



Definitions

Information System Task Force
8th January 2016

What we are trying to define

- For GLUE 2
 - Logical CPUs
(GLUE2ExecutionEnvironmentLogicalCPUs)
 - Benchmark Value (GLUE2BenchmarkValue)

Existing definitions in WLCG (I)

- Definitions in the Installed Capacities document:
https://twiki.cern.ch/twiki/pub/LCG/WLCGCommonComputingReadinessChallenges/WLCG_GlueSchemaUsage-1.8.pdf
- i.e. `GlueSubClusterLogicalCPUs`
 - *Total number of cores/hyperthreaded CPUs in the SubCluster.*
 - *In other words, LogicalCPUs counts the number of computing units seen by the OS on the WNs. Sites typically configure one job slot per logical CPU, but some sites allow more than this (e.g. five jobs per four cores) to allow for the fact that jobs spend some time waiting for i/o. The 1.3 GLUE schema does not allow such an over allocation to be published explicitly.*
 - *Please note that again LogicalCPUs is a static number manually configured by the system administrator at a site and does not reflect the dynamic state of the WNs*

Existing definitions in WLCG (II)

- What about Benchmarks?
 - At the time the Installed Capacities document was written, it was the transition between SpecInt2000 and HEPSPEC
 - The proposal is in any case to find the power per core (dividing the total score by the number of cores in the box)

More documentation referenced by WLCG

- How to define SI00, HEP-SPEC, MaxCPUTime, MaxWallClockTime
 - <https://wiki.egi.eu/wiki/Tools/Manuals/TS190>
- For historical reasons, even if a Glue SubCluster is an homogeneous set of WNs, sites normally define one heterogenous subcluster per cluster per CE and publish an average WN information
 - Double counting problem (!)
 - Can we get rid of this in GLUE 2?

Proposal by Andrew McNab [1]

- *GLUE2ExecutionEnvironmentLogicalCPUs: the number of single-process benchmark instances run when benchmarking the Execution Environment, corresponding to the number of processors which may be allocated to jobs. Typically this is the number of processors seen by the operating system on one Worker Node (that is the number of "processor :" lines in /proc/cpuinfo on Linux), but potentially set to more or less than this for performance reasons. This value corresponds to the total number of processors which may be reported to APEL by jobs running in parallel in this Execution Environment, found by adding the values of the "Processor" keys in all of their accounting records.*
- *GLUE2BenchmarkValue: the average benchmark when a single-process benchmark instance is run for each processor which may be allocated to jobs. Typically the number of processors which may be allocated corresponds to the number seen by the operating system on the worker node (that is the number of "processor :" lines in /proc/cpuinfo on Linux), but potentially set to more or less than this for performance reasons. This should be equal to the benchmark ServiceLevel in the APEL accounting record of a single-processor job, where the APEL "Processors" key will have the value 1.*

Proposal by Brian Bockelman [2]

- **GLUE2ExecutionEnvironmentLogicalCPUs:** *the number of single-process benchmark instances run when benchmarking the Execution Environment.*
- **GLUE2BenchmarkValue:** *the average benchmark result when $\$(GLUE2ExecutionEnvironmentLogicalCPUs)$ single-threaded benchmark instances are run in the execution environment in parallel.*

Feedback from sys admins

- Several sys admins contacted

1. *Describe with as many details as possible the HW you will use in this example with the terminology you normally use. Add as many explanations and details as possible so we can understand your numbers and how you calculate things.*

2. *Read definitions in [1] and provide the numbers for your example according to what you have understood:*

- *GLUE2ExecutionEnvironmentLogicalCPUs*
- *GLUE2BenchmarkValue*

3. *Do the same for [2]*

- *GLUE2ExecutionEnvironmentLogicalCPUs*
- *GLUE2BenchmarkValue*

4. *Which definition you prefer?*

5. *Is it crystal clear on what needs to be provided after reading the definition? If not, what is it? What are you missing?*

- Very useful feedback from Steve Jones, Alessandra Doria and Manfred Alef.
 - Thanks very much!

Feedback from Steve Jones

- Both definitions give the same figures
 - Preference for definition [1] with benchmarking process properly documented
 - See attachment in Indico with benchmarking process description
- Very useful GridPP publishing tutorial
https://www.gridpp.ac.uk/wiki/Publishing_tutorial
 - We should have something similar!
 - It contains references to Accounting and includes other attributes that may be also relevant to define in this context
 - *i.e. Scaling factors*
- Notes on how these attributes are published at Liverpool
 - Both GLUE 1 and GLUE 2 published in the same way
 - 3 separate clusters
 - CREAM/TORQUE (YAIM), ARC/CONDOR (manual) and VAC (manual)
 - All clusters contain heterogeneous resources
 - All nodes are scaled to a baseline machine (reference to which real nodes are scaled)
 - Average values for logical CPUs and HEPSPSPEC for the nodes inside the cluster is published

Feedback from Alessandra Doria

- Both definitions give the same figures
- Both definitions are clear but she would make some changes
- The definitions focus on the benchmarks
 - Benchmarks are important but do not determine the number of Logical CPUs
 - Good to clarify that benchmarks must be run for the chosen number of processors which may be allocated to jobs as done in [1]
 - Good to make references to APEL
- Notes on how these attributes are published in Napoli
 - Both GLUE 1 and GLUE 2 published
 - 2 separate clusters/subclusters of PBS/CREAM
 - Both clusters contain heterogeneous resources
 - Average values for logical CPUs and HEPSPEC for the nodes inside the cluster is published

Feedback from Manfred Alef

- Both definitions give the same figures
- Preference for definition [1] with some minor modifications
- Notes on how these attributes are published at KIT:
 - Only GLUE 1 published so far
 - 8 CEs/clusters/subclusters of CREAM/SGE
 - In reality 1 single heterogeneous cluster
 - Average values for logical CPUs and HEPSPEC for the nodes inside the cluster are published

Next steps

- What about writing something like the GridPP Publishing tutorial for GLUE 2?
 - We should check whether it is fine for non GridPP sites
 - We will use the definitions for Logical CPUs and benchmark agreed within the TF
 - Could we aim at publishing one Execution Environment per set of homogeneous resources? Or do we let the site decide what they prefer to publish?
 - Are average values of heterogeneous nodes good enough?
 - Do we need to document also the benchmarking process or is this done (or could be done) by each site in a different way?
- Then we can do a validation campaign
 - We need to define validation criteria for this!