

Initial notes about Scoreboard

Marco Mambelli, Tanya Levshina
2/12/2014

Here we investigate the need and describe the scope of Scoreboard, a tool to provide Fermilab experiments with easy access to accounting information for their planning and operational needs. This document is the summary of the meetings we had with Scientific Computing Division management, Intensity Frontier experiments liaisons and researchers.

Use cases

We were able to identify some common use cases:

- Capacity planning.
The most common need is to use accounting data for capacity planning: to provide aggregate and by experiment high-level views of resource usage. Experiments need to document the resources they used to evaluate future needs, to do internal planning and to make the case to DOE, OSG or the Lab Computing Division. Aggregated experiment data includes average and peak CPU use and space allocated on different kind of storage. The same information is useful also for the Lab resource managers to plan allocations, to verify if the requests match a real need. Capacity planning normally is performed two, three times a year, for annual reviews. Someone mentioned also the presentation of status updates at collaboration meetings, every three or four months.
- Pledge monitoring.
Identify users of the resources to monitor resource pledges and usage for sub-projects. Graphs plotting which activities consume more resources, e.g. simulation or analysis. And graphs documenting the computing cycles contributed by Universities, the Lab and other facilities can be used at the collaboration meetings to tribute proper recognition.
- Production trends.
CPU and storage resource usage is used by some experiments also to monitor trends, e.g. on a weekly basis, or to discover problems, suggested by unusual usage patterns.
- Job efficiency and failure monitoring.
It is common also the desire to keep track of job efficiency (ratio between CPU time and wall time) and of failed jobs. Identify patterns and correlations to find and remove inefficiencies and failures, to have more successful and efficient jobs. Where are inefficient jobs? Who sent them? What do they have in common? How many jobs failed? Are they all on the same host or resource? This information is used continuously and historic data helps for long-term patterns.
- Experiment accounting summary.

Many experiments asked for an experiment dashboard, to have a quick view on their status, progress and trends. They are sometime vague on the content but they agree on it being a quickly available snapshot collecting multiple information.

Preferred resources

Being familiar with a tool, knowing how to use it and what it can provide is more important than all the features. To use Steve Wolbers's words "It's really a question of getting familiar with a tool. I'm looking at Keith's page because I'm used to it"

Each person interviewed has it own set of preferred tools and plots, frequently collected in a start page or in the browser bookmarks.

Here follow some of the recurring tools with links to the Web pages, positive and negative remarks and notes on using them for FIFE accounting.

| |
|--|
| CDF CAF monitoring |
| http://cdfcaf.fnal.gov/caf/cdfgrid/history (requires authentication cdf_auth) |
| Margaret's favorite. Shows what's running, what's waiting, how many users there are. Shows how many jobs are not using CPU (inefficiencies): both in general (summary) and at the user level. There are very nice by user plots/tables: generally there are 1-2 users that are doing something inefficient and you can spot it. There is a fairly good historic plot from the monitoring page. |
| Access to the most pages it is restricted to CDF members. It seemed very specific to CDF. |

| |
|---|
| Fermigrid view (Keith's plots) |
| http://fermigrid.fnal.gov/monitor/fermigrid-merged-total-slots.html |
| Consulted daily by Steve W. It uses the job history from HTCondor on FermiGrid. Shows how people are using opportunistically the cluster. Can compare different VOs: CDF, D0, ... |
| It can show only the batch (HTCondor) slots on Fermigrid |

| |
|--|
| FIFEmon (Joe Boyd's plots) |
| http://fifemon.fnal.gov/minerva/ http://fifemon.fnal.gov/nova/ |
| These plots use RRD databases at different scales and javascript. You can see how efficiently CPUs are used and have by user plots, to find inefficiencies and correct them. |
| Most experiments use these and like the plots. Some would like more flexibility. Some complain that the dynamic plots take too long to be calculated. |
| These document only Fermilab resources and it is difficult for collaborations using also outside resources |
| Provide only a partial view and the time intervals are fixed |

