

Workflow generator and tracking at the rescue of distributed processing. Automating the handling of STAR's Grid production.

Monday 23 March 2009 16:50 (20 minutes)

Processing datasets on the order of tens of terabytes is an onerous task, faced by production coordinators everywhere. Users solicit data productions and, especially for simulation data, the vast amount of parameters (and sometime incomplete requests) point at the need for a tracking, control and archiving all requests made so a coordinated handling could be made by the production team.

With the advent of grid computing the parallel processing power has increased but traceability has also become an increasing problematic due to the heterogeneous nature of Grids. Any one of a number of components may fail invalidating the job or execution flow in various stages of completion and re-submission of a few of the multitude of jobs (keeping the entire dataset production consistency) a difficult and tedious process. From the definition of the workflow to its execution, there is a strong need for validation, tracking, monitoring and reporting of problems.

To ease the process of requesting production workflow, STAR has implemented several components addressing the full workflow consistency. A Web based online submission request module, implemented using Drupal's Content Management System API, enforces ahead that all parameters are described in advance in a uniform fashion. Upon submission, all jobs are independently tracked and (sometime experiment-specific) discrepancies are detected and recorded providing detailed information on where/how/when the job failed. Aggregate information on success and failure are also provided in near real-time. We will describe this system in full.

Presentation type (oral | poster)

oral

Authors: Dr LAURET, Jerome (BROOKHAVEN NATIONAL LABORATORY); Mr HAJDU, Levente (BROOKHAVEN NATIONAL LABORATORY); Dr DIDENKO, Lidia (BROOKHAVEN NATIONAL LABORATORY)

Presenter: Mr HAJDU, Levente (BROOKHAVEN NATIONAL LABORATORY)

Session Classification: Distributed Processing and Analysis

Track Classification: Distributed Processing and Analysis