Contribution ID: **88** Type: **poster**

Data Analysis through the DIANA Meta-Scheduling Approach

Wednesday, September 5, 2007 8:00 AM (20 minutes)

We introduce the concept, design and deployment of the DIANA meta-scheduling approach to solving the challenge of the data analysis being faced by the CERN experiments. The DIANA meta-scheduler supports data intensive bulk scheduling, is network aware and follows a policy centric meta-scheduling that will be explained in some detail.

In this paper, we describe a Physics analysis case study using the DIANA meta-scheduler. We demonstrate that a decentralized and dynamic meta-scheduling approach is an effective strategy to cope with the increasing numbers of users, jobs and datasets. We present "quality of service" related statistics for the physics analysis through the application of a policy centric fair-share scheduling model. The DIANA meta-schedulers create a peer-to-peer hierarchy of schedulers to accomplish resource management. They employ scheduling approaches that are able to change with evolving loads and adapt following the dynamic and volatile nature of the resources. The DIANA meta-scheduler also acknowledges the important role of networks in the modern day distributed systems and treats them an equally important resource in the scheduling decisions along with the compute and data resources.

This topic formed the contents of a Ph.D. thesis done partly at CERN and partly at the University of the West of England in the United Kingdom. The Physics analysis case study was based on the computing model that was presented in the CMS Technical Design Report. Our results show that Physics analysis scheduling can be made robust by employing fault tolerant and decentralized meta-scheduling.

Primary author: Prof. MCCLATCHEY, Richard (UWE)

Co-authors: Dr ANJUM, Ashiq (UWE); Dr WILLERS, Ian (CERN)

Presenter: Prof. MCCLATCHEY, Richard (UWE)

Session Classification: Poster 2

Track Classification: Distributed data analysis and information management