WLCG Resource Usage Metering Policy

A proposal to be discussed in the GDB

Draft 2
May 30, 2006

Introduction

This is a first draft on a WLCG Resource Usage Metering Policy. Many of the issues have been discussed on several occasions in the GDB but have never been collected together in one document. The ideas herein on user-level accounting were originally presented during the Rome GDB by Dave Kelsey representing the JSPG http://agenda.cern.ch/fullAgenda.php?ida=a057704 . An essential contribution also came from Holger Marten from FZK who discussed this issue with the data privacy commissioner of their center and produced a report. Most comments and recommendations from this report have been taken into account.

Accounting is needed:

- For the applications to find out how much of their resource assignment has been used and for what. This allows them to plan the analysis of the data properly.
- For the resource providers to find out how the resources they contribute are used and by whom. This allows them to (re-)assign their resources properly and plan purchases timely.
- By the WLCG collaboration to find out if the pledged resources as described in the Computing MoUs have indeed been provided and properly used by the applications. This allows for a better planning of the project as a whole.

The above needs indicate that accounting numbers are needed with a precision of order one percent. A better precision may be difficult to achieve but is not really needed for the purposes described above. It is however important that accounting is done the same (again to a precision on order one percent) for all resources and all applications. This requires the metering and registering to be done automatically following the same procedures everywhere. Only in this way we can be sure that corrections and calibrations are done in a correct way and the numbers can be compared.

In accounting resources have a value and can be traded and ultimately an independent unit (call it money) can be used to exchange resources of different value. At this stage that is not what is needed for WLCG. Really what we want is ‘metering’, a normalisable number for the usage of the resources without adding a value to those resources. This does not imply that we will not try to do real accounting at some later stage of the project.

Accounting has been done as long as computing resources exist but almost always within a resource center or other local or national boundaries. Often these numbers were for internal use only and were treated with some care. With accounting on a world wide grid one has to treat accounting numbers with at least the same or even more care as they may display internal or national policies and priorities that were not meant to be made visible in public. Most countries also have rules for the protection of personal information and data retention and those rules have to be respected as well. We have no tradition in this sense in HEP as we mostly 'owned' all our resources and could do whatever we wanted. However, on a grid while using someone else's resources another attitude is mandatory.

Responsibilities for the GOC

- Each job submitted to one of the WLCG grid infrastructures produces an accounting record. The schema for this accounting record is an international standard and is discussed and defined by the Usage Record Working Group UR-WG in GGF http://www.psc.edu/~lfm/PSC/Grid/UR-WG/ People working on accounting in WLCG are actively involved in the URWG and can make sure the schema suits our purposes well.
- Usage Records of all grid jobs are stored in one central data base at the GOC, the Grid Operation Center at RAL in the UK http://goc.grid-support.ac.uk/gridsite/accounting/ The GOC is responsible for the maintenance, protection and curation of that data.
- The GOC is also responsible for the generation of the aggregated statistics from the data and the publication on its portal. It shall take care that all data are corrected and calibrated such that all data can be compared. It is also responsible for the accessibility and protection of that data at the level defined further down in this document.
- The GOC is responsible for the automatic generation and distribution of the printed reports on resource usage for the WLCG Overview Board (OB) and the CERN Computing Resource Review Board (C-RRB).
- For the purpose of accountability within WLCG the only VOs that are considered are ALICE, ATLAS, CMS, LHCb and dTeam.
- The GOC is responsible for deleting the individual accounting records after one year and for preserving the aggregated data for the whole duration of the project.

Responsibility of the Sites

- Each active site in WLCG has to report accounting of all resources, jobs submitted via the grid and storage used for the VOs mentioned above. This refers to the Tier-0, the Tier-1's and the Tier-2's as far as work is concerned for the VO as described in the Computing MoU.
- If a site wants to report into the GOC data base work being done for a VO but not submitted via the grid it may do so but only after agreement with the VO computing management beforehand. This exception is only valid until January 1 2007 after which no non-grid work may be reported any more.

- A site is supposed to preserve its own accounting to be able to check the aggregated accounting statistics as generated by the GOC before being printed and published.
- A site is responsible for sending its usage records to the central data base once per day (24 hours). Sites not following this rule will be flagged on the GOC site monitoring page and alerted by the GOC on duty.

Storage Usage Monitoring

At the moment of writing this memorandum only the usage of CPU resources are monitored and reported to the GOC. Automatic storage metering has a lower priority not because it is less important but because it is more difficult and the massive usage of storage is only just starting now. An additional difficulty is that the discussion on storage classes still has not converged. As we will probably want to monitor storage also according to storage class this issue will have to be taken up later.

Most sites have a conservative approach to purchasing hardware and would like to wait as long as possible and only buy when the storage space is really needed. Typical purchasing and installation times are three months. So to have a timely availability of the storage the plans of the experiments have to be known reasonably precise three months ahead in time. This does not have to be atomized or stored in a central data base, a simple wiki would suffice and as a matter of fact such plans already exist but maybe not with sufficient precision yet. The procedure to set up and maintain a wiki page that can serve this purpose needs to be discussed in further detail with the experiments.

