Report of the Computing Resources Scrutiny Group

CRSG report to the CRRB

Jonathan Flynn

CRSG

16 April 2013

Contents

Report on use of resources in 2012, requests for 2014 and some discussion of 2015.

Contents

LHC running conditions

LHC live time in Ms for 2012 and anticipated for 2015

RRB year	pp	HI
2012	6.6	0.7
2013	—	—
2014	—	—
2015	5.2	0.7

- pp running at 8 TeV CM energy in 2012
  - maximum luminosity more than double 2011 maximum
  - run was extended
- pPb run in Jan–Feb 2013
- 25 ns bunch spacing demonstrated at end of run

LHC running conditions

Effective overall event rates

ALICE	pp 180 Hz, pPb 280 Hz
ATLAS	prompt 450 Hz, delayed 140 Hz
CMS	prompt 450 Hz, delayed 360 Hz
LHCb	4 kHz plus 1 kHz deferred

- Not necessarily the same as trigger rates (eg: ALICE has to wait until luminosity declines before recording)
- ATLAS and CMS delayed or parked data is recorded for reconstruction during LS1
- LHCb deferred events are buffered in HLT for processing in inter-fill periods

Contents

Overall assessment

- WLCG resources intensively used
- Experiments continue to evolve computing models and optimise use of resources
  - reduced CPU use per event
  - fewer reprocessings
  - reduced number of copies of data
  - fewer data types saved on tape
- Hierarchical distinction of which tasks run where becoming less rigid
- Use of HLT farms by ATLAS, CMS and LHCb for offline tasks during LS1 and plans to do so during times without beam from 2015
- Benefit from use of resources outside WLCG

Overall assessment

- Experiments to review computing models and analysis strategies over Summer 2013
- Experiments launching software development programmes
  - faster algorithms, faster libraries
  - code improvements
  - reduced memory consumption
  - adaptation to changing architectures

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

Overall assessment

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

- Looks achievable if funding can be maintained at current level
- Anticipated increase in resources has *already* been moderated by assuming software improvements will work, problems like increased pileup and out-of-time events solved, . . .
- Experiments anticipate being more certain about 2015 requests by October

Contents

Overall used/pledged Jan–Dec 2012

		average	end of year
CPU	CERN	61%	—
	T1	116%	—
	T2	171%	—
Disk	CERN	104%	111%
	T1	135%	141%
	T2	—	—
Tape	CERN	88%	101%
	T1	71%	86%

- From WLCG accounting (T2 disk info not available)
  - Average uses time-integrated CPU or storage
  - End of year uses capacity
  - Pledges here include efficiency factors
- Similar to 2011 but more use of pledged tape

2012 fulfilment of pledges: installed/pledged

CPU	
CERN	100%
T1	111%
T2	169%†

Disk	
CERN	85%
T1	113%
T2	—

Tape	
CERN	100%
T1	100%

- Situation at end of 2012, from WLCG accounting
- † T2 CPU percentage is delivered/pledged over 2012 from WLCG T2 reports

Resource use at CERN plus T1s

End of 2012

	CPU	Disk	Tape	% CPU at CERN
ALICE	12%	13%	10%	50%
ATLAS	55%	44%	39%	14%
CMS	24%	33%	42%	30%
LHCb	10%	10%	10%	22%

- First three columns are division of resource use between experiments
- CPU is time-integrated over the year; storage is capacity in use at year-end
- Last column is percentage of total CPU consumption by each experiment which was at CERN (column need not sum to 100%)
- Pattern similar in 2011

T2 CPU usage

Distribution of time-integrated CPU consumption by experiment

	2012	2011
ALICE	7%	9%
ATLAS	53%	53%
CMS	35%	30%
LHCb	5%	7%

CPU efficiency by experiment (CPU time/wall time)

