Contribution ID: **9** Type: **oral presentation**

EasyGrid: a job submission system for distributed analysis using grid

Thursday, 10 May 2007 14:30 (15 minutes)

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).

We automate job submission system to grid farms with EasyGrid, an intermediate layer between grid middleware and user software that provides
functionality to perform data and functional parallelism. Users without grid skills
were able to use grid. The case studies were hadronic tau decays distributed
analysis and neutral pion discrimination using genetic programming algorithms.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.

Data gridification: The first benchmark was eta(540) reconstruction to test what
is the best approach to data distribution. The second benchmark was tau
decays to neutral pions. This benchmark selected events over 482 million real
events and generated 5 million MC events. The third benchmark was search for
anti-deuterons in all events available in BaBar Run 3 (1,500 million events in
one week using 250 computers in parallel).
Functional gridification: Genetic programming was used to discriminate
reconstruction of real neutral pions from background evolving a mathematical
model with maps the variables hyperspace to a real value, an algebraic function
of pions kinematics variables. Applying the discriminator to a given pair of
gammas, if the discriminate value is bigger than zero, the pair of gammas is
deemed to come from pion decay. Otherwise, the pair is deemed to come from
another (background) source.
More information see http://www.hep.man.ac.uk/u/jamwer/

With a forward look to future evolution, discuss the issues you have encountered (or that you expect) in using the EGEE infrastructure. Wherever possi-

ble, point out the experience limitations (both in terms of existing services or missing functionality)

Easygrid has performed its tasks submitting, recovering results, and providing
listings for further analysis when something goes wrong. However, data
services (store, recover and transfer data) still a bottleneck. Most worker nodes
were running distributed analysis (an IO bond application) with 50% IOWAIT,
and consequently 50% CPULOAD. This is a severe efficiency problem that needs
to be tackled before CERN distributed analysis production starts next year.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications

The added value was a user's transparent framework for reliable data
gridification (support execution and results'recovery of many copies of the
same binary code running independently and at same time in many computers
with different data files) and functional gridification (one binary code running
distributed in many grid computers at same time).
Data gridification can be used to run Monte Carlo Events generation, raw data
analysis, any Root application, or any other generic software.
Functional parallelism is done through a library with several functions to run
conventional software on the grid with minor changes in the source code. It
provides an efficient and secure communication mechanism to allow data
transfer between jobs in different worker nodes. If any node goes down, the
master program re-submits the task to another server. For more information
see http://www.geocities.com/jamwer2002/gridgeral.pdf

Author: Dr WERNER, James (University of Manchester)

Presenter: Dr WERNER, James (University of Manchester)

Session Classification: Workflow

Track Classification: Workflow