Requirements to report WLCG Installed Capacity information

Editor: Flavia Donno, John Gordon, Ruth Pordes

Date: 1/28/2009

Version: 3.1.2

Contributors/Authors: WLCG Management Board

Purpose of this document

The goal of this document is to define the requirements for LHC Tier-1 and Tier-2 sites to report installed capacity information and resource usage by the four WLCG VOs.

Background

The WLCG MOUs Current Pledge and Resources table defines the pledges of resources from Tier-1 and Tier-2 sites to meet the needs of the LHC experiments. Both this table and the needs of the experiments are updated annually. The management and oversight bodies are charged with assessing whether the agreed upon set of resources are indeed available, whether the needs of the experiment are and can be met, and do capacity planning and negotiation for the future. The WLCG has been reporting the CPU usage of available resources on a monthly basis for Tier1s for several years and for Tier2s for the last year. However, there has been no automated reporting of the usage of archival and disk storage. Additionally, there is no ongoing reporting of the resources actually available at any site. The funding agencies and oversight boards require this information in order to do their assessments. The management of ALICE, ATLAS, CMS and LHCB have asked for the information to be available on an ongoing basis in order to assess the installed base of resources and plan for the future. The manual reporting done by Tier1s will not scale to all Tier2s so it is an MB requirement that the data collection is automatic.

The WLCG baseline services define the requirements for publishing of dynamic information to enable the resource brokers, job and schedulers to ensure maximal efficiency in the use of the installed base of resources. It is therefore imperative that accurate and complete information be published about the capacity and current availability of the resources available.

Published disk resources should include disk cache in front of tape and other s requested by the experiments, not just disk for permanent files.   Other resources used internally by a site for optimization reasons, testing or operation do not have to be reported.

Requirements

There are two classes of requirements: the first covering the needs for monthly publishing of installed capacity and resource usage; the second providing timely information to be used by the experiments for ongoing resource monitoring and scheduling. The requirements are:

1. *Provide the WLCG management with a monthly view of the "total installed capacity" at a site*. This includes resources that are not currently in use but are available to be deployed at short notice. Measurement errors associated with jobs running over month boundaries and other uncertainties will limit the accuracy. This should be accurate to within 10%.
2. *Publishing information to provide the WLCG and VO management with a monthly view of the resource assignment per experiment at sites*. This includes only resources that have been configured and explicitly assigned to a VO. An exception is made for shared resources. In this case it is accepted to provide only aggregated information for the common usage.
3. *Publishing information to allow experiment operators to monitor the experiment usage of the resources*. This is specifically important for monitoring the usage of storage staging areas or to monitor the usage of disk only areas.
4. *Support for users to make maximal use of the installed resource base.* Any changes made to the information service schema or semantics should not impact existing uses of the information service. Such changes should either be backwards compatible or be made after careful consultation with users and application developers.

Notes:
a. Since the measurements are only required to be monthly it cannot be assumed that any values published are up to date. An agreement needs to be reached on the period during which data are valid. E.g. data could be held always to refer to the previous calendar month.
b. A decision needs to be taken on how disk storage is measured for each month. Possibilities include: the last measurement taken each month, the highest value, the average, a time integral.