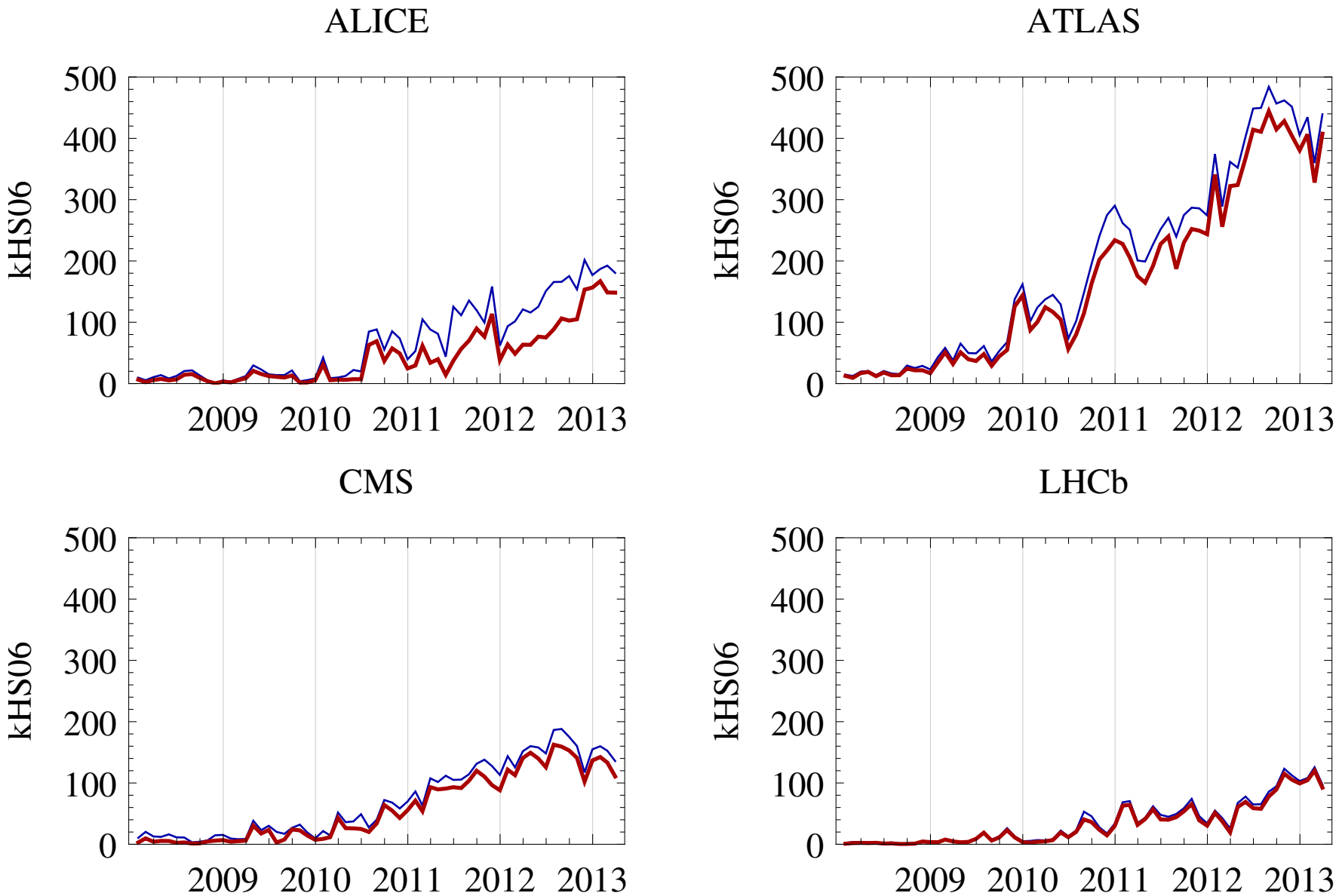
	2012	2011
ALICE	64%	60%
ATLAS	88%	88%
CMS	83%	82%
LHCb	95%	98%

