Contribution ID: **275** Type: **oral presentation**

DIANA Scheduler

Wednesday, 15 February 2006 14:20 (20 minutes)

Results from and progress on the development of a Data Intensive and Network Aware (DIANA) Scheduling engine primarily for data intensive sciences such as physics analysis is described. Scientific analysis tasks can involve thousands of computing, data handling, and network resources and the size of the input and output files and the amount of overall storage space allotted to a user necessarily has significant bearing on the scheduling of data intensive applications. If the input or output files must be retrieved from a remote location, then the time required transferring the files must be taken into consideration when scheduling compute resources for the given application. The central problem in this study is the coordinated management of computation and data at multiple locations and not simply data movement. However, this can be a very costly operation and efficient scheduling can be a challenge if compute and data resources are mapped without network cost. This can result in performance degradation if the advantage of recent advances in networking technologies and bandwidth abundance based on optical backbones is not delegated to a scheduling engine. To incorporate these features, we have implemented an adaptive algorithm within the DIANA Scheduler which takes into account data location and size, network performance and computation capability to make efficient global scheduling decisions. DIANA is a performance-aware as well as an economy-guided Meta Scheduler. It iteratively allocates each job to the site that is likely to produce the best performance as well as optimizing the global queue for any remaining pending jobs. Therefore it is equally suitable whether a single job is being submitted or bulk scheduling is being performed. Results suggest that considerable performance improvements are to be gained by adopting the DIANA approach and this makes it a very suitable Meta Scheduler for Physics Analysis.

Primary author: Mr ANJUM, Ashiq (University of the West of England)

Co-authors: Prof. ALI, Arshad (NUST); Prof. WILLERS, Ian (CERN); Prof. MCCLATCHEY, Richard (University of the West of England)

Presenter: Mr ANJUM, Ashiq (University of the West of England)

Session Classification: Distributed Data Analysis

Track Classification: Distributed Data Analysis