PIC
port d'informació
científica

Comparison of accounting and pledges for Tier1 sites

WLCG Management Board – 16/02/2016

J. Flix

EXCELENCIA SEVERO OCHOA
Institut de Física d'Altes Energies
IFAE
Ciemat
Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

PIC
port d'informació
científica

Acknowledgments...

- Many people from **experiments** and **sites** (T1s & T2s) have been involved in these discussions. Individuals listed here, with no order of preference:
  - Jeff Templon, John Gordon, Alessandra Forti, Brian Bockelmann Oxana Smirnova, Maarten Litmaath, Alessandro Di Girolamo, Philippe Charpentier, Helge Meinhard, Latchezar Betev, Andrew McNab, Manfred Alef, Michael Ernst, ShaoTing Cheng, Felix Lee, Concezio Bozzi, Sébastien Gadrat, Stephen Burke, Renaud Vernet, Catherine Biscarat…
- **Lof of feedback received!!!**
  - Only **CPU accounting** is going to be discussed in this 10' talk...

PIC
port d'informació
científica

Minutes from MB meeting 18/Nov/2014

Ian Bird reported that at the last RRB, following the recommendation of the CRSG, it was proposed to remove efficiency factors from WLCG resource accounting. These factors were originally introduced to ensure that the funding agencies were aware of what level of usage of CPU and storage was reasonable to expect. This is now generally well understood, and the actual usage is very high, and the continued use of the factors leads to some confusion. Therefore, there is general agreement that these efficiency factors have served their purpose and now should be removed from resource accounting reports. The experiments will continue to use the agreed efficiency factors in estimating their resource needs. Ian Bird invited MB members to let him know of any objections to this proposal, in the absence of which this will be considered as approved.

PIC
port d'informació
científica

Removing the efficiency factors

*Ian Bird reported that at the last RRB, following the recommendation of the CRSG, it was proposed to **remove efficiency factors from WLCG resource accounting**. These factors were originally introduced to ensure that the funding agencies were aware of what level of usage of CPU and storage was reasonable to expect. This is now generally well understood, and the actual usage is very high, and the continued use of the factors leads to some confusion. Therefore, there is general agreement that these efficiency factors have served their purpose and now should be removed from resource accounting reports. The experiments will continue to use the agreed efficiency factors in estimating their resource needs. Ian Bird invited MB members to let him know of any objections to this proposal, in the absence of which this will be considered as approved.*

PIC
port d'informació
científica

Removing the efficiency factors

- All of the efficiency factors were removed from the accounting reports since Apr. 2015. For CPU:
  - **Before Apr. 2015:** the CPU pledge for a site was scaled (down) by applying the CPU efficiency factor (0.85 for T1s, 0.7 for T2s); then, the consumed CPUtime in the year was compared to this scaled down CPU pledge
    - *Either we were scaling down the pledge, or scaling up the CPUtime usage, to account for the inefficiency of the jobs...*
  - **After Apr. 2015:** the CPU pledge for a site is not scaled down anymore; then, the consumed CPUtime in the year is compared to the CPU pledge...
    - But, as jobs are not 100% efficient, the jobs used WALLtime should indeed be compared to the CPU pledge **[see next slides]**

Efficiency inflation on experiment requests

Ian Bird reported that at the last RRB, following the recommendation of the CRSG, it was proposed to remove efficiency factors from WLCG resource accounting. These factors were originally introduced to ensure that the funding agencies were aware of what level of usage of CPU and storage was reasonable to expect. This is now generally well understood, and the actual usage is very high, and the continued use of the factors leads to some confusion. Therefore, there is general agreement that these efficiency factors have served their purpose and now should be removed from resource accounting reports. ***The experiments will continue to use the agreed efficiency factors in estimating their resource needs****. Ian Bird invited MB members to let him know of any objections to this proposal, in the absence of which this will be considered as approved.*

PIC
port d'informació
científica

Efficiency inflation on experiment requests

- Contacted **ALICE**, **ATLAS**, **CMS** and **LHCb**:
  - All the experiments **confirm** they are indeed including the efficiency factors in their internal calculations
  - The pledges are expressed in **"walltime-corrected" HS06** (no eff. factors applied)
  - They report to CRSG the used walltime at the sites wrt pledges
- Contacted **many Tier-1s**:
  - They **confirm** they are providing CPU pledges, understood as **"walltime-corrected" HS06** (no eff. factors applied)
- The experiments expect to execute tasks at the sites at these pledge levels, or even beyond (if there are resources available and not used by other VOs)