Data from EGI accounting portal

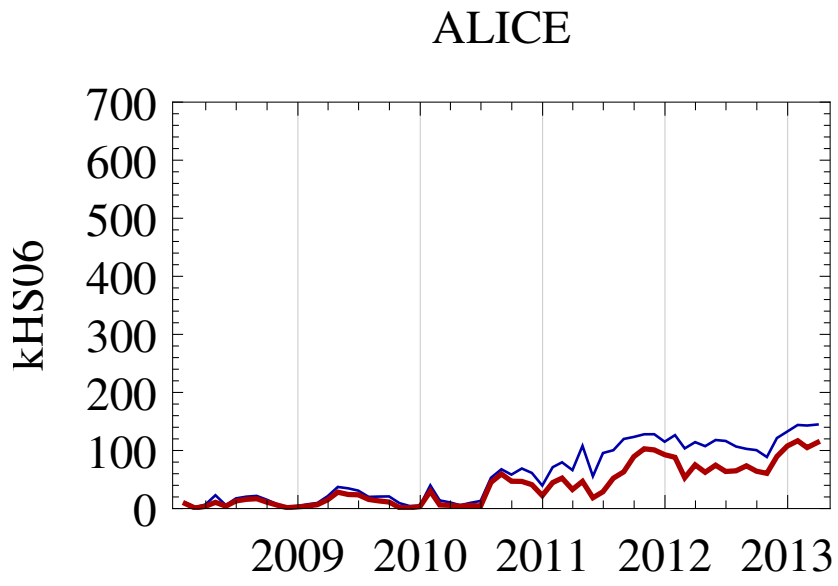
CPU history

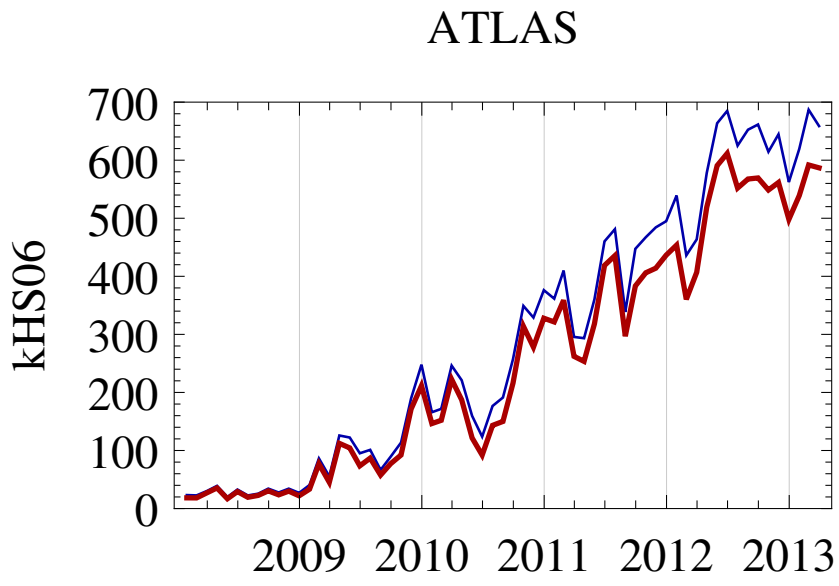
- Look at CPU history, Jan 2008 to March 2013
- Red: Monthly normalised CPU time (HS06·hrs), divided by month-length to get effective CPU power used
- Blue: Monthly normalised elapsed time (HS06·hrs), divided by month-length to get effective CPU power tied up
- Data from EGI accounting portal

