Contribution ID: **39**   Type: **oral presentation**

DIAL: Distributed Interactive Analysis of Large Datasets

Monday 13 February 2006 14:00 (20 minutes)

DIAL is a generic framework for distributed analysis. The heart of the system is a scheduler (also called analysis service) that receives high-level processing requests expressed in terms of an input dataset and a transformation to act on that dataset. The scheduler splits the dataset, applies the transformation to each subdataset to produce a new subdataset, and then merges these to produce the overall output dataset which is made available to the caller. DIAL defines a job interface that makes it possible for schedulers to connect with a wide range of batch and grid workload management systems. It also provides command line, root, python and web clients for job submission that enable users to submit and monitor jobs in a uniform manner. Scaling to very large jobs can be handled with a scheduler that does partial splitting and submits each subjob to another scheduler. I will give the current status of DIAL and discuss its use in the context of the ATLAS experiment at the CERN LHC (Large Hadron Collider). There we are looking at submission to local batch systems, globus gatekeepers, EGEE/LCG workload management, ATLAS production, and PANDA. The latter is a U.S. ATLAS framework for data production and distributed analysis (thus the name) that may also use DIAL for its internal scheduling.

Primary author: ADAMS, David (BNL)

Co-authors: TAN, Chun Lik (University of Birmingham); HARRISON, Karl (University of Cambridge)

Presenter: ADAMS, David (BNL)

Session Classification: Distributed Data Analysis

Track Classification: Distributed Data Analysis