The Open Science Grid and BNL

Brookhaven National Laboratory (BNL) is a multi-disciplinary laboratory with major current responsibilities for coordinating computing in support of very large High Energy Physics (HEP) and Nuclear Physics (NP) projects. We are, consequently, very supportive of the establishment of a persistent production Grid capable of marshalling computing resources for these HENP projects. The projects with major BNL responsibility are the Relativistic Heavy Ion Collider (RHIC) experiments at BNL and the ATLAS detector at the CERN Large Hadron Collider (LHC). Both are prototypical of a new brand of very large, internationally distributed collaboration which brings with it access rights to major independently funded and managed distributed computing resources. For RHIC, BNL provides a large data analysis facility; for ATLAS, BNL is one of the five Tier-1 facilities in the world. While the immediacy of these HENP needs militates for very prompt Grid deployment, likely future needs in other areas of BNL multi-disciplinary activity, such as accelerator related computing, computational biology, lattice gauge computing, etc. requires a more deliberate approach. A premium must be put on a persistent production Grid which is properly constructed and so can serve as the foundation on which other large science projects can build as their Grid needs emerge over the next few years.

For US ATLAS, the immediate programmatic need is the support of Data Challenges 2 (starting in April 2004) & 3 (starting in October 2005), which involves three major functional aspects. First, we need to marshal distributed US ATLAS computing resources and federate them with ATLAS resources worldwide for the production of large volumes of Monte Carlo data. Second, we must afford access to data sets both on a demand basis and by maintaining replicas of specific data sets at locations selected to optimize availability. Third, we must empower individual users to effectively utilize distributed resources for physics analysis.

Within the RHIC project, the most immediate programmatic need is to replicate event summary data currently being produced by reconstruction farms at BNL to remote sites where substantial resources are available to support final analysis. While this activity is already being performed prototypically by STAR between BNL and LBNL using Grid middleware components, the need exists to more fully integrate this activity in a standardized way with site fabrics and to harden the process so as to improve performance and reduce the substantial human effort currently required to maintain the replication. The goal of this standardization is that STAR, PHENIX, other RHIC and non-RHIC experiments along with their respective remote analysis sites can efficiently make use of common solution. More generally, the RHIC experiments have an immediate need to effectively harvest computing resources in support of analysis from collaborating sites distributed around the world.

The operation of a persistent, production-quality Grid requires the definition of appropriate standardized interfaces for the various required services, the creation of middleware satisfying these standards and the adaptation of the fabrics of the participating sites to utilize that middleware. While much of the work that has been done (and remains to be done) is technical computer science work, specification of the needs and interfaces as well as the adaptation and deployment of middleware developed elsewhere will involve major site and scientific project-specific effort. BNL is specifically concerned about and involved in the following Grid issues, 1) cyber security, authentication, authorization and accounting, 2) defining, supplying and optimizing storage services, 3) Grid connected fabric monitoring, 4) distributed analysis, 5) furnishing support for Grid deployment at remote sites, and 6) Grid related wide area networking.

The scale of the Grid needs which must be satisfied for RHIC and ATLAS at BNL is expected to grow by at least an order magnitude every 3 years into the long-term future. In addition as the general trend of basic science toward large distributed collaborations continues, user communities from additional disciplines at BNL will surely find profit in integrating their resources into the Open Science Grid.