

CDF Distributed Analysis

The CDF experiment at Fermilab is undergoing an aggressive Trigger and DAQ upgrade which will increase the maximum event rate out of the detector from 80Hz to 360Hz by 2006. The rationale for this increase is to be able to maximize physics output at any luminosity by dynamically filling the bandwidth with low threshold triggers.

The most important physics measurement that benefits from the dynamically filled bandwidth is the measurement of Bs mixing. However, a broad range of other physics topics will also benefit. The increased data volume translates into a doubling of computing needs which had previously not been budgeted for. To satisfy those needs, CDF has embarked on an ambitious global computing program with the goal of having 50% of all of CDF's computing outside of FNAL by Summer 2005.

The CDF situation is somewhat unique among currently running experiments in particle physics in that its computing need is clearly dominated by user analysis computing. Since January 2004 we have put into production 9 computing centers across the world, all of which are accessible by all of the close to 800 CDF physicists registered at the central system at FNAL. No user logins are required. Instead users interact with the distributed system from their desktop as if their jobs were running locally. Authentication, authorization, and accounting are all based on user credentials, and no user accounts need to be maintained at any of the sites.

For the 2004 summer conference season resources outside of FNAL accounted for roughly 25% of CDF's global resource consumption, and six of the nine centers are located outside the US.

The software infrastructure in CDF today is based on the CDF Analysis Farm (CAF) and SAM projects. CAF provides job management, user interfaces, and monitoring & accounting with SAM providing data management. In addition, there are efforts (FNAL & JHU) underway to deploy a squid based distributed database caching system for calibration metadata. The CAF effort is a UCSD/INFN collaboration while SAM in CDF receives effort from FNAL, the UK, Germany, Rutgers, and INFN. Operations support is co-ordinated by TTU.

PPDG has contributed to CDF via Condor, Globus, and SAM, and efforts are under way to understand both technical as well as support issues with regard to adding storage elements with SRM interfaces at each of the sites. At present, the main FNAL as well as the UCSD site are based on Condor, and the remaining sites are expected to transition to Condor within the next 6 months. Condor effort has been provided in three areas so far: scalability & hardening, computing on demand, and security & encryption. In addition, work on policy implementation and "pull model" for job management is ongoing. CDF benefits from Globus in the sense that all data movement between sites is based on gridFTP. CDF benefits from SAM in that all data access outside of FNAL is based on SAM.

Future directions are focused in the US on SAMGrid within the context of the Open Science Grid. However, as 50% of future CDF hardware funding presently appears to be located outside the US, interoperability between and federation of grids will be a primary concern for CDF in the future.

URL's:

CDF – www-cdf.fnal.gov

PPDG – www.ppdg.net