Then after the storage hardware has been installed, the experiments have to make a firm reservation for space so the system administrators can make sure that the space is available when needed. This space reservation has to be accompanied by a period also, so like “ALICE reserves 500 TByte of storage at RAL for 6 months starting from July 1st 2007”. For this a web form can be prepared which puts it in a visible format on the web and at the same time stores it in the GOC data base. The GGF UR Schema can be simply adapted to measure different resources so using it for storage is not a problem.

Reserved space by a VO needs to be counted for as that space can not be used for anything else. However for the experiments to better manage their storage reservation and usage they also need to know how much of the reserved space is actually being used. This quantity should be stored in the GOC database automatically. For this to happen tools have to be provided for the various storage systems DPM, dCache and CASTOR which allow measuring the storage used per VO and per storage class. Once a day these tools should be used to measure the storage used and a Usage Record should be created and sent to RAL. Storage usage can then be reported much the same way as CPU usage is. The outstanding issue is how such resources are aggregated after they are reported. Unlike jobs which are single items which can be counted and summed, storage records will be measurements of a continuously-varying value. Algorithms for evaluating an integral of TB-days will need to be developed.

It has to be seen if the storage classes to be defined match well the data classes from the experiment’s models. If they don’t the experiments may want an account of how much storage is used for Raw, ESD and AOD data to say it in the Atlas language. Moreover it has to be seen if the storage classes distinguish well the data on disk and the data on tape otherwise monitoring of this usage may be required as well.

N.B. All this seems far from the ‘global space reservation’ which should work on the ideal grid where storage is reserved without ever knowing where the data will go and be stored. It seems however beyond the horizon of what is realistically possible for the first LHC running and will not be further discussed in this document.

Data privacy and protection

For many data centers (at least in Europe) laws apply for data privacy and protection and for the storage of and access to data which is linked to a user name. These laws have to be respected. It must be noted however that for the monitoring purposes described in this document no sensitive user information is involved. Generally what is meant by ‘sensitive’ is the user name connected with an (email-) address, phone number or a day of birth. In the case of monitoring only the user name is stored. However any type of data that can potentially be used to record and control the work of individual persons and that allows conclusions about her/his working methods, results, performances, etc is sensitive information and has to apply to strict regulation.

It is therefore sufficient to encrypt the user name in the Usage Record before it is send across the network to the ROC database. From just one Usage Record one cannot derive any ‘sensitive’ data in the sense described above.

However in the database the aggregated data of a user becomes ‘sensitive’ data and must be highly protected. We therefore propose that all user data in the database is anonymous in the sense that a user data can not easily be connected to a user name. A special tool is available where this connection can be made and only one person per experiment has access to this tool. This could be organized by making the special VOMS role of Resource Manager which only one person in the VO can have. Access to the tool can then be controlled by the credentials of this person’s certificate. GridSite the front-end of the GOC database has such capabilities.

The VO Resource Manager will have to sign a document in which is stated what she/he can and can not do with the user related information from the database. It must be clear it will be forbidden (for example) to present on a slide which is shown during the ATLAS general collaboration meeting the usage of the resources by a particular (named) user over the past 6 months. This sort of information must be handled with care and only be used to resolve problems. The usage policy document saying what the VO Resource Manager can and cannot do will be prepared by the JSPG group.

There is another group of people which by the nature of their job have access to the user related data, the GOC developers. These GOC developers will have to sign the same or a very similar document as the VO Resource Managers. By signing the document they become authorized GOC managers and have access to parts of the database that un-authorized people have no access to. It is the JSPG group that will decide to draft a separate usage policy document or make one that works for both groups, the VO Resource Managers and the authorized ROC managers.

There is one more rule to be obeyed: any user has the right to access her/his own usage records and the same mechanism must be implemented as for the VO Resource Manager to access the aggregated user information. The GridSite front-end can make this information available based again on the user's credentials. Not only must the individual user to be able to see which data is stored it must also be possible to change that data if she/he can justify/prove that the stored data is wrong. In that case she/he must contact the corresponding VO Resource Manager and the authorized GOC managers and with such a request. In case of agreement the data in the database must be corrected. In case no agreement can be reached, the VO Resource Manager decides.

Priorities

At the time of writing of this memorandum also the priorities of the work to be done were discussed.

1. A first accounting report is needed for the next POB on June 12. With the comments received at that meeting a second and improved report will than have to be submitted to the C-RRB on September XX. The first report will be generated from the data in the GOC database as far as CPU resources are concerned. For storage usage the data will be collected from the sites by the agreed spreadsheet.
2. The second priority is to provide aggregated data on CPU usage per Group, Role as defined in VOMS, the Virtual Organization Management Service. The Schema exists that has a FQAN, a Fully Qualified Attribute Name where the Group and Role can be specified. The Subject DN already existed in the old schema definition to put the Name from the certificate of the person who submitted the job but was never used. For this to work the new gLite-3.0 software has to be operational. Access to Group/Role aggregated data has to be restricted to members of that VO.
3. The third priority is to provide aggregated data by user name. The access to this data will be only for people from the same VO but the name will be encrypted. Access to a portal that allows transforming the encrypted name into a person's name will be restricted to one person in the VO, the Resource Manager.
4. As fourth priority storage accounting will be added to the GOC data base.
5. N.B. All the above is really about metering. Accounting proper is for after this basic metering works and is beyond the scope of this memorandum.