

Physics results from the STAR experiment at RHIC benefit from production Grid data services

The STAR experiment at Brookhaven National Laboratory, Long Island, New York, is one of four experiments at the Relativistic Heavy Ion Collider (RHIC) accelerator whose primary motivation is the observation and characterization of the Quark Gluon Plasma (QGP). The QGP is a form of matter consisting of a hot and dense soup of quarks and gluons thought to have existed briefly when the universe was about one microsecond old, just before a phase transition formed the protons and neutrons that make up atomic nuclei in the universe today. Data taking started in 2000 and by March 2004 a wealth of scientific results have come out including 26 publications in refereed journals.

STAR utilizes primarily two computing facilities, one at Brookhaven Lab, the RHIC Computing Facility (RCF), and another at Berkeley Lab, the National Energy Research Scientific Computing Facility (NERSC). As part of the Particle Physics Data Grid project, STAR has been developing its datagrid connection between these two facilities since 2001 so that it is now routine to have automated transfers of 1,000's of gigabytes of data (10,000's of files), and it's associated metadata (description of the data) over a day's time. This allows "next day" access to fresh data for analysis and physicists using the facilities at Berkley and Brookhaven are able to collaborate more effectively on the analyses that have led to the recent physics results.

This datagrid connection is implemented using Storage Resource Manager and Globus Toolkit software, developed with support from other DOE middleware projects, as well as the open source database software, MySQL. The data transfer procedures used before the current grid-based implementation were tedious, manpower intensive and error prone. It would take 10 days to transfer 1,000 gigabytes of data, not because of insufficient network bandwidth, but due to a lack of automation and fault tolerance. After transferring a dataset it was common to have an inconsistency in the files of 1% between the two sites, compromising the integrity of the data. Today this level of inconsistency is less than 0.02% (50 times improvement).

In addition to bulk data transfer STAR has interfaced its computational workload job submission tool, the STAR Unified Meta Scheduler (SUMS) with the grid and has been using it to run 1000's of simulation jobs at NERSC and automatically transport and archive the output at the RCF, making the results available to the collaboration for data analysis. This utilizes the Condor middleware from the University of Wisconsin as well as the Globus Toolkit.

URL's:

STAR – www.star.bnl.gov

PPDG – www.ppdg.net