| |
|---|
| Gratia |
| http://gratiaweb.grid.iu.edu/gratia/ https://sourceforge.net/projects/gratia/ https://github.com/ashuguru/OSG-Measurements-Metrics |
| Contains accounting information but people do not know about it and it is difficult to know the query for the desired plot |
| The gratia database contains a wealth of accounting information but the current displays are insufficient for IF experiments: the BIRT has been deprecated for time and is being removed, the Web display is tailored for OSG needs. |

| |
|---|
| MCAS |
| http://nusoft.fnal.gov/minos/mcas/over.html |
| This was a Fermilab project, no more supported, that a scientist decided to maintain to have an overview of the essential stats for the Minos experiment. The page is quick to load (it refreshes periodically a set of images) and provides a nice customizable bird eye view. |
| The idea of a set of images ready for an overview is good but the tool would not help much in displaying Gratia data. |

| |
|--|
| Scoreboard |
| http://fermigrid.fnal.gov/scoreboard/ |
| Scoreboard, its current version, is discussed separately in its own section below since it is a proof of concept for the project discussed in this document. |
| The current pages are limited but received positive feedback and it is using Gratia data. |

As suggested also by Panagiotis Spentzouris we looked into what other collaborations and big experiments are using. CDF monitoring has been mentioned above. Groups like ATLAS, CMS or other Grids like XSEDE devoted time and developers for tools for similar purposes. Here follow some of these tools listed in a similar format.

| |
|--|
| APEL and EGI/WLCG Accounting Portal |
| https://wiki.egi.eu/wiki/APEL http://accounting.egi.eu/egi.php |
| APEL (Accounting Processor for Event Logs) is an accounting tool with sensors, a database and a viewing portal. It is similar in its architecture and functions to Gratia. |
| It is tailored to the WLCG needs. It is an useful comparison but not a solution, specially if we want to leverage the data available in Gratia |

| |
|--|
| ATLAS |
| http://panda.cern.ch:25980/server/pandamon/query?dash=prod http://dashb-atlas-job.cern.ch/dashboard/request.py/dailysummary |

<http://dashboard.cern.ch/>

The first link, the Panda dashboard, provides an overview of ATLAS grid activity. The second is the Experiment Dashboard, It is a nice Python Web Framework (with ORM, MQ and MVC), has nice views and drill-down reports, it is already known and familiar to the HEP community and developers may be known and involved in other CERN projects.

The first is very tied to the Panda workflow manager used by ATLAS, so is not suitable. The second one is one instance of the (ARDA) Experiment Dashboard, a flexible framework used for job and transfer dashboards at CERN. It uses the Oracle DB at CERN but probably may be adapted to adopt other information sources. It evolved considerably in the last couple of years. It would need some evaluation to see how flexible and adaptable it is and if it could be adapted to the IF experiments.

CMS

https://cmsjobmon.fnal.gov/cpu_usage.html

http://cmsdcam3.fnal.gov:8081/DOEmetrics/SiteCommissioningGlobalStatus_Sites.html

<http://dashboard.cern.ch/cms/index.html>

The first two pages are summaries from different systems providing an overview on CMS running jobs, efficiency and the status of the different resources, similar to the MCAS mashup. The last one is a dynamic dashboard using the Experiment Dashboard mentioned in the ATLAS section (Thanks to Burt Holtzman for providing the links!)

The Dashboard, 3rd link, has been discussed in the ATLAS section. The first two pages are summaries The 2 pages with summaries use links to images that are generated from the batch system monitoring or the Dashboard. Probably those are custom coded HTML.

XSEDE Metrics on Demand (XDMoD)

<http://xdmod.ccr.buffalo.edu/>

It is a nice display using HTML5, JavaScript and a PHP framework underneath developed to display XSEDE accounting information.

There are two versions, the XSEDE one and an Open source one. The XSEDE version (with some more feature than the open source version) displays already some data from Gratia. This tool was interesting enough to receive a more detailed evaluation in a separate document [1].

