In this section we need to describe all transformation which happens to the accounting data at all levels.

How CMS and ATLAS Dashboard convert raw wallclock into HS06

For both ATLAS and CMS, we also have plots in historical views for HEPSPEC06 over pledges. For this, - we first collect once per month the installed capacities of all the sites from REBUS: logical cpus and total hepspecs. By dividing these two number together, we get the average HEPSPEC06 power for one CPU core for a site. We store this in our database (with the complete monthly history). - in the CPU/WC HEPSPEC06 plot, we multiply the CPU/WC consumption with the HEPSPEC normalisation factor that was calculated above and then we transform this in hours.

How Physical CPUs, Logical CPUs and total HEPSPEC get into REBUS

The total number of Physical CPUs, Logical CPUs and HEPSPEC in a site is calculated using the numbers published via BDII in the following GLUE 1.3 variables:

- GlueSubClusterLogicalCPUs
- GlueSubClusterPhysicalCPUs
- GlueProcessorOtherDescription

The total numbers are calculated adding the values of these variables for all the GlueSubcluster objects published by the site in the BDII. However, if two or more subclusters publish the same number of logical CPUs, the values for each of the three variables are taken into account only once, as REBUS considers that the subcluster is doublecounting resources. The exact piece of code in REBUS taking care of this is available here (lines 251-255).

HEPSPEC06 is https://gitlab.cern.ch/exp_dashboard/cosmic/blob/master/django-installed-capacities/config/cron/get-installed-capacitie

so it seems that it is GlueSubClusterLogicalCPUs multiplied by reg_expr = '^.*Benchmark=([\.\d]+)-HEP-SPEC06' that is available in the

processorotherdescription

Note from Alessandra

looking at some of the sites accounting I've found that REBUS and APEL use different values of HS06 (though the value may be the same for many sites).

There are 3 values benchmarking values in the BDII for each computing element. This are three CEs from different sites and within a site depending how the CEs are configured there may be identical or different values in each CE.

GlueCECapability: CPUScalingReferenceSI00= 2445 GlueHostBenchmarkSI00: 0
GlueHostProcessorOtherDescription: Cores=16.42, Benchmark=11.85-HEP-SPEC06

GlueCECapability: CPUScalingReferenceSI00=2220 GlueHostBenchmarkSI00: 2200
GlueHostProcessorOtherDescription: Cores=12, Benchmark=8.88-HEP-SPEC06

GlueCECapability: CPUScalingReferenceSI00=2500 GlueHostBenchmarkSI00: 10355
GlueHostProcessorOtherDescription: Cores=12,Benchmark=10-HEP-SPEC06

APEL uses the scaling reference and if that is not there it uses the BenchmarkSI00, REBUS uses BenchmarkSI00. ATLAS picks the HS06 values from REBUS i.e. the total HS06/total Logical CPUs. So the first site above has 0 HS06 in REBUS and in the ATLAS dashboard I think this publishing is plain wrong; the second is my site we publish the same values, because we don't scale and when we compare the accounting numbers both the wallclock and the HS06 are in line, the third is RAL and they scale and the wallclock shows big discrepancies between the two dashboads while the "work" (HS06 hours) go back in line via two different routes: ATLAS payload walltime*HostBenchmark and EGI dashboard scaled walltime*Scaling Reference. I thought we should use a uniform value between APEL and REBUS but if we do sites like RAL will have the wrong HS06hours value in one of the dashboards. So we need to keep them both because the two dashboards calculate that part differently. We should review what sites publish though because UNIBE is clearly wrong and affects the ATLAS dashboard. Also it may help to identify who scales and who doesn't.

Note from Pepe

- Since the farms are composed by many CPU types, one select a reference, and scale all of the CPUtimes and WallTimes to this reference. This is the CPUScalingReferenceSI00 value. These "seconds" or "hours" are not real time, it's scaled time, that goes to APEL, the the EGI accounting portal. When one selects "CPU Time", we cannot compare the results obtained at different sites, since each site uses different references. Hence, the sum of "times" provided by the EGI accounting portal, and the relative %, are simply WRONG, since the values cannot be compared at all. This table is useless:

The following table shows the distribution of Sum CPU time grouped by TIER1 and VO (only information about **LHC VOs** is returned).

TIER1	alice	atlas	cms	lhcb	Total	%
CA-TRIUMF	0	2,754,550	0	0	2,754,550	2.79%
CH-CERN	6,299,187	1,488,417	1,515,481	1,867,570	11,170,655	11.33%
DE-KIT	1,717,676	3,226,270	610,284	1,877,121	7,431,352	7.54%
ES-PIC	0	1,250,735	742,818	589,017	2,582,570	2.62%
FR-CCIN2P3	1,943,360	4,182,375	1,672,661	2,813,140	10,611,536	10.77%
IT-INFN-CNAF	2,078,560	2,429,679	2,274,949	2,593,821	9,377,010	9.51%
KR-KISTI-GSDC	1,892,382	0	0	0	1,892,382	1.92%
NDGF	661,612	1,390,294	0	0	2,051,906	2.08%
NL-T1	1,241,653	2,540,314	0	1,261,807	5,043,774	5.12%
NRC-KI-T1	1,978,024	1,882,311	0	1,146,558	5,006,893	5.08%
RU-JINR-T1	0	0	1,193,619	0	1,193,619	1.21%
TW-ASGC	0	1,651,324	0	0	1,651,324	1.68%
UK-T1-RAL	3,093,349	10,984,268	2,337,671	9,817,572	26,232,861	26.62%
US-FNAL-CMS	0	0	5,037,856	0	5,037,856	5.11%
US-T1-BNL	0	6,517,383	13	0	6,517,396	6.61%
Total	20,905,804	40,297,921	15,385,352	21,966,607	98,555,684	
Percentage	21.21%	40.89%	15.61%	22.29%		

Click here for a CSV dump of this table
Click here for a Extended CSV dump of this table
Click here for XML encoded data

- The EGI accounting portal uses CPUScalingReferenceSI00/250 to scale these values and get the HS06·hours, which is called the Normalised CPU (or Elapsed) time. This is correct, since the different references are then taken into account, and then we can compare the values at the different sites, and get the relative contributions, etc...