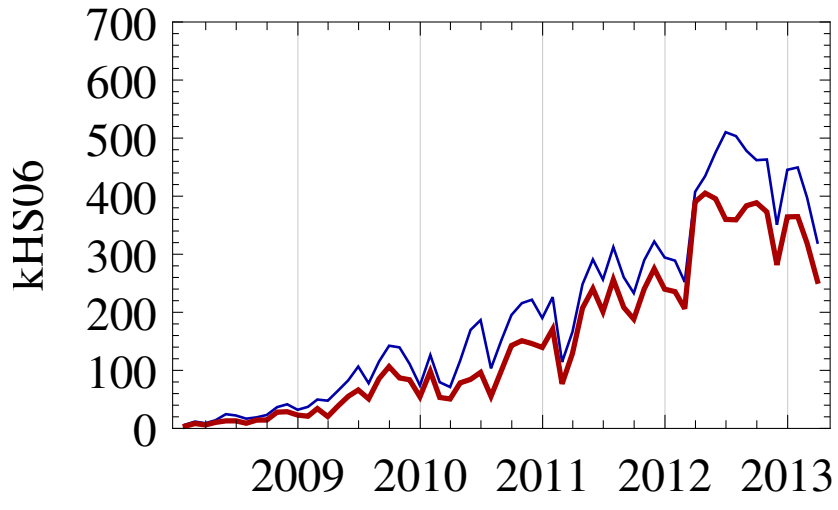
CPU history: CERN plus T1s

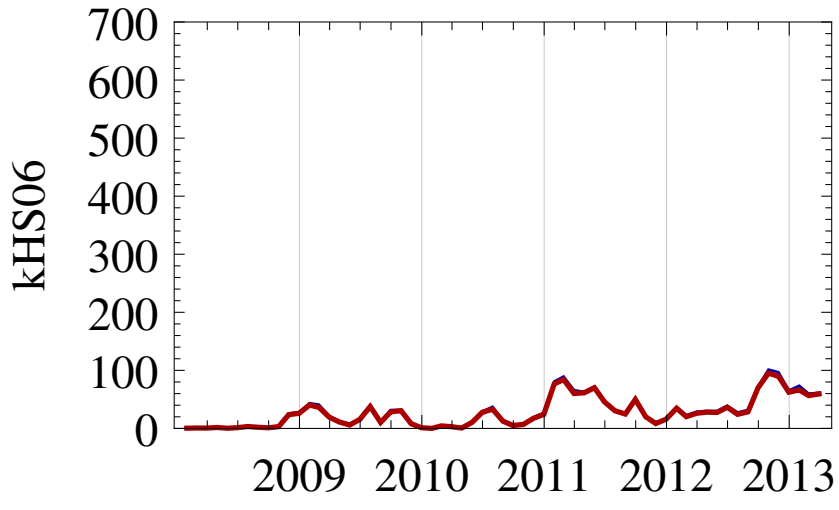


CPU history: T2s

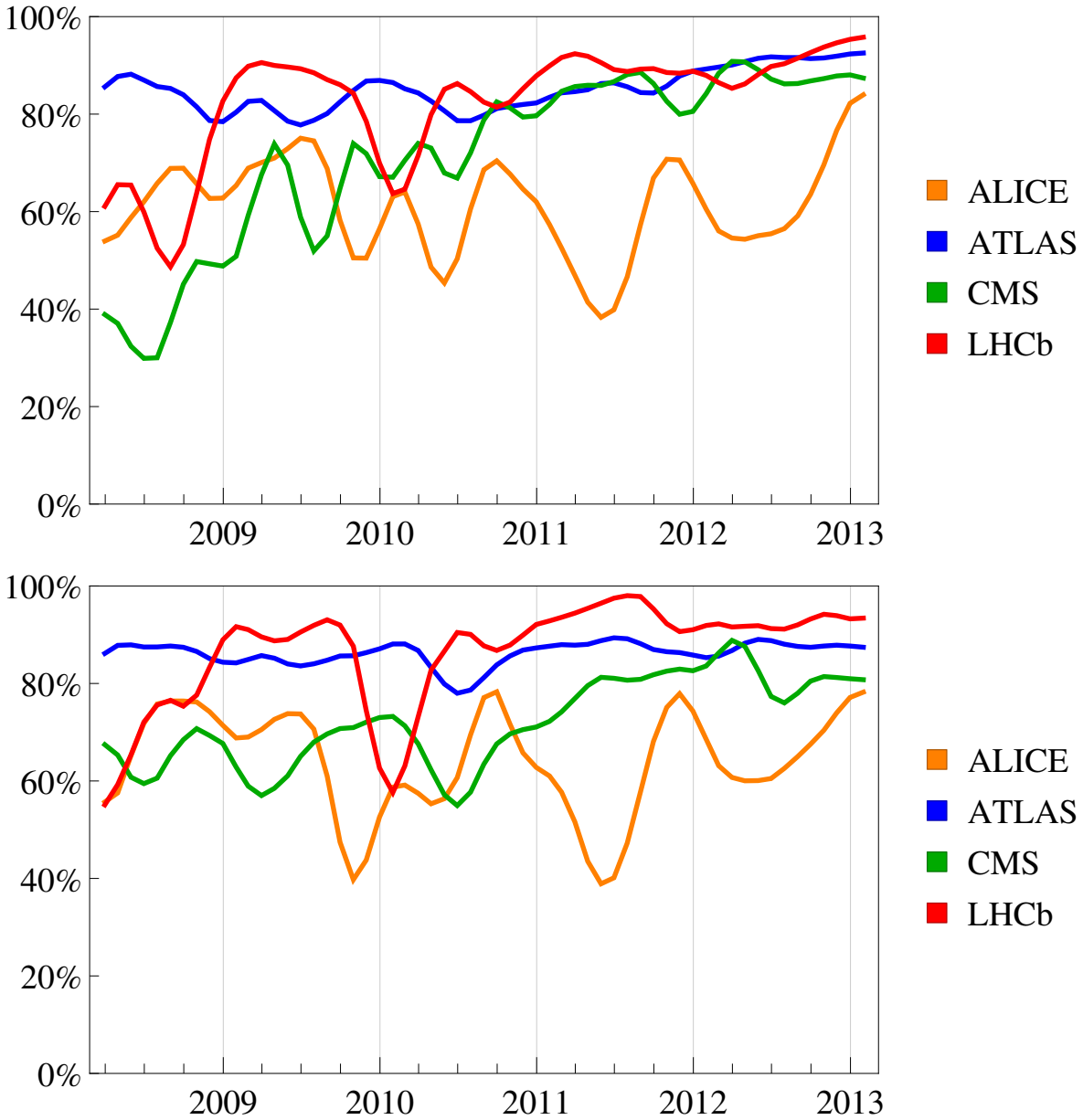








CPU history: efficiency



Top: CERN plus T1
Bottom: T2

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

Gaussian-smoothed monthly values.

Data from EGI accounting.

ALICE 2012

		2012 pledge	2012 used
CPU/kHS06	T0	90	78
	T1	113	83
	T2	178	189
Disk/PB	T0	8.1	8.1
	T1	8.1	9.0
	T2	10.9	10.1
Tape/PB	T0	20.0	10.2
	T1	12.0	4.6

- Resource use April 2012 – March 2013
- T0, T1 disk includes buffers for tape system
- Much reduced tape use (change in model: no simulation data on tape); carries forward to 2014

ATLAS 2012

		Pledge	Used	$\frac{\text{Used}}{\text{Pledge}}$	Avge CPU efficiency
CPU/kHS06	T0	111	111	100%	89.8%
	T1	285	420	147%	92.1%
	T2	332	634	191%	88.2%
Disk/PB	T0	9	10	111%	
	T1	30	47	157%	
	T2	45	52	116%	
Tape/PB	T0	18	29	161%	
	T1	38	31	82 %	

- Very successful use of resources beyond pledges
- Reduced CPU for reconstruction and simulation
- Improved distribution of data/jobs among T1s and T2s

CMS 2012

		Pledge	Used	$\frac{\text{Used}}{\text{Pledge}}$	Avge CPU efficiency
CPU/kHS06	T0	121	75	62%	90%
	T1	137	142	104%	88%
	T2	320	429	134%	83%
Disk/PB	T0	7	6	84%	
	T1	21	21.5	102%	
	T2	27	25	93%	
Tape/PB	T0	23	22	96%	
	T1	47	42	89%	

