

**Particle Physics Data Grid
Collaboratory Pilot,
Grid Physics Network,
International Virtual Data Grid
Laboratory**

Integrating Datagrid Technology with Physics
Experiment End-to-end Applications.
ppdg-exec@ppdg.net



www.ppdg.net



www.griphyn.org



www.ivdgl.org

Members of the CMS experiment working in concert with PPDG, iVDGL, and GriPhyN have carried out the first production-quality simulated data generation on a data grid comprising sites at Caltech, Fermilab, the University of California-San Diego, the University of Florida, and the University of Wisconsin-Madison. This is a combination of efforts supported by DOE SciDAC, HENP, MICS, and the NSF as well as the EU funded EU-DataGrid project.

The deployed data grid serves as an integration framework where grid middleware components are brought together to form the basis for distributed CMS Monte Carlo Production (CMS-MOP) and used to produce data for the global CMS physics program. The middleware components include Condor-G, DAGMAN, GDMP, and the Globus Toolkit packaged together in the first release of the Virtual Data Toolkit.

The CMS-MOP distributed production system employs a tier-like hierarchy in which a production manager at a Tier-1 center distributes production jobs to several remote Tier-2 sites. Once generated at the Tier-2 sites, the simulated data is automatically published back to the Tier-1 center as well as replicated to selected Tier-2 sites.

This integration exercise showed that the Grid still presents significant challenges in harnessing distributed resources. Issues of data and security had to be overcome, e.g. how do you get your software &

data to many remote systems and know it's there? How do you get your results back? How do you allow people on the Grid to trust you? Issues of heterogeneity and error recovery had to be addressed, - to use other sites' resources, you need to interface with many batch systems; the Grid means more errors, more crashes, more mysterious failures. Unanticipated errors were handled such as: Key machines crashing in the middle of a run; Grid credentials expiring in the middle of a run; Jobs successfully completing but their results being lost before they got sent back; Various pieces of middleware did the unexpected; The network would go down.

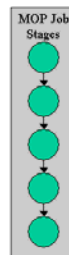
Despite these challenges, over 50,000 proton-proton collision events inside the CMS detector have been simulated using CMS-MOP and validated for use by CMS physicists. Production of another 150,000 simulated events is underway.

This achievement represents an extremely challenging and important milestone in the integration of grid middleware components within the current "real world" LHC computing environment!

http://grid.fnal.gov/test_beds/USCMSTestBed.htm



MOP Job Stages



- > Stage-in - get the program and its data to a remote site
- > Run - run the job at the remote site
- > Stage-back - get the program logs back from the remote site
- > Publish - advertise the results so they will be sent to sites that want it
- > Cleanup - clean up remote site

