}$
CPU (kHS06)	T0	126	90	94	104%
	T1	101	101	122	121%
	T2	188	184	144	78%
Disk (PB)	T0	8.3	8.1	7.1	88%
	T1	10.1	7.8	9.9	127%
	T2	12.8	12.8	7.7	60%
Tape (PB)	T0	12.0	22.8	9.2	40%
	T1	6.0	14.2	5.6	39%

- Using new T1 at KISTI
- Above-pledge CPU use at T1 compensates lower pledge at T0. Not all pledges installed (CPU, T2 disk)
- Reduced tape use (model change: no ESD/AOD on tape)
- Storage underused at sites with poor connectivity

ATLAS 2013

		Pledge	Used	$\frac{\text{Used}}{\text{Pledge}}$	Avge CPU efficiency
CPU (kHS06)	T0	111	111	100%	94%
	T1	333	503	151%	96%
	T2	404	729	180%	92%
Disk (PB)	T0	10	9	90%	
	T1	36	38	105%	
	T2	49	47	96%	
Tape (PB)	T0	27	29	107%	
	T1	41	38	93%	

- Successful use of HLT and of resources beyond pledges
- Simulation framework improved (speedup, fast/full sim mixing)
- T1 and T2 disk close to saturation

CMS 2013

		Pledge	Used	$\frac{\text{Used}}{\text{Pledge}}$	Avge CPU efficiency
CPU (kHS06)	T0	121	87	72%	80–85%
	T1	150	124	83%	85%
	T2	420	407	97%	80%
Disk (PB)	T0	7	6	85%	
	T1	24	23	97%	
	T2	30	31	103%	
Tape (PB)	T0	26	28	108%	
	T1	48	45	93%	

- CPU use at T0 ramped up in second half of year
- Use of HLT farm with upgraded bandwidth to T0
- Data popularity agent introduced to monitor disk use

LHCb 2013

		Pledge	Used	$\frac{\text{Used}}{\text{Pledge}}$
CPU (kHS06)	T0	34	16.3	48%
	T1	92	74.7	81%
	T2	52	90.8	175%
Disk (PB)	T0	4.0	3.0	75%
	T1	7.0	6.8	97%
	T2		0.2	
Tape (PB)	T0	6.5	5.9	91%
	T1	9.5	8.8	93%

- Planned computing tasks completed; T0,T1 underuse compensated at T2
- Successful restripping with large-scale recall from tape
- Tier 2 disk commissioning at selected sites
- HLT used extensively for simulation

Contents

ALICE

		2014 ALICE	2014 CRSG	2015 ALICE	2015 CRSG
CPU (kHS06)	T0	135	135	175	175
	T1	110	110	120	120
	T2	190	190	200	200
Disk (PB)	T0	8.3	8.3	14.5	14.5
	T1	10.1	10.1	17.8	17.8
	T2	12.8	12.8	22.7	22.7
Tape (PB)	T0	12.0	12.0	16.2	16.2
	T1	6.0	6.0	10.2	10.2

- CPU and storage for Run 2 increased by 25% (beam energy and pileup)
- PbPb and pPb events include TPC data; raises reco and sim times

ALICE

- Major demands come with heavy-ion running towards end of year.
  - T0 CPU needed for heavy-ion reconstruction before following year’s pp run.
- Sum of T1 and T2 resources more important than precise division between them.
- Some significant jumps in requests for 2015 and on to 2017.
- HLT farm being upgraded; expected to be operational at end of 2014. Planned use for offline tasks in Run 2.

ATLAS

		2014 ATLAS	2014 CRSG	2015 ATLAS	2015 CRSG
CPU (kHS06)	T0	111	111	205	205
	T1	355	355	462	450
	T2	390	390	530	520
Disk (PB)	T0	11	11	14	14
	T1	33	33	39	36
	T2	49	49	55	53
Tape (PB)	T0	27	27	33	33
	T1	44	44	65	65

- Request essentially the same as last October
- Disk: reduced pre-placement and more aggressive deletion of unused data
- Multi-core capable software, new analysis format, removal of a data-copy step

ATLAS

- CRSG strongly supports software development. *Benefits needed* to constrain future resource needs
- CRSG welcomes more aggressive policy for deleting unused data, but maintains pressure to make more effective use of disk with small reduction in T1 and T2 disk
- Acknowledge successful use of HLT farm; but we think its use should be included in requests (hence CPU reduction)