- Estimated resource usage proved generally accurate
- Longer pp run compensated by reducing parked data and by below-expected pileup

LHCb 2012

		2012 pledge	2012 used	
CPU/kHS06	T0	34	17.5	(26)
	T1	110	64.3	(115)
	T2	48	45.1	(94)
	other		13.0	(30)
Disk/PB	T0	3.5	2.6	
	T1	7.3	5.6	
Tape/PB	T0	6.4	5.2	
	T1	5.5	8.1	

- CPU used is year average (peak in parentheses); disk use does not include tape buffers, disk pledge does
- Twice as many events taken in 2012 cf 2011
- pp run extension
  - ~ 40% more raw data ⇒ pressure on storage
  - reduced copies on disk/tape; still pressure on tape

Contents

ALICE

		2013 needed	2014 needed	2015 needed
CPU/kHS06	T0	126	135	190
	T1	101	110	110
	T2	188	190	200
Disk/PB	T0	8.3	8.3	10.8
	T1	10.1	10.1	13.6
	T2	12.8	12.8	16.1
Tape/PB	T0	12.0	12.0	27.0
	T1	6.0	6.0	21.0

- Disk and tape flat in 2013 and 2014
- CPU increases below 10% in 2013 and 2014
- Larger jumps expected in 2015, notably for tape

ALICE

- Request to trade tape for disk in 2013 following change in model; we endorse this (might help meet disk increase for 2015)
- Encouraged by better CPU efficiency at end of 2012; strongly support software efforts to improve implementation of computing model
- ALICE concerned that T1 CPU/disk pledges for 2013 are below requests and below 2012 pledges
- ALICE has reduced requests at behest of CRSG to better match anticipated pledges; CRSG hopes that funding agencies can, in principle, fully fund the scrutinised requests

ATLAS

		2013 ATLAS	2013 CRSG	2014 ATLAS	2014 CRSG
CPU/kHS06	T0	111(111)	111	111(111)	111
	T1	316(333)	319	373(327)	355
	T2	360(396)	350	408(399)	390
Disk/PB	T0	10(10)	11	11(11)	11
	T1	35(36)	33	36(33)	33
	T2	51(49)	49	56(47)	49
Tape/PB	T0	25(27)	23	31(31)	27
	T1	42(41)	40	53(43)	44

- CPU: stable at T0 in 2014; ~ 20% increase for sum of T1 and T2 (use of HLT farm helps)
- Total disk (tape) request grows 7% (25%) 2013 to 2014

Pledges in parentheses

ATLAS

- Larger increases anticipated for 2015; largest is doubling CPU at T0 by end of 2015
- CRSG strongly supports activity to adjust computing model and improve software to reduce resource use. *Benefits needed* to constrain resource needs from 2015.
- Acknowledge use of HLT farm in 2013 and 2014. Encourage ATLAS to develop further the ability to use this during periods without beam from 2015.
- ATLAS supplied information on data replication and popularity; we are pushing them to limit disk use.

CMS

		2013 CMS	2013 CRSG	2014 CMS	2014 CRSG
CPU/kHS06	T0	121(121)	121	121(121)	121
	T1	165(150)	165	165(150)	175
	T2	350(400)	350	400(390)	390
Disk/PB	T0	7(7)	7	7(7)	7
	T1	26(23)	26	26(22)	26
	T2	26(29)	26	27(29)	27
Tape/PB	T0	26(26)	26	26(25)	26
	T1	50(48)	50	56(45)	55

- Requests stable in 2013 and 2014 apart from T2 CPU up 14% and T1 tape up 11%
- HLT for simulation mitigates CPU growth

