EOSC-hub

# The Accounting Portal

*Iván Díaz Álvarez (CESGA)*

*November 3, 2020*

eosc-hub.eu

@EOSC_eu

**Dissemination level**: Public

EOSC-hub receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 777536.

# Overview

- The Accounting Portal is a service and web portal to access and explore Accounting data from scientific computation at sites around the world.
- Covers mainly computational resources for High Performance Computing (Grid) and Federated Cloud sites.
- Currently including 185 HPC sites and 24 official Federated Cloud sites in 44 countries, and many more historical.
- The data from these sites is collected by APEL, summarized to make it computationally efficient and correlated with other information sources.

# Information Sources and metrics

- These sources include:
  - APEL for the raw Accounting data
  - EGI's  GOCDB and WLCG's CRIC for topological data.
  - EGI Operations Portal to get VO and scientific discipline data.
  - EGI CheckIn for AAI, user VO membership and privileges.
- Metrics are the numerical data measured in accounting, these are direct measures or derived, and include:
  - Number of jobs or Virtual Machines.
  - CPU time, this can be raw or normalized, which uses a benchmark to make hours congruent from machine to machine.
  - Wall CPU hours, both normalized and not.

# Information Sources and metrics (II)

- Normalised Wall time (Elapsed) * number of processors. This metric further improves comparison by having the number of logical cores in account.
- Efficiency, this calculates from CPU and Wall time how much time has been involved in actual computation instead of I/O and housekeeping.
- Monetary Cost, which estimates a equivalent cost for the accounting based on a per-site rate.

- For Cloud only:
  - Inbound and Outbound Traffic.
  - Memory

# Variables

- These data and metrics can be grouped by:
  - Month, Quarter, Half Years and Years, the base granularity is monthly data.
  - Country and NGI, with separate topologies for both.
  - VO, these can be All, top 10, LHC (4 VOs), and EGI Official.
  - Number of Processors or Nodes
  - SubmitHost.
  - Virtual Machine Status.

# Special Views

- The Discipline View displays data pertaining to VOs that fit in a particular scientific discipline, disciplines are arranged in a three level hierarchy and a given VO can belong to any number of disciplines.
- Restricted views are only accessible with AAI credentials, since they show UserDN information that can be sensitive:
    - VO Admin views: This view shows all the accounting for a VO, including detailed user records.
    - Site Admin: This restricts accounting for a site, but shows user data.
    - User View: This shows the data for a individual user, but not anything from other users.
    - All these require appropriate credentials in CheckIn, User Data is anonymized automatically after 18 months+

# Reports

- Reports use the same data tables as the previous views, but collate and aggregate them in custom manners to facilitate reporting to organizations:
- Currently we have the following reports active in production:
  - **InterNGI**: It uses the UserDN information anonymously to assign each user to an institution and then to a NGI, and calculates how many resources they consume from each other.
  - **Resource Centres**: Does a listing of the consumption of sites by top level scientific disciplines, it can show the top 10, 100, or all alphabetically or by country.
  - **Disciplines**: Shows several metrics aggregated by VOs from a number of selected disciplines.
  - **VO Metrics**: Shows the level of activity of VOs.

# Infrastructures

- Also, the Portal has a Infrastructures menu, this includes:
  - WLCG Support, including Tier1 and Tier2 for both HPC and cloud. These views have a topology derived from CRIC, and the metrics have names and definitions different than the usual views.
  - OSG view, this supports the Open Science Grid sites, with the WLCG metrics and their own topology. OSG have their own accounting service, but pass their accounting data already summarised to APEL periodically

# Design

- The Portal is implemented using Django and the Dojo Js library, using technologies like AJAX, HTML5, CORS, JSON.
- New Interactive graphs with support for zooming, panning and data point query, and dynamic capabilities on page elements (e.g. sortable tables).
- Human-friendly URL structure for easy customization and bookmarking.
- Breadcrumb capabilities and backtracking capabilities, you can easily jump backwards or forwards on your browsing history
- Tables generate navigable links, so you can go down hierarchically or jump to other sites depending on the context.
- Apart from the reporting views, which have to process millions of lines and have appreciable delay, most views are generated quickly.

# Design (II)

- For the relational database backend we use MySQL, along with Apache and the backend code and scripts in a self-contained VM.
- Many of the table schemas are determined by APEL and other operational tools, but there is extensive index optimizations, since writes are very sporadic.
- Raw data available for users in JSON/CSV formats.
- There is also a API to access raw data directly, and endpoints for use of other Core Tools like the Ops Portal, and for monitoring probes and other tools.

# Some Recent Developments

- AAI Implementation using mod_auth_mellon and SAML that allows users to be identified using different credentials and identity providers as VO or Site Admins in order to access the Restricted Views.
- Now the cloud topology uses the Scope modifier in GOCDB, so only sites officially in Fedcloud are displayed, since some sites published data without being registered
- Changed the topology backend to support CRIC as a new topology source, and new WLCG long form federation names.
- Support for Jupyter Notebook sites.
- Support multipliers in metric units.
- Removed inherited Tier1 REBUS report.

# Some Recent Developments (II)

- Implemented EUDAT storage accounting as a static report that doesn’t use inconsistent date information.
- Storage accounting is implemented, but waiting for summarisation to be moved to APEL to move to production.
- Improved indexation of some tables to speed up response.
- Added a new Cloud based graph to the Welcome Page.
- Fix Date display on locales that are on the GMT - hemisphere. That caused small errors on the ending of a month only on non-european machines.
- Improved PDF creation, legend positioning and timestamping on some reports.
- Better unit support for Cloud UserDN and units in Storage accounting.

# Future

- Improve performance and responsiveness of the portal
- Automated testing using Selenium framework
- Implement memcached functionality
- Create a separate model Layer Unit Test suite
- Automated regression testing
- Improve virtualization layer
- Port code to Python 3
- Evaluate other Web server possibilities
- Optimize queries
- Stress testing
- Security Testing

## Thank you for your attention!

*Questions*?

## Contact

Accounting Portal developer mailing list: grid-admin@cesga.es

Also: idiaz@cesga.es

EOSC-hub

eosc-hub.eu @EOSC_eu