

SLAC • FERMILAB DOE/Office of Science

PLUGGING INTO THE GRID



As the amount of BABAR data rapidly increases, the need to exploit the existing computing resources of geographically dispersed collaborating institutions becomes critical. The DOE funded Particle Physics Data Grid (PPDG) project is addressing the analyst's needs for efficient access to vast amounts of BABAR data located around the world.

THE GRID

What is a Grid?

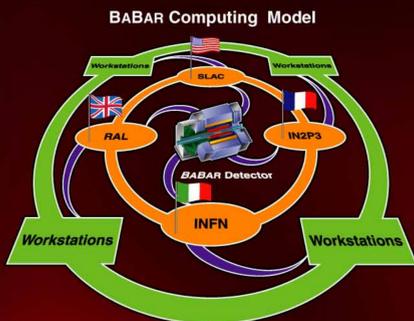
A Grid is a widely distributed set of resources capable of coordinated use.

Why do we need a Grid?

Grid Technologies are well suited to the problem of distributing multi-terabyte data samples to geographically disparate locations.

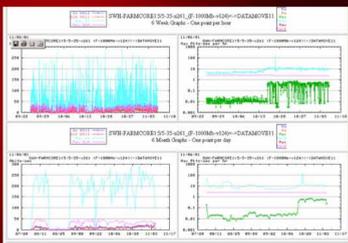
What does the Grid offer that we cannot do with existing technology?

Grid Technologies offer the promise to standardize the large-scale data sharing environment; thus making it easier to deploy storage resource intensive applications across a wide area network and multiple administrative domains.



The BABAR computing model uses a highly distributed pattern for data storage to fully exploit the available computing resources of the collaboration. Overlapping slices of data will be stored across the collaboration (eg. in the data centers at SLAC, INFN, RAL, IN2P3) to facilitate data access by collaborators working at their home institutions.

Transatlantic Data Transfer



Graphs of the history of data transfer (US to France).

SLAC AND THE PPDG

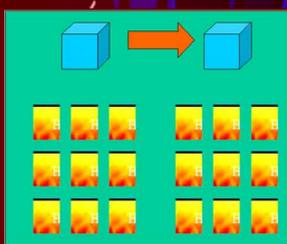
What is the PPDG?

The Particle Physics Data Grid is a DOE sponsored multi-lab/university project aimed at employing Grid technologies to devise vertically integrated solutions to the problem of data sharing problems apparent in High Energy and Nuclear Physics (HENP).

What is SLAC doing as part of the PPDG?

SLAC is currently undertaking a study to understand how Grid technologies can best be used to address the problem of inter- and intra-site BABAR data distribution. The experience and additional tools developed to use these technologies will be fed back to the PPDG hopefully benefiting future high volume HENP experiments. Two technologies that SLAC is currently looking at are the Storage Resource Broker (SRB) developed at SDSC and the GLOBUS toolkit developed at Argonne. The SRB is a vertical solution to the problem of replicating data on demand. The GLOBUS toolkit is a horizontal solution to using widely distributed computational and data resources within a common security infrastructure.

Grid Enabled Data Processing

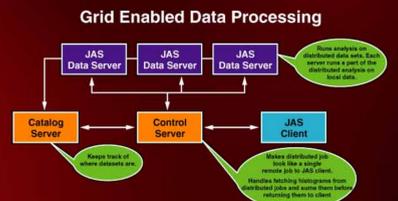


Moving data with Grid tools and analyzing data with JAS servers.

WORKING WITH DISTRIBUTED DATA

How do you analyze distributed data?

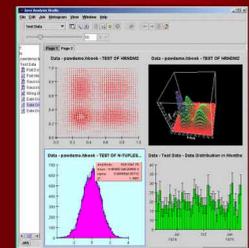
The demo below shows interactive analysis of data distributed over many machines. A prototype system is in use which allows analysis tasks to be transparently distributed across a set of machines, exploiting the aggregate CPU power and IO bandwidth of all the machines.



How Does this relate to the GRID?

In this demo Grid tools are used to produce multiple copies of data at different sites. In future the system could be closely integrated with the Grids authentication, resource location, and data cataloging services.

Interactive Analysis Tools



JAS Client: Interactive Analysis controller/viewer

What is JAS?

JAS is a general-purpose Java-based distributed data analysis system. It provides a GUI to allow scientists to take advantage of the graphics capabilities of their desktop machines, while accessing the IO and CPU power of remote data servers. JAS exploits the capability of Java to simply distribute data analysis tasks to a farm of possibly inhomogeneous machines.