# WLCG Accounting Task Force Update

## Julia Andreeva CERN IT
on behalf of the WLCG Accounting Task Force

## GDB
## 11.01.2017

# WLCG Accounting Task Force

- Task Force has been setup following the outcome of the WLCG accounting review at the GDB in April 2016

- Composition of the task force team: members from the experiments, sites, WLCG operations, APEL and EGI accounting portal developers

- All activities of the task force are documented on the twiki

# Main objectives of the task force

- Validation of the WLCG accounting data presented by the EGI accounting portal

- Coordinate with the EGI accounting portal developers in order to deploy  a dedicated WLCG accounting view providing all necessary information in a table and/or graph form available also through APIs. Should be bug-free and user-friendly

- Enable generation of the T1 and T2 accounting reports by the EGI accounting portal. The reports should use data reported by sites to APEL. Get rid of manual updates of the CPU accounting data in REBUS. Report generation should be possible on demand for any given time range with monthly granularity.

- For space accounting, agree with the experiments and sites on the functionality which is required for the WLCG storage space accounting system considering the use case of the high level overview of the total and available space

- Understand whether experiments are interested in enabling accounting of the opportunistic resources in the EGI accounting portal and propose possible implementation

# Validation of the WLCG accounting data presented by the EGI accounting portal (1)

- Data in the accounting portal was validated by comparison with data in the experiment-specific accounting systems for all 4 experiments.
- Experiments measure CPU consumption of their payloads, while APEL gets data from the batch systems which implies that it accounts pilot CPU consumption.Therefore,100% consistency is not expected due to pilot overhead.
- Two metrics were compared

  - wallclock time. In case of APEL  wallclock time as it is reported by the batch system, could be scaled by the batch system. In case of experiments, payload time as it is measured by the experiments accounting systems, always raw, no scalability factors applied.

  - wallclock work, same as "Normalised Sum elapsed * Number of processors" in the old version of the accounting portal. In case of APEL represents wallclock time as reported by the batch system and multiplied by benchmarked HEPSPEC06 power of a given CPU resource(published in BDII) and by number of processors. In the experiment-specific accounting systems raw wallclock time is multiplied by number of processors and by a benchmarking factor. Not all experiments provide this metric. Benchmarking factor is not necessary consistent between APEL and experiment accounting system.

# Validation of the WLCG accounting data presented by the EGI accounting portal (2)

- For ALICE, ATLAS and LHCb the checks showed good consistency (difference ~5%, at least for wallclock time metric).
- For CMS the difference is considerably higher. After investigation with the CMS experts this difference was explained by a higher pilot overhead (more details)

**Atlas T2 wallclock time comparison for April 2016**
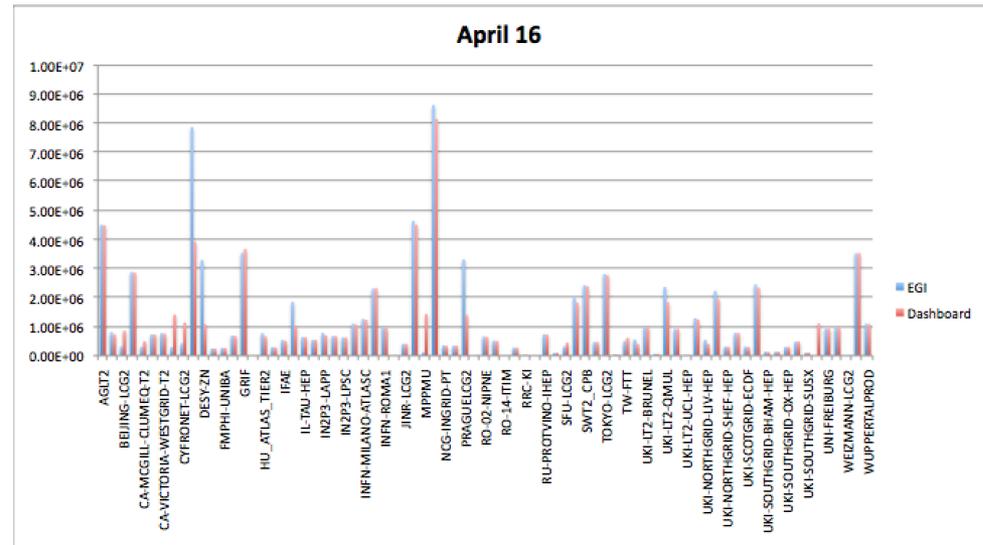


- More details about data validation can be found here

- The fact that data in the experiment-specific accounting systems well agrees with the data provided by the EGI accounting portal allowed to make a conclusion that data provided by the EGI portal is trustworthy.

