



Contribution ID: 227

Type: oral presentation

DZero Data-Intensive Computing on the Open Science Grid

Monday, September 3, 2007 2:40 PM (20 minutes)

High energy physics experiments periodically reprocess data, in order to take advantage of improved understanding of the detector and the data processing code. Between February and May 2007, the DZero experiment will reprocess a substantial fraction of its dataset. This consists of half a billion events, corresponding to more than 100 TB of data, organized in 300,000 files.

The activity utilizes resources from sites around the world, including a dozen sites participating to the Open Science Grid consortium (OSG). About 1,500 jobs are run every day across the OSG, consuming and producing hundreds of Gigabytes of data. OSG computing and storage resources are coordinated by the SAM-Grid system. This system organizes job access to a complex topology of data queues and job scheduling to clusters, using a SAM-Grid to OSG job forwarding infrastructure.

For the first time in the lifetime of the experiment, a data intensive production activity is managed on a general purpose grid, such as OSG. This paper describes the implications of using OSG, where all resources are granted following an opportunistic model, the challenges of operating a data intensive activity over such large computing infrastructure, and the lesson learned throughout the few months of the project.

Primary author: Dr GARZOGGIO, Gabriele (FERMI NATIONAL ACCELERATOR LABORATORY)

Co-authors: BARANOVSKI, Andreii (FERMI NATIONAL ACCELERATOR LABORATORY); Dr ABBOT, Brad (University of Oklahoma); Dr DIESBURG, Mike (FERMI NATIONAL ACCELERATOR LABORATORY); MHASHILKAR, Parag (FERMI NATIONAL ACCELERATOR LABORATORY); Dr KURCA, Tibor (IN2P3)

Presenter: Dr BOEHNLEIN, Amber (FERMI NATIONAL ACCELERATOR LABORATORY)

Session Classification: Grid middleware and tools

Track Classification: Grid middleware and tools