WLCG monthly accounting reports

- Monthly WLCG Office accounting reports still compare CPUtime to the pledges, for both T1s and T2s. This might be a pure technicality that was left forgotten, but it adds to the confusion and **needs to be fixed**. It also has strong political impact.
  - Universal agreement from all of the parties involved

WLCG Accounting

January to December 2015

Centre: CA-TRIUMF

MoU pledges	apr-14 mar-15	apr-15 mar-16
CPU (HS06-years)	38,500	46,200
Disk (Tbytes)	3,500	3,900
Tape (Tbytes)	5,300	6,500

Standard efficiency factors	
Disk utilisation	100%
Mass store utilisation (tape)	100%
Scheduled cpu used	100%
Scheduled cpu (until Mar 2015	85%

CPU used - HEPSPEC06-days	Jan 2015	Feb 2015	Mar 2015	Apr 2015	May 2015	Jun 2015	Jul 2015	Aug 2015	Sep 2015	Oct 2015	Nov 2015	Dec 2015	Total ****	% MoU
Grid ALICE cpu	0	0	0	0	0	0	0	0	0	0	0	0	0	
Grid ALICE wall	0	0	0	0	0	0	0	0	0	0	0	0	0	
Grid ATLAS cpu	1,055,143	1,110,895	1,656,668	1,875,509	1,874,573	1,972,247	2,031,620	1,863,888	1,914,261	1,935,006	1,890,587	1,599,592	20,780,009	
Grid ATLAS wall	1,351,108	1,257,928	1,876,848	2,106,454	2,149,360	2,183,583	2,241,889	2,192,256	2,025,456	2,134,267	2,104,582	2,069,125	23,712,856	
Grid CMS cpu	0	0	0	0	0	0	0	0	0	0	0	0	0	
Grid CMS wall	0	0	0	0	0	0	0	0	0	0	0	0	0	
Grid LHCb cpu	0	0	0	0	0	0	0	0	0	0	0	0	0	
Grid LHCb wall	0	0	0	0	0	0	0	0	0	0	0	0	0	
Grid TOTAL cpu	1,055,143	1,110,895	1,656,668	1,875,509	1,874,573	1,972,247	2,031,620	1,863,888	1,914,261	1,935,006	1,890,587	1,599,592	20,780,009	134%
Grid TOTAL wall	1,351,108	1,257,928	1,876,848	2,106,454	2,149,360	2,183,583	2,241,889	2,192,256	2,025,456	2,134,267	2,104,582	2,069,125	23,712,856	
Non-Grid ALICE cpu	0	0	0	0	0	0	0	0	0	0	0	0	0	
Non-Grid ALICE wall	0	0	0	0	0	0	0	0	0	0	0	0	0	
Non-Grid ATLAS cpu	0	0	0	0	0	0	0	0	0	0	0	0	0	
Non-Grid ATLAS wall	0	0	0	0	0	0	0	0	0	0	0	0	0	
Non-Grid CMS cpu	0	0	0	0	0	0	0	0	0	0	0	0	0	
Non-Grid CMS wall	0	0	0	0	0	0	0	0	0	0	0	0	0	
Non-Grid LHCb cpu	0	0	0	0	0	0	0	0	0	0	0	0	0	
Non-Grid LHCb wall	0	0	0	0	0	0	0	0	0	0	0	0	0	
Non-Grid TOTAL cpu	0	0	0	0	0	0	0	0	0	0	0	0	0	0%

PIC
port d'informació
científica

WLCG monthly accounting reports

- Monthly WLCG Office accounting reports still compare CPUtime to the pledges, for both T1s and T2s. This might be a pure technicality that was left forgotten, but it adds to the confusion and **needs to be fixed**. It also has strong political impact.
  - Universal agreement from all of the parties involved

WLCG
Worldwide LHC Computing Grid

Tier-2 Accounting Summary — **December 2015**

Efficiency factor for Tier-2 sites - utilisation 100% of pledge as agreed at CRRB Oct 2014 / MB Nov 2014

CPU usage in month (HEPSPEC06-Hrs)