# Validation of the WLCG accounting data presented by the EGI accounting portal (3)

- However, validation allowed to detect problematic sites, which do not publish correctly accounting data to Apel or benchmarking information to BDII, these sites are being followed up and fixed
- Some of the encountered problems:
- ARC/HTcondor parser: when flocking is active the current version doesn't see all the jobs anymore → under reporting in APEL
- Wrong internal scaling on a big portion of resources → over reporting in APEL
- Wrong HS06 in the BDII → wrong reporting in APEL and ATLAS
- Resources not reported in the BDII → work done disappears from ATLAS
    - For example sites with cloud resources or non traditional sites using VAC
        - VAC and traditional batch system are affected if they don't adjust the BDII to include them
        - VAC only sites have no BDII → work is not reported in ATLAS similar to the next problem
- BDII seemingly correctly publishing but REBUS doesn't see the site → ATLAS doesn't record any work though the wall clock time is there
- Wrong DN/missing service in GOCDB → missing resources in APEL
- Site capacity misreported in REBUS → ATLAS numbers are affected
    - Typically sites on university clusters with highly variable usage
- Wrong benchmark: official benchmark is still HS06 **32bit not 64bit** → overeporting both in ATLAS and APEL
    - The comparison ATLAS - APEL may be green but the accounting is still wrong
- Event Service in ATLAS (still under investigation) → ATLAS under reporting
- APEL clients stops publishing → under reporting in APEL
- Very nice accounting FAQ for site administrators from Alessandra Forti

# Automation of data validation

- In order to facilitate detection and debugging of the corner cases, an automatic validation procedure has been put in place using Site Status Board (credits to CERN openlab summer student Dimitrios Christidis).

- Data from the EGI accounting portal and experiment-specific accounting systems is being uploaded monthly into SSB. Two metrics (wallclock time and wallclock work) are considered. Ratio between measurements from two sources is calculated. Colour for a particular metric depends on the detected difference between two measurements. Latest snapshot and historical distributions are available.



- More details about automation of data validation can be found [here](here)
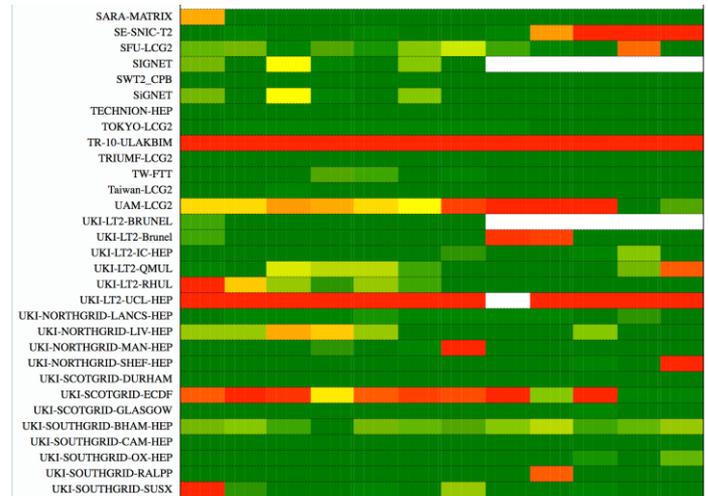
# T0 accounting

- Due to complexity of the heterogeneous T0 cluster, up to now manual updates of the T0 accounting data were required.
- This issue has been addressed by setting up of the new T0-Apel report generation procedure (summary records) which takes into account all types of T0 resources (credits to Miguel Coelho dos Santos)

## Architecture
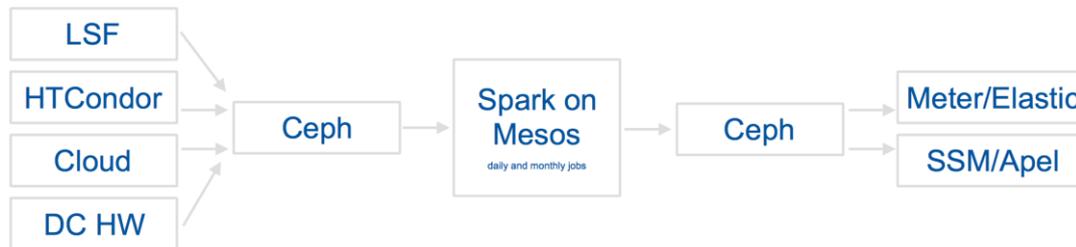


- Validation of data generated by the new T0 accounting procedure has been performed (still ongoing) in collaboration with the experiment experts.
- CERN accounting data has been re-published to APEL starting from April 2016
- See more  more details about new T0-APEL report generation

# New WLCG view in the EGI accounting portal

- New EGI accounting portal has been deployed for validation in the beginning of summer.

