



Contribution ID: 26

Type: **Workshop**

## **An overview of CMS Workload Management for data analysis.**

*Wednesday, 13 April 2011 11:00 (30 minutes)*

### **Overview**

The Compact Muon Solenoid (CMS) is one of the four main experiments (VOs) at LHC and relies heavily on grid computing. The CMS computing model defines how the data is to be distributed and accessed to enable physicists to efficiently run their analyses over the data. The CMS Remote Analysis Builder (CRAB) is the specific Workload Management system, that allows the end user to transparently access data in heterogeneous pool of resources distributed across several continents.

### **Impact**

CRAB is currently starting the re-writing process with the aim to better cope with the foreseen growing amount of data and user community. This process is providing an extremely important opportunity for developers to share a set of functionality already implemented by other Workload Management Systems (WMS). These latter are the systems which have been developed by other LHC Experiments to perform Grid submissions, such as Panda, Dirac, and Alien and GANGA.

### **Description of the work**

CRAB adopt a client server architecture implemented in python. The client, which provide the user with a batch-like command line application, has the responsibility of generating the user's proxy, packaging up the pre-compiled user code, and submitting the job package to the CRAB server. The intermediate Analysis Server has the responsibility of submitting the jobs to the chosen Grid middleware, resubmitting jobs as needed, and caching the users' output. CRAB design allow to transparently interacts with different Grids and batch system. CRAB has been in production and in routine use by end users since Spring 2004. It has been extensively used between the initial definition of the CMS Computing Model in 2004 and the start of high energy collisions in 2010, attending numerous scaling tests and service challenges. CMS is currently in the process of replacing its workload management system. This next generation tool is named WMCORE, a Python based workflow library which is used for so called task lifecycle management. The workflow library is the result of a convergence of three subprojects that respectively deal with scientific analysis, simulation and real time data aggregation from the experiment. CRAB will be part of this migration.

### **Conclusions**

The CRAB Workload Management system is widely used for accessing CMS distributed data by end users. During the first year of the data taking the system coped well with the physics needs. In order to reduce the operations load and improve the scalability CRAB is going to be re-written. This process has also the aim to

share common functionality with other existing WMS. The system is being developed in a very close collaboration with the developers of other Workload Management system as well as the Experiment Dashboard team.

**Primary author:** Dr SPIGA, Daniele (CERN)

**Co-authors:** VAANDERING, Eric (FNAL); FARINA, Fabio (INFN of MilanoBicocca); FANZAGO, Federica (INFN of Padova); RIAHI, Hassen (University&INFN of Perugia); CINQUILLI, Mattia (CERN)

**Presenter:** Dr SPIGA, Daniele (CERN)

**Session Classification:** User Environments

**Track Classification:** User Environments - Applications