- NOTE: Sites report the accounting values of the pilot jobs, for those VOs running pilots, which are the 'jobs' that the farms see. The dashboard of the experiments show the information coming from the payloads. So, one cannot simply go to the dashboard and get the core-hours and compare to these scaled times reported by EGI, since you are measuring different things, and in different scales (the jobs running at the sites report the exec. time on the node, no scale factors are applied).

- We can argue if the HS06 is the best benchmark we have, but it's pretty trivial with the MJF enabled at the sites for each payload to report to the dashboard about the consumed HS06·hours. This would be great.

In PIC we have a reference of 3050 SI00. Our average reported power (GlueHostBenchmarkSI00) is GlueHostProcessorOtherDescription.Benchmark(12.1205)*250 = 3030. But, as of today, we have switched on more machines, and the average HS06/core is 13.1, which corresponds to an average power of 13.1 HS06/core. This is a small effect, and indeed, we can check if the sites are publishing correctly this info, by simply ldapsearch this info:
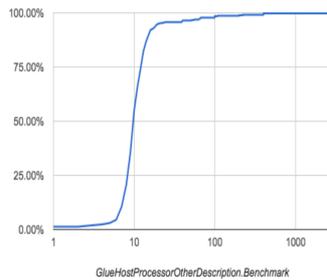
ldapsearch -x -h lcg-bdii.cern.ch -p 2170 -LLL -b o=grid | grep "HEP-SPEC06" | grep "Benchmark" | cut -d"=" -f3 | cut -d"-" -f1 | sort -n | uniq -c

95% of the published CE's have GlueHostProcessorOtherDescription.Benchmark < 20:

Note from Alessandra
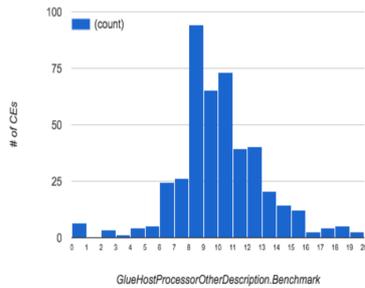
There are CE's with values that are REALLY big, and to me, these should be corrected if we want to draw conclusions from this:

The most dangerous value to be wrong is the CPUScalingReferenceSI00, since it has direct influence in the EGI accounting portal, if the times are not well scaled. Having a look to this value CPUScalingReferenceSI00/250, this is what I observe from the BDii:
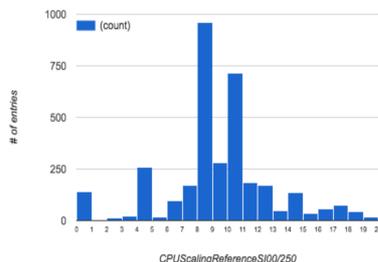
98% of the published values (multiple entries per CE, but here we look how the numbers are) are <20. 2% of the reference values seem unreasonable, with max. at 236 HS06/core! So, if these sites are really referencing to these high values, their reported scaled cpu time in APEL should be really small, since they are referencing to a very powerful (inexisting!) CPU type. Let's see if this indeed happens.

The site reporting CPUScalingReferenceSI00=59000 is ULAKBIM. For May 2016, the EGI "Sum CPU time" value was 278.893 "hours". The EGI "Sum Normalised CPU time" was of 65818773 HS06·hours. You see that the ratio is 236 (59000/250), consistent with this finding.

Then, is the reported "scaled" CPU time correct? or... Is the reported reference incorrect? You see the impact in the accounting if this is wrong... The site might be reporting a factor x10 more used CPU time than indeed done... how to verify this?

[jflix@ui03 ~]$ ldapsearch -x -h kalkan1.ulakbim.gov.tr -p 2170 -LLL -b o=grid | egrep "CPUScalingReferenceSI00|GlueHostBenchmarkSI00|GlueHostProcessorOtherDescription" | sort | uniq GlueCECapability: CPUScalingReferenceSI00=59000 GlueHostBenchmarkSI00: 59000 GlueHostProcessorOtherDescription: Cores=12,Benchmark=5.19-HEP-SPEC06

The site reports an average power of 5.19 HS06/core, and it's referencing to 236 HS06/core, however the GlueHostBenchmarkSI00 is incorrect, it should be 5.19*250 = 1298.

So, at this point, we need to think how to proceed. To me, since the reference is arbitrary, it would be really useful that all of the sites reference to the same CPU type, or the same value. This would simplify things a lot, and it would be really useful to detect publication errors, since this have a direct impact on the delivered HS06·hours reported in the EGI accounting portal.

Also, if we could have MJF enabled and the payloads would report the HS06·hours to the dashboard, it would be really really useful.

-- JuliaAndreeva - 2016-06-23

This topic: LCG > DataNormalization
Topic revision: r2 - 2016-07-12 - JuliaAndreeva