- A dedicated WLCG view : accounting-next.egi.eu/wlcg

- Validation of the WLCG view in the portal has been performed. Very good collaboration with portal developer Ivan Diaz Alvarez. Quick bug fixes and implementation of the feature requests

- After validation, the new portal has been publicized to the sites and experiments via the WLCG operations coordination meetings. Asked for feedback. So far feedback is positive

- Multiple changes have been implemented: introduced possibility to chose time/work units, tooltips and mouse-over help provided, new metric naming convention

# New accounting reports

- Currently accounting reports are generated by two different systems: REBUS (T1) and EGI portal(T2)

- Generation of the T1 accounting reports moved to the EGI portal. REBUS code has been re-used

- Main changes in the reports:

  - wallclock work (wallclock multiplied by number of processing cores and by benchmarked HEPSPEC06 power of a given CPU resource) became the main CPU consumption metric which is compared to pledges (instead of CPU work)

  - installed CPU capacity removed from the reports

  - add possibility to switch between HEPSPEC06 hours/days units

  - new metric naming convention has been introduced

  - reports can be generated on demand for any given time range with monthly granularity

- New reports have been validated by the members of the task force

- Two sets of reports (current and new ones) have been already sent for November

- If no serious problems are detected, new reports become official starting from January 2017

# Storage space accounting (1)

- There are two main use cases for storage space accounting which experiments currently deal with:

    - Very detailed space accounting with a possibility of the full storage dump. Implies experiment-specific requirements and data organization

    - High level space accounting (few numbers : total/used/free per VO and where/if possible with some level of details for high level quotas)

- As much as WLCG infrastructure is concerned the second objective is more important and looks to be easier to address in a common way

# Storage space accounting (2)

- The goal of the task force was to assess the possibility to enable WLCG high level space accounting which is not based on SRM

- Agree with the experiments on high level accounting requirements ( Proposal and discussion details)

- Agree with the experiments on the common format of the 'space usage record' (in progress). Has been discussed at the pre-GDB data management meeting.

- Work with storage providers in order to enable generation of the required space usage reports (work will be coordinated by the WLCG data management steering group)

- Collect, store and visualize storage space accounting information. This work is planned for this year though it is out of scope of the WLCG Accounting Task Force.

# Accounting of the opportunistic resources

- Some of the experiments would like to be able to account all resources used by them via the central accounting portal

- In order to make it possible and to transparently integrate opportunistic usage in the WLCG accounting, two issues to be resolved : topology and benchmarking of the opportunistic resources

- Computing Resource Information Catalogue (CRIC) will provide topology for the opportunistic resources

- Benchmarking working group is focusing on the benchmarking solution which hopefully can be shared by all experiments

- Already now experiments do account usage of the opportunistic resources at least in terms of wallclock time. If experiment accounting systems provide both wallclock time and wallclock work metrics, they can generate accounting reports similar to ones provided by OSG, CERN, NDGF and NIKHEF which can be digested by APEL. This is a currently foreseen scenario, though more work is required in order to make it happen.

# Things to be followed up

- Further improvements of the dedicated WLCG view of the accounting portal (some additional functionality was proposed including new plots)
- Benchmarking. Can we have a common benchmarking procedure shared by all experiments and APEL?
- Publishing of benchmarking information and keeping it uptodate, in particular if we stop using BDII
- HTCondor accounting in APEL
- How to improve debugging of the accounting problems? Having raw wallclock time published for all sites in APEL would certainly help

# Summary

- The main tasks of the WLCG Accounting Task Force have been successfully accomplished (data validation, dedicated WLCG view and new accounting reports have been enabled)

- Established close collaboration with APEL and EGI accounting portal development teams, which allowed to progress quickly. We hope that this collaboration will continue

- There are still tasks which require more work, namely storage space accounting and integration of the opportunistic usage in the WLCG accounting system for experiments which expressed their interest.