Contribution ID: **111** Type: **Oral**

CRAB, the CMS tool to allow data analysis in a distributed environment

Wednesday, February 13, 2008 12:00 PM (20 minutes)

The CMS experiment will produce few PBytes of data each year to distribute and store in many computing centres spread in the countries participating to the CMS collaboration and made available for analysis to world-wide distributed physicists.CMS will use a distributed architecture based on Grid infrastructure to analyze data stored at remote sites, to assure data access only to authorized users and to ensure remote resources availability.Data analysis in a distributed environment is a task that assume to know which data are available, where data are stored and how to access them.To simplify analysis job creation and management the CMS collaboration is developing CRAB (CMS Remote Analysis Builder) a tool to allow users with no specific Grid knowledge to be able to run analysis in the distributed environment as data were in their local farm. CRAB is developed as tool standalone and client-server to improve the throughput, the scalability and to automatize most of CRAB functionalities

3. Impact

Users have to provide CRAB with the name of the dataset to analyze and the total number of events, their analysis configuration file and libraries.They must belong to the CMS Virtual Organization and have a valid Grid certificate.CRAB creates a wrapper of the analysis executable including CMS environment setup and output management.CRAB finds data location querying specific CMS catalog and splits the number of events in jobs according with data block distribution.CRAB packs the user code and send it to remote resources together with the wrapper.The job submission is done using Grid workload management commands.The resources availability, status monitoring and output retrieval of submitted jobs are fully handled by CRAB.For job submission CRAB is interfaced with gLite WMS and with OSG, based on condor_g.CRAB uses the voms-proxy server to create the user proxy certificate and its delegation.CRAB uses the LB Api to check the status of jobs and the UI command to manage jobs.

4. Conclusions / Future plans

During the last year the number of users and jobs submitted via CRAB increased.This result shows that CRAB is useful to run analysis in Grid environment and the development of server-client architecture is needed to guarantee scalability.Our experience using CRAB shows some weakness of some Grid services as WMS constrains, problem with sandboxes dimension, problem with the protocol for copy the produced output to remote mass storage.Remote sites need continuous checks to guarantee availability

Provide a set of generic keywords that define your contribution (e.g. Data Management, Workflows, High Energy Physics)

CMS distributed data analysis, workload management, high energy physics, Grid

1. Short overview

The CMS collaboration is developing a tool to allow physicists to access and analyze data stored in geographically distributed sites, simplifying the data discovery and hiding details related analysis job creation, execution and monitoring in the Grid environment. With this presentation we would like to show the progress of our work and some statistics about its usage.

Primary authors: VAANDERING, Eric (FNAL); FANZAGO, Federica (CERN-CNAF); FANFANI, alessandra (university bologna); KAVKA, carlos (INFN-TS); SPIGA, daniele (INFN-PG); FARINA, fabio (Univesity Milano Bic-

occa); CODISPOTI, giuseppe (university bologna); CORVO, marco (CERN-CNAF); CINQUILLI, mattia (INFN-Perugia); LACAPRARA, stefano (INFN-LNL)

Presenter: FANZAGO, Federica (CERN-CNAF)

Session Classification: Grid Access

Track Classification: Application Porting and Deployment