



CRAB: a tool for CMS distributed analysis in grid environment.

Wednesday 1 March 2006 18:00 (30 minutes)

The CMS experiment will produce a large amount of data (few PBytes each year) that will be distributed and stored 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 complex computing task, that assume to know which data are available, where data are stored and how to access them.

The CMS collaboration is developing a user friendly tool, CRAB (Cms Remote Analysis Builder), whose aim is to simplify the work of final users to create and to submit analysis jobs into the grid environment. Its purpose is to allow generic users, without specific knowledge of grid infrastructure, to access and analyze remote data as easily as in a local environment, hiding the complexity of distributed computational services.

Users have to develop their analysis code in an interactive environment and decide which data to analyze, providing to CRAB data parameters (keywords to select data and total number of events) and how to manage produced output (return file to UI or store into remote storage).

CRAB creates a wrapper of the analysis executable which will be run on remote resources, including CMS environment setup and output management. CRAB splits the analysis into a number of jobs according to user provided information about number of events. The job submission is done using grid workload management command.

The user executable is sent to remote resource via `inputsandbox`, together with the job. Data discovery, resources availability, status monitoring and output retrieval of submitted jobs are fully handled by CRAB.

The tool is written in python and have to be installed to the User Interface, the user access point to the grid.

Up to now CRAB is installed in ~45 UI and about ~210 different kind of data are available in ~40 remote sites.

The weekly rate of submitted jobs is ~10000 with a success rate about 75%, that means jobs arrive to remote sites and produce outputs, while the remnant 25% aborts due to site setup problem or grid services failure.

In this report we will explain how CRAB is interfaced with other CMS/grid services and will report the daily user's experience with this tool analyzing simulated data needed to prepare the Physics Technical Design Report.

Summary

Report about CRAB, a tool for CMS analysis in grid environment: how it is interfaced with CMS/grid services and user's experience.

Authors: FANFANI, Alessandra (Bologna University); SPIGA, Daniele (Perugia University); FANZAGO, Federica (INFN-PADOVA); CORVO, Marco (CERN/CNAF); LACAPRARA, Stefano (INFN Legnaro)

Co-authors: CIRAOLO, Giovanni (Firenze University); DE FILIPPIS, Nicola (Bari University); SMIRNOV, Nikolai (INFN-PADOVA)

Presenter: FANZAGO, Federica (INFN-PADOVA)

Session Classification: 1b: Astrophysics/Astroparticle physics - Fusion - High-Energy physics

Track Classification: Astroparticle physics - Fusion - High-Energy physics