User interface notes

Familiarity plays a big role in using a tool, hence the variety of tools currently used. Anyway here is a list of features that would make a tool more liked.

- **Tailored information:** different people or even the same person when “wearing different hats” would like uncluttered pages, dashboards with

information important for them in that role. Things like experiment dashboards and role based customizations would help providing a better personal experience.

- **Responsive overviews.** Overview pages should load very quickly to allow a quick glance to the experiment status.
- **Both fixed and flexible time selections.** Ready-made intervals, e.g. today, yesterday, this and last week, last month, are quick, provide standard views and allow easy comparisons. Selectable time intervals allow more flexibility, e.g. to prepare the exact plot needed for a report.

Missing information

Here some of the data that scientists would have used but is not available. Gratia collects and stores already a lot of data about the IF experiments, most of the accounting data that we heard interest about. Here follows a list of additional data mentioned during the meetings. An IF accounting system would have to include all data currently in Gratia and also this additional data.

- The **computing capacity**, total number of available resources including the unused ones, is frequently not available. Some clusters provide it but it is not collected centrally. This does not make sense for elastic (on demand) resources like clouds or borrowed opportunistic slots, but it is important for owned or reserved resources, to know if there is possibility of expansion.
- Detailed **FTS (SAM File Transfer Service) accounting.** Data like number of transfers per time period, number of GB per time period, number of errors per time period, would be useful for experiment reviews.
- **Monitoring of storage usage.** Beside the amount of data transferred, the experiments would like to know the actual space used on the different storage systems they have access to: space used, available space and quotas, if enforced. There are already probes for dCache that could collect this information for Gratia. Enstore has its own reporting but people would like to look at various historical and aggregated graphs similar the ones produced by GratiaWeb. Many people expressed the wish to have similar reporting for BlueArc.

Comments on the current scoreboard pages

Generally interviewed people like the current Scoreboard pages and consider them a good start to build up on. "It is good to have a summary page". Most would like an experiment page, with summaries specific to an experiment/VO, one would prefer a personally customizable page, where accounting data would go as part of a personal mashup.

Everyone liked the enlarged picture when clicking on the summary even if they'd prefer the title to be preserved there as well (the cells in the summary table have a title, beside the one in the PNG file).

Generally they would like more consistency across the plots, specially on how different experiments are represented:

- the legends should have entries in alphabetical order
- all the experiments should appear in the legend, even if no data is available (to preserve order and position)
- the color scheme should be consistent across plots and over time (same experiment, same color)
- one scientist (Art) suggested the possibility to highlight the data of one VO, e.g. clicking on it in the legend

Then some terms or units are different, specially across different monitoring or accounting system. Would be preferable to have uniform names and units or at least clear conversions, to compare the results.

Most interviewed were not familiar with all the terms and had question on what exactly the data corresponded to, e.g.:

- What is BestMan?
- How this relate to BlueArc or dCache
- What's included in the "Transfer volume per VO on Fermi facilities"? Are copies to BlueArc (dd and cp) included?
- Why is the computing measured in CPU hours?

These questions were answered mostly with short explanations that could be added at the bottom of the plots.

They found the tables and CSV with monthly summaries very useful. Some presentation changes could improve usability:

- Possibility to hide columns or at least remove the columns with the deltas
- Fix the first row when scrolling the table
- Possibility to sort according specific columns

The additional plots (Grid, Cloud, Storage) sometime were not responsive and there is no progress bar or spinning wheel to signal activity.

Once the calendar widget were not working on one of the tabs, maybe a glitch, and a malformed date results in a Python traceback, leaving the user perplexed. They were eager to report if some data looked suspicious or strange, e.g. if values were not matching other monitoring or reporting systems.

Plots in the scoreboard page are limited and GratiaWeb, that could provide more plots, is not easy to use. Some people do not know about it, other do not know how to get the data they need and it is too much effort to learn. The general reply when showing some plot was something like "Interesting. I like this. This would have been nice to have."