CMS

		2014 CMS	2014 CRSG	2015 CMS	2015 CRSG
CPU (kHS06)	T0	121	121	271	271
	T1	175	175	300	300
	T2	390	390	500	500
Disk (PB)	T0	7	7	3+12	15
	T1	26	26	27	26
	T2	27	27	31	29
Tape (PB)	T0	26	26	31+4	35
	T1	55	55	74	74

- 2015 requests unchanged since last October
- CMS takes account of use of HLT in requests

CMS

- Efforts to constrain CPU requirements
  - Software efficiency improvements. CRSG strongly supports this.
  - T0 setup to be more like T1 to allow prompt reconstruction at T1s from 2015
  - Fewer reprocessing passes
  - Reduction in ratio of simulated to real events — may hurt physics output
- CRSG acknowledges these efforts
- As for ATLAS, still push for aggressive cleanup of unused data to make more effective use of disk → small reduction in T1 and T2 disk

LHCb

		2014 pledge	2015 LHCb	2015 CRSG
CPU (kHS06)	T0	34	36	36
	T1	110	118	118
	T2	62	66	66
	HLT + Yandex		10+10	
Disk (PB)	T0	4.0	5.5	5.5
	T1	11.7	11.7	11.7
	T2	1.1	1.9	1.9
Tape (PB)	T0	8.5	11.2	11.2
	T1	11.0	23.7	23.7

- More use of T2 for simulation and analysis; introduction of T2 disk
- Use of HLT and Yandex accounted for in request

LHCb

- Several changes for 2015
  - No further reprocessing of Run 1 data in 2015
  - Postpone *re*processing of raw data to LS2
  - Omitted reconstruction pass and reduced stripping
  - Reduced ratio of full DST to microDST to reduce storage
- Jump in tape for 2015, including significant space for data preservation
- Bigger jumps in CPU/disk/tape anticipated for 2016
- LHCb noted that common LHC running assumptions used here may be pessimistic

Contents

Comments and recommendations

1. Run 2 requests made with assumption of flat budget (not inflation-adjusted)

Comments and recommendations

1. Flat budget assumption built in
2. Data preservation. Distinguish ability to read/reanalyse old data from requirements for open/public access (both storage and human effort)

Comments and recommendations

1. Flat budget assumption built in
2. Data preservation: distinguish reuse of old data from open access
3. CRSG acknowledges use of HLT farms during LS1 and plans to use them during technical stops and shutdowns in Run 2. CRSG does not consider this use to be opportunistic.

Comments and recommendations

1. Flat budget assumption built in
2. Data preservation: distinguish reuse of old data from open access
3. Use of HLT farms
4. Improving software efficiency (ultimately physics per euro) is essential to constrain growth in requests. The resulting gains are already assumed. CRSG strongly supports this and recommends that sufficient effort is funded.

Comments and recommendations

1. Flat budget assumption built in
2. Data preservation: distinguish reuse of old data from open access
3. Use of HLT farms
4. Support software engineering
5. Effectiveness of disk use only partly captured by occupancy. CRSG welcomes experiments' efforts to purge obsolete or unused data and thanks them for supplying data popularity information.

Comments and recommendations

1. Flat budget assumption built in
2. Data preservation: distinguish reuse of old data from open access
3. Use of HLT farms
4. Support software engineering
5. Disk efficiency
6. Good networking has been exploited to reduce disk use (fewer pre-placed copies of data) and move processing between tiers. Danger that poorly-networked sites will be underused and possible cost implications of providing network capacity.

Comments and recommendations

1. Flat budget assumption built in
2. Data preservation: distinguish reuse of old data from open access
3. Use of HLT farms
4. Support software engineering
5. Disk efficiency
6. Importance of networking
7. Scrutiny schedule. First scrutiny of 2016 requests in October, revisiting 2015 only if necessary. We do not intend to report usage to the RRB in October.

Comments and recommendations

1. Flat budget assumption built in
2. Data preservation: distinguish reuse of old data from open access
3. Use of HLT farms
4. Support software engineering
5. Disk efficiency
6. Importance of networking
7. Scrutiny schedule

Contents

CRSG membership

T Cass (CERN)
J Flynn (UK, chairman)
M Gasthuber (Germany)
J Kleist (Nordic countries)
G Lamanna (France)
D Lucchesi (Italy)
J Marco (Spain)
H Meinhard (CERN/IT sci sec)
T Schalk (USA)
J Templon (Netherlands)
M Vetterli (Canada)

G Lamanna (France) and M Gasthuber (Germany) will stand down after this scrutiny round and will need replacing. We thank them for their contributions.

We thank the experiments for their dialogue with us and the CERN management for support.