

PARTICLE PHYSICS DATA GRID

Collaboratory Pilot



PI's: Richard Mount, Miron Livny, Harvey Newman

www.ppdg.net

GOALS

Enabling and enhancing data intensive experimental particle physics by "vertical integration" of Grid middleware components into HENP experiments' ongoing work.

Collaboration on specification, development and hardening of grid middleware by pragmatic development of common Grid services and standards - data replication, storage and job management, monitoring and planning.

TECHNICAL CHALLENGES

Make Grids usable and useful for the real problems facing international physics collaborations and for the average scientist in HENP.

Improve the robustness, reliability and maintainability of Grid software through early use in production application environments.

Address these challenges by:

- Integration, deployment of robust data replication
- Development, integration and deployment of intelligent job placement and scheduling
- Integration and deployment of storage resource management
- Specification, integration and deployment of global monitoring and information services
- Deployment of Grid technologies and services in production (24x7) environments with stressful performance needs.

OTHER CHALLENGES

Within PPDG - collaboration among numerous physics and CS groups (the social challenge)

Construct environment and workplan where a collection of effective binary/tertiary CS-Expt. partnerships lead to constructive development

External connections / coordination (the socio-political challenge)

- HENP Intergrid Coordination Board - representing PPDG, GriPhyN, iVDGL, EU-DataGrid, DataTag, CrossGrid, GridPP, INFN Grid, France, Germany, Japan,...
- SciDAC connections with DOEScienceGrid, SDM Center, ...
- Contacts identified and conversations underway

CURRENT STATE OF THE ART

Awkward but workable infrastructure solutions allow existing experiments to carry out today's data intensive computational challenges.

- Little or no re-use between centers
- Security issues

Manual (& labor intensive) data replication between centers.

Network is primary data transport mechanism

- Tapes stay in the robot and used only for massive storage capacity (Hurray!)

DEPENDENCIES ON OTHER SciDAC PROJECTS

High Performance Datagrid Toolkit, Security and Policy for Group Collaboration

- Relying on deployment in PPDG

SRM/SDM

- Relying on deployment in PPDG

DOE Science Grid - PKI Certificate Authority

- Relying on CA for subordinate PPDG RA
- Relying on international cooperation and coordinated development of policy and certificate acceptance

CONTACTS WITH SciDAC PROJECTS

Earth System Grid II - Arie Shoshani, LBNL

Collaboratory for Multi-Scale Chemical Science - Richard Mount, SLAC

National Fusion Collaboratory - Doug Olson, LBNL

DOE Science Grid - Doug Olson, LBNL

Pervasive Collaborative Computing Environment - Miron Livny, U. Wisc.

Reliable and Secure Group Communication - Miron Livny, U. Wisc.

A High-Performance Data Grid Toolkit - Ian Foster, ANL

Middleware Technology to Support Science Portals - Miron Livny, U. Wisc.

CoG Kits - Ian Foster /Mike Wilde, ANL

Scientific Annotation Middleware - Reagan Moore, SDSC

Storage Resource Management for Data Grid Applications - Arie Shoshani, LBNL

Middleware to Support Group to Group Collaboration - Harvey Newman, Caltech

Distributed Security Architectures - Doug Olson, LBNL

SIGNIFICANT RESULTS AND BREAKTHROUGHS

2002

- Deploy reliable file transfer, replica catalog
- Storage resource management of local resource
- Resource monitoring & status reporting
- Integrate CMS simulation production with Condor-G / Globus
- Job description language specification
- Deploy PKI Registration Authority under DOESG CA with policies accepted by international HENP community
- Test prototype VO authorization service
- Integration of file transfer, replica catalog and job control in some experiments' production services

SIGNIFICANT RESULTS AND BREAKTHROUGHS

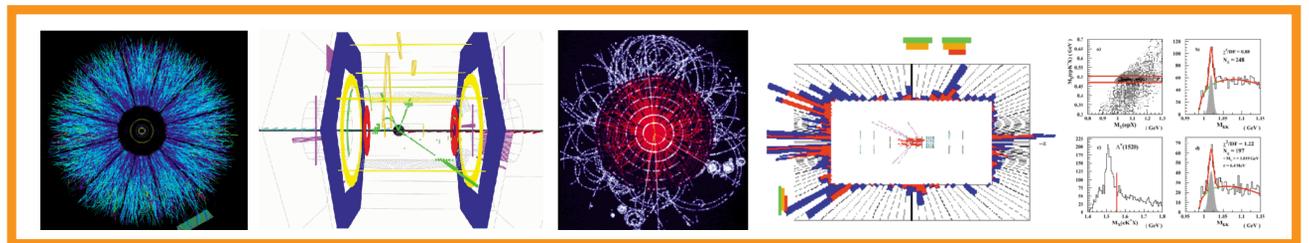
2003

- Deploy file replication services with reliable file transfer and cache management
- Storage resource discovery and prototype storage resource scheduling
- Integrate PKI RA with experiments' own registration & authorization functions
- Deploy VO authorization service
- Monitoring and status reporting of resource availability, quotas, priorities and cost estimation
- Remote job submission and production computing management
- Test and refine job description language
- Initial deployment of distributed data analysis and remote job submission for end-user physicists
- All experiments running some grid services in production

SIGNIFICANT RESULTS AND BREAKTHROUGHS

2004

- Deploy standardized job description language for experiments production computing integrated with:
 - replica management services
 - storage resource discovery & scheduling
 - monitoring and status reporting of jobs and resources
 - correlated storage & compute resource scheduling
- Experiment specific event & object collection replication in addition to file-based replication service
- Deployment of production distributed data analysis and remote job submission for end-user physicists
- Expanded set of common grid services in production for experiments



DESKTOP

CLUSTERS & CACHE DISK

HPCC & MSS

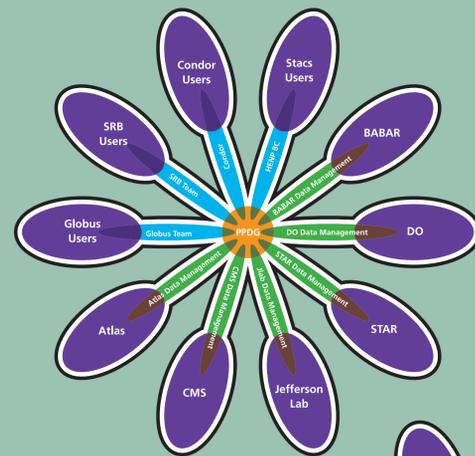
Grid Middleware & SciDAC Projects



DOE SCIENCE GRID



CHALLENGES OF PROJECT AND INTER-PROJECT COORDINATION



The PPDG Coordination Challenge

The HENP Grid-Project Coordination Challenge