Pledges in parentheses

CMS

- 2014 request allows for complete reprocessing pass on all real and simulated data accumulated to date, leaving a consistent legacy archive.
- Plan to use HLT farm to augment T1 capacity by ~ 40% in 2013/14. CRSG very encouraged by speed of reconfiguring farm and plans to use it in no-beam periods from 2015 (perhaps even between fills).
- Substantial increases anticipated for 2015 (doubling CPU at T0 and T1)
- CRSG strongly supports software development during LS1; *improvements are already assumed* in planning for next LHC run

LHCb

		2013 pledge	2014 request
CPU/kHS06	T0	34	40
	T1	92	110
	T2	47	47
	unpledged+HLT		10+30
Disk/PB	T0	4.0	6.4
	T1	7.0	14.0
Tape/PB	T0	6.5	6.6
	T1	9.5	11.1

- 2013 → 2014: CPU ↗ 14%, disk ↗ 43%, tape ↗ 15%
- Substantial use of HLT and other unpledged resources
- CPU request not anticipated to grow in 2015, but near-doubling in tape

LHCb

- 2014 requests solidly based on experience from 2012 processing and known amount of data; little margin for reduction remaining.
- More disk in 2014 to hold simulation data and use CPU. Both freed up to deal with new data in 2015.
- Use of unpledged resources acknowledged; LHCb even rely on getting them.
- LHCb can do all processing via the Grid so less sensitive to precise location of resources between T0 and T1 (provided disk-to-cpu ratio not too small).
- LHCb concerned that scrutinised T1 request cannot be met (~ 70% of collaboration associated with T1 centres); CRSG encourages RRB to address this.

Contents

Comments and recommendations

1. Strong expectation that exploitation of the LHC and the experiments from 2015 will require significantly increased computing resources. Looks achievable if stable funding can be maintained.

Comments and recommendations

1. 2015 reachable with stable funding
2. ALICE and LHCb’s scrutinised requests have not been fully met at T1. It even looks impossible in principle for 100% of the scrutinised levels to be met. We think this problem should be addressed urgently by the RRB.

Comments and recommendations

1. 2015 reachable with stable funding
2. Meeting scrutinised requests
3. CRSG is willing to work with WLCG management to review the request/review/pledge process in light of experience over the first period of LHC running.

Comments and recommendations

1. 2015 reachable with stable funding
2. Meeting scrutinised requests
3. Review r/r/p cycle
4. Improving software efficiency is essential to constrain 2015 requests. The resulting gains are already assumed in making those requests. CRSG strongly supports this and hopes that sufficient effort can be funded.

Comments and recommendations

1. 2015 reachable with stable funding
2. Meeting scrutinised requests
3. Review r/r/p cycle
4. Support software engineering
5. Efficiency factor of 0.7 has been used when calculating disk requirements. In practice disks are used more efficiently (ATLAS now prefers to make explicit calculation of disk space). CRSG welcomes a change to reflect more efficient use, but maintain ability to compare experiments on an equal footing.

Comments and recommendations

1. 2015 reachable with stable funding
2. Meeting scrutinised requests
3. Review r/r/p cycle
4. Support software engineering
5. Disk efficiency
6. The effectiveness of disk usage is only partly captured by disk occupancy figures. A metric which also takes account of frequency of access would be highly desirable.

Comments and recommendations

1. 2015 reachable with stable funding
2. Meeting scrutinised requests
3. Review r/r/p cycle
4. Support software engineering
5. Disk efficiency
6. New disk usage metric
7. CRSG welcomes the new documentary description of the experiments' computing models, expected before next RRB.

Comments and recommendations

1. 2015 reachable with stable funding
2. Meeting scrutinised requests
3. Review r/r/p cycle
4. Support software engineering
5. Disk efficiency
6. New disk usage metric
7. New computing model descriptions

Contents

CRSG membership

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

- Previous chairman D Espriu (Spain) stood down at end of 2012; replaced by Flynn as chairman
- M Vetterli (Canada) replaces W Trischuk
- J Marco (Spain) replaces Espriu
- D Groep (Netherlands) and B Vinter (Nordic Countries) to stand down and be replaced following this scrutiny