Summary

None of the tools examined can be taken off the shelf and used for the Intensity Frontier experiments without any development. Gratia contains already most of the desired accounting information and with limited additional development could collect the additional information. It makes sense to keep using Gratia to store the accounting information and to find a solution that can display the Gratia content.

The GratiaWeb [4] tool, the current frontend used by Gratia is tailored to OSG and inappropriate for the IF experiments. It would require significant changes to

display the new information and mostly to improve usability and ease to find and access the desired information. Being maintained by OSG and the GOC we'd have a good leverage to get support in the development or at least have our changes accepted in the repository. It would be the quickest way to provide a set of new plots with corrected labels (titles and scales) and some pages collecting significant plots. Further enhancements, like dynamic plots, resizing, exporting data in multiple formats, access control and customization would require substantial new development.

The XDMoD [2] deployment used by XSEDE is fed also with information coming from Gratia, so there should be already filters that ease the interaction. The development to start plotting IF accounting information would be more than in the GratiaWeb case, but the framework would bring features like access control, user roles (different views for different roles), possibility to customize the dashboard (when you login you see your favorite plots). In a test installation during its evaluation [1] we were able to load and display job information from Gratia. The developers offer to collaborate. This could be a valid alternative starting point, especially if some of the advanced features are important.

The (ARDA/Experiment) Dashboard [3] has a generic data access layer currently supporting other databases but not MySQL, Gratia's database. This framework is used also for the accounting of the LHC experiments that uses data from Gratia, so there may be some filter but I have to investigate. It is generating nice displays that can adapt to the available space and can export the data in multiple formats (images, CSV, XML). Allows access control but does not seem to support user customization. Has a very interesting capability to filter the data and select what to display. The developers were very responsive to some initial questions and are used to different customized versions of the Dashboard. There are many instances, all different in part, some used as main Dashboard for the four LHC experiments, some monitoring data movement, some for specific applications. Like XDMoD also adapting the ARDA Dashboard would require a bigger initial development (compared to GratiaWeb) but would allow adding more easily advanced features. I would need a short evaluation to assess how much Dashboard is customizable and if it is a better alternative to XDMoD or not. All three solutions proposed above would satisfy the basic requirements mentioned in the meeting, with the last two covering more than the first one. Whichever solution is chosen, it is important to document well the tool and to let people know about its features because adoption is driven by familiarity. People take advantage of what they know how to use.

References

- [1] XDMoD Evaluation, Fermilab DocDB, CS-doc-5303-v1
- [2] Open XDMoD <https://sourceforge.net/projects/xdmod/>
- [3] ARDA/Experiments Dashboard <http://dashboard.cern.ch/>
- [4] GratiaWeb <http://gratiaweb.grid.iu.edu/gratia/>
- [5] Scoreboard wiki <https://twiki.grid.iu.edu/bin/view/Accounting/GratiaScoreboard>

Acknowledgements

This document is the result of meetings that Tanya Levshina and Marco Mambelli had with different scientists. Some are part of the computing division, managing and allocating the Fermilab resources, some more with administration/managerial roles, some are FIFE experiment stakeholders (directing/running the activities for experiments). Some experiment are running production and were more clear on their needs, some are in early stages and they repeated that this would be the initial wish list that will need some more iterations

Thank you to all the people that agreed to discuss their needs and the tool they use:

- Stuart C Fuess
- Margaret Votava
- Steve Wolberg
- Panagiotis Spentzouris
- Thomas Junk, Elizabeth S Sexton-Kennedy, Maxim Potekhin (LBNE)
- Adam Lyon (g-2)
- Brian Rebel (LArIAT)
- Heidi Schellman, Rick Snider (Minerva)
- Arthur Kreymer (Minos)
- Rob Kutschke (Mu2e)

We tried to contact representatives from all experiments to reduce bias. The following did not reply to the emails:

Andrew Norman, Gavin S. Davies, (NOvA)
Herb Greenlee (MicroBooNE)

Michael Kirby, coordinating FIFE, suggested most of the experiment liaisons and will participate in the process as well.

You can find more material in the Scoreboard wiki page [5]