

To/Mail Code: Barish, Saulson
From A. Lazzarini
for the LSC Computing Committee
Phone 626-395-8444
FAX: 626-304-9834
Refer to: M040139-00-Z
Date: 6 May 2004

A.L.

Re: Analysis Computing Support Centers within the LIGO Scientific Collaboration

This memo proposes a policy for the LSC on the subject of the hierarchically tiered computing infrastructure that has been implemented by the LSC. This infrastructure, modeled on the grid computing paradigm, is termed the LSC Data Grid. Now it consists of four LIGO Laboratory sites and two LSC institutions, PSU and UWM. Looking to the future, there will come a time when new institutions ask to join this Data Grid. Many institutions perceive that being able to boast hosting of a Data Grid site at one's institution provides certain benefits with regard to NSF recognition that outweigh the drawbacks of having to provide such service to other LSC institutions. In order to ensure equitable opportunity to those institutions wishing to make the significant investment in personnel and resources to become a Data Grid member site, this memo provides a plan by which the Data Grid may be expanded. It includes guidelines to be followed by institutions seeking either to establish a new site or to upgrade their sites through federal funding. Further, the LSC Computing Committee will develop a plan, to be updated yearly, that tracks site resources age and needs. This will be used to coordinate the development of proposals for new sites and upgrades to existing sites.

Statement of Policy

Grid Nomenclature

LIGO follows the concepts and nomenclature for geographically dispersed computing support that has emerged in recent years from the movement toward a grid computing environment in high energy physics and related fields. This focuses on several tiers of computing centers. The breadth of the community they support defines the tiers. Each tier provides support and data for the next lower level (higher numbered) tier. In high energy physics, the Tier 1: Tier 2: ... : Tier N ratios of resources, whether measured in terms of CPUs, TB, etc., typically follow a power law, e.g. 100:10:1:.. distribution of resources.

This note explains the tier concept and defines the tiers in the context of the LSC Data Grid (LDG). The jargon associated with grids is widely prevalent among funding agencies in the US and Europe. Therefore, having clear definitions of LDG tiers is important so that requests for funding for a Tier x computing center from a particular university does not conflict with requests from other institutions or with the LSC's broader plans for funding its computing support.

Tier 0 centers appear to exist only in the context of CERN LHC experiments. For these experiments the raw data archives and the support for them at CERN are the Tier 0 center. Tier 1 centers for the LHC experiments are major national centers that support a whole country or region. Many of the larger European countries are planning to have Tier 1 centers. The US will have one Tier 1 center each for Atlas (BNL) and CMS (Fermilab). These contain full (or at least very large) mirrors of the Tier 0 data and significant computing to support analysis.

The LSC Data Grid

Within the LSC, LIGO Laboratory facilities at Caltech presently play the role of both Tier 0 and Tier 1 and we tend to refer to it now as Tier 1. However, one could imagine that in the future, an LSC European (and even an Australian or Japanese) Tier 1 center might be established. Then Caltech would be the Tier 0 center while also providing Tier 1 services for North America (and Japan and/or Australia, for example).

LSC Tier 2 Sites

A Tier 2 center provides support for subsets of a Tier 1's community. These subsets are fluid and may be defined geographically or along lines of scientific effort. However, the support is expected to be for people at a number of dispersed institutions beyond the local one.

An LSC Tier 2 center is expected to have at least the following resources:

1. 128 processor CPUs, each at the high end of available desktop computing capabilities and with commensurate memory.
2. Storage for the full L3 data set from the two most recent science runs
3. User storage costing the equivalent of 15 TB in 2004

4. An OC3 or higher network connection so that remote users and Tier 3 centers can readily access data.
5. Personnel: at least 1 FTE to support site operations and at least 1 FTE to support users
6. Full and equal access to *all* LSC members who are entitled to access and use LIGO data, without regard to institutional affiliation.
7. Software commitments:
 - Must keep current with OS as specified by DASWG and LSC Computing Committee;
 - Must keep current with all DASWG supported analysis environments (or a subset negotiated with the Chairs of DASWG and the LSC Committee Committee).

Institutions that wish to provide Tier 2 services to the LSC must meet these minimum requirements. PIs who plan to develop proposals requesting federal funding for a Tier 2 site should coordinate by notifying the LSC directorship (LIGO Laboratory Director, LSC Spokesman) and the LSC Computing Committee (by notifying the Chair) at least 6 months in advance of the anticipated proposal deadline. An early draft of the proposal should be circulated to the LSC Computing Committee to allow sufficient time for comments and suggestions. The Computing Committee continuously evolves a plan prioritizing needs for LSC Tier 1 and Tier 2 level resources. The Committee will endorse proposals that are consistent with this plan.

LSC Tier 3 Sites

An LSC Tier 3 center is expected to support only the local institution. Funding requests for a Tier 3 center *should not state or suggest* that there will be broader support for the LSC community.

Therefore, a Tier 3 center would not normally be configured with more than a third of the minimum hardware and networking resources listed above for Tier 2 centers. If at any time a Tier 3 site's capacity is enhanced (through, e.g., an upgrade) to go beyond this nominal capacity, then the guidelines outlined above will become applicable.

LSC Tier 4 Resources

AN LSC Tier 4 resource represents a typical individual scientist's desktop. These computers may be clustered with the rest of a Tier 2 or Tier 3 center's resources, but they do not count toward the minimums or maximums in the Tier 2 and Tier 3 definitions.

At present, institutions maintaining Tier 3 or Tier 4 centers are not subject to any guidelines.