



Contribution ID: 31

Type: **not specified**

CRAB: a user friendly application for distributed data processing for the Compact Muon Solenoid experiment at the LHC

Tuesday, 27 March 2012 11:00 (30 minutes)

Description of the Work

Taking the experience gained from previous CRAB versions, developers plan to release a new version of the tool which aims to improve the sustainability of the service besides solving known issues and bottlenecks. CRAB will be centrally deployed as an online service exposing a Representational State Transfer (REST) interface. Services offered by the server will be accessible by the end user through a lightweight client, which will send requests to the server REST interface. The server is composed by a multi-tiered architecture where each tier takes care of performing specific functions in the chain. The WorkQueue tier takes care of providing a central queue for all the user requests, and manages the priorities between users/requests themselves. Interactions with the underlying Grid layer are handled by the so called Agent tier. The Agent pulls user requests from the Workqueue, it splits them in several jobs, and submit jobs to the Grid. Finally, the tier called AsyncStageOut handles the output produced by user's jobs.

By using CRAB a user can abstract from the technical details of the Grid infrastructure, and just focus on his primary activity: the analysis of the data collected by the Large Hadron Collider. Features like automatic resubmission of failed jobs and automatic handling of Grid computational and storage resources, considerably simplify the user's work. From the maintainance point of view, the new implementation aims at reducing of the sustainability cost. In fact, the tool has been rewritten on top of a commonly developed library (named WMCore), which is also used for other use cases in CMS. In the paper, new features of CRAB will also be described.

Conclusions

At the time of writing the new version of CRAB is on the process of consolidating the basic functionalities. It is close to enter the commissioning phase after which CMS will start the transition to the new version. We present the status of the project and the achieved experience during the integration period.

Impact

CMS will be producing scientific result for at least a quarter of a century. Based on the experience of these first years of data taking, the experiment have to produce a model that makes CMS Computing sustainable in the future. Reliability, usability and scalability of the analysis system can represent a crucial aspect for the success of the whole experiment. The reduction of the human effort needed for the analysis operations represents a key aspect of a sustainable model. The commissioning of the new version of CRAB is extremely important to start the deprecation process of the previous version. A global usage of the new CRAB represents a step towards the sustainability of the CMS Computing.

Overview (For the conference guide)

The CMS Remote Analysis Builder (CRAB) application addresses the needs of the CMS community, allowing the users to easily access the Grid resources. CRAB interacts with the local user environment, the Data Management services and with the Grid middleware, limiting the knowledge of the technical details required of the end user. CRAB has progressed from a limited initial prototype nearly 5 years ago, to a system heavily employed by the whole CMS collaboration to prepare over 100 analysis papers. CMS observes more than 400 unique users submitting CRAB jobs per week, with close to 1000 individuals per month. Up to 200,000 CRAB jobs per day run on the Grid.

The CRAB team has an ambitious program planned in 2012: to release a new generation of CRAB that aims to make a step towards a SaaS architecture. This work will present the joint CMS experiment and CERN IT-ES effort to realize such project, highlighting the impact on the service maintenance and first experiences dealing with beta users.

Primary author: Dr SPIGA, Daniele (CERN)

Co-authors: Dr VAANDERING, Eric (FNAL); Mr RIAHI, Hassen (INFN); Mr CINQUILLI, Mattia (CERN); Dr MASCHERONI, marco (CERN)

Presenter: Dr SPIGA, Daniele (CERN)

Session Classification: Community-tailored Services

Track Classification: Software services for users and communities