This is cputime, not walltime

Federation - Accounting Name	April 15 - Mar 16 at 2015 CPU Pledge (HS06)	pledge inc. efficiency (HS06-Hrs)	Site(s)	ALICE	ATLAS	CMS	LHCb	Total	used as % of pledge
Australia, University of Melbourne			Australia-ATLAS		5'344'412			5'344'412	
AU-ATLAS	**10'700**	**7'960'800**			**5'344'412**			**5'344'412**	**67%**
Austria, Austrian Tier-2 Federation			HEPHY-UIBK		175'884			175'884	
			Hephy-Vienna			4'376'964		4'376'964	
AT-HEPHY-VIENNA-UIBK	**6'857**	**5'101'608**			**175'884**	**4'376'964**		**4'552'848**	**89%**
Belgium, Belgian Tier-2 Federation			BEgrid-ULB-VUB			3'020'996		3'020'996	
			BelGrid-UCL			3'906'108		3'906'108	
BE-TIER2	**23'100**	**17'186'400**				**6'927'104**		**6'927'104**	**40%**
Brazil, SPRACE, São Paulo			SPRACE			4'722'964		4'722'964	
BR-SP-SPRACE	**15'000**	**11'160'000**				**4'722'964**		**4'722'964**	**42%**
Canada-East Federation			CA-SCINET-T2		7'353'124			7'353'124	
			CA-McGill-CLUMEQ-T2		2'665'528			2'665'528	
CA-EAST-T2	**13'250**	**9'858'000**			**10'018'652**			**10'018'652**	**102%**
Canada-West Federation			CA-VICTORIA-WESTGRID-T2		7'757'616			7'757'616	
			SFU-LCG2		3'015'280			3'015'280	
CA-WEST-T2	**13'250**	**9'858'000**			**10'772'896**			**10'772'896**	**109%**
China, IHEP, Beijing			BEIJING-LCG2		5'646'856	3'066'452		8'713'308	

PIC
port d'informació
científica

WLCG monthly accounting reports

- Monthly WLCG Office accounting reports still compare CPUtime to the pledges, for both T1s and T2s. This might be a pure technicality that was left forgotten, but it adds to the confusion and **needs to be fixed**. It also has strong political impact.
  - Universal agreement from all of the parties involved

- It could be convenient if both T1 and T2 monthly reports show the CPU accounting values (both CPUtime and WALLtime) using the same units: **HEPSPEC06-days**

PIC
port d'informació
científica

We indeed measure the CPU eff. factors

Ian Bird reported that at the last RRB, following the recommendation of the CRSG, it was proposed to remove efficiency factors from WLCG resource accounting. These factors were originally introduced to ensure that the funding agencies were aware of what level of usage of CPU and storage was reasonable to expect. This is now generally well understood, and the actual usage is very high, and the continued use of the factors leads to some confusion. ***Therefore, there is general agreement that these efficiency factors have served their purpose and now should be removed from resource accounting reports****. The experiments will continue to use the agreed efficiency factors in estimating their resource needs. Ian Bird invited MB members to let him know of any objections to this proposal, in the absence of which this will be considered as approved.*

PIC
port d'informació
científica

We indeed measure the CPU eff. factors

- We are **indeed measuring the CPU efficiency factors**:
  - The CPUTime/WallTime reported to WLCG reflects the actual CPU Efficiencies of the jobs at the sites
  - They indeed include all of the inefficiencies, with contributions coming from (empty) pilots, half-full/half-empty multicore pilots, high memory jobs in multicore pilots or alone, ...
- We **should regularly use/check the CPUEff. Information:**
  - to cross-check that the observed values are reasonable
  - to know if the efficiencies that the experiments use are realistic or not

PIC
port d'informació
científica

(next) for the CPU accounting

- We are now scheduling very **different types of job**s + experiments would like to **pass all of the job parameters** to the batch systems
- We should re-define how we make the accounting
  - **[a]** 1 job asking for 2 GB of mem. != **[b]** 1 job asking for 12 GB of mem.
  - Indeed, job requirements != resources blocked by the jobs
  - In [b] 6 cores might be blocked – but current CPU accounting can yield the same result of cputime and walltime for both [a] and [b]
- Experiments and sites consulted agree that we should move towards an accounting that is done on **locked resources** at the sites
- The need of clear procedures/guidelines for these job submissions to **maximize farm utilizations** (such as WLCG multicore Task Force)
  - These inefficiencies can hardly be charged to the experiments

PIC
port d'informació
científica

Conclusions

- (partial) conclusions:
  - Proposal to correct the monthly LCG Office accounting reports since Apr. 2015
    - use CPU WALLtime usage when comparing to CPU pledges
  - Proposal to create a group to drive the discussions on accounting for the future (nice ideas tha have emerged in these discussions)
    - Accounting based on **locked resources**
    - **Installed capacities** through BDii or the new WLCG IS
    - **Disk buffers** in front of tapes are in the exp. Requests?
    - Some discussions (might) be needed for **Disk and Tape**
    - **"Economic models"**: for example, mechanisms to expose the "memory cost" to experiments so they can make choices on what workload to send where...