



Update on data validation

Julia Andreeva CERN

WLCG Accounting Task Force meeting
23.06.2016

Main objectives

- To have by the end of August :
 - EGI accounting portal with a dedicated WLCG view which contains all necessary information in a table and/or graph form available also through APIs. Should be bug-free and user-friendly
 - Data exposed through the portal should be correct. We should have in place an automatic data validation procedure, which would allow us to detect eventual problems with data publishing or/and processing
 - The accounting reports with an agreed content should be generated by the EGI portal. This should be possible to do on demand for any given time range with monthly granularity.

First results of data validation (1)

- We do not compare exactly the same thing.
 - Experiment systems measure real payloads, APEL measures pilots, therefore EGI should have slightly higher numbers
 - We do not always have raw wall clock in APEL, since some batch system rather report scaled wall clock. However in this case normalized metrics should still agree (RAL, Prague)
- Agreement we see so far is pretty good and we can conclude that overall accounting data is reliable
- However we detected quite some problems with certain sites which need to be understood

First results of data validation (2)

Experiment	Performed comparison scaled/raw wall clock multiplied by number of cores	Agreement EGI/Experiment	Performed comparison normalized (HS06) wall clock multiplied by number of cores	Agreement EGI/Experiment	Comment
ATLAS	YES	~105%	YES	~95%	Using ATLAS Dashboard API
ALICE	YES	~104%	NO	-	Normalized metrics are not available
CMS	YES	~132%	YES	~132%	Using CMS Dashboard API
LHCb	NOT	-	NOT	-	Waiting for API or data extraction from DIRAC

First results of data validation (3)

- **Cases to be investigated:**
 - EGI usage is much lower (more than 30-40%) both in terms of raw and normalized metrics compared to the experiment systems
 - EGI usage is several times higher both in terms of raw and normalized metrics compared to the experiment systems
 - Big discrepancy for normalized metric, while raw metric does agree
- **Most striking discrepancies for ATLAS are already understood**
- **Enabling automatic publishing of the comparison results to SSB**

Normalization activity

- Sub-group (coordinated by Alessandra and Miguel and Pepe) which will investigate and document all data transformations happening at all levels (site, APEL, experiment-specific monitoring systems)

raw wall clock (time) -> scaled wall clock (time) ->
HS06 normalized wall clock (work)

Conclude whether procedures are correct/accurate enough, derive recommendations regarding normalization, benchmarking and APEL/experiment systems reporting.