

ThIS on the Grid

Wednesday, February 13, 2008 12:15 PM (15 minutes)

Our main requirements concern computing resources and data management.

The simulation is split into sub-jobs. Each sub-job uses a different random seed number, allowing to be statistically independent and to be run concurrently. By dividing one simulation into hundreds of different sub-jobs, computation time can be reduced from more than one day to less than 1 hour if computing resources on the grid are rapidly available.

Data requirements are also important. ThIS needs about 50Mo of input data and can produce between 15 to 150Mo of output data per each sub-job.

ThIS is based on the Geant4 toolkit and consequently uses Geant4 and CLHEP libraries. In order to keep our application independent on the installed software on the grid, we decided to compile it statically. Thus, each time a complete simulation is run, we provide the executable and all the needed input files as a tar ball stored on a storage element (SE) of the grid. Once completed, output data are retrieved and merge on a SE.

3. Impact

Porting ThIS on the grid is making use of a large number of grid services, from basic ones like the file catalogue to more evolved services proposed by the Workload Management System (WMS). Among the latter, we can cite the submission of parametric jobs, as well as the possibility to specify input data stored on the SE of the grid.

A parametric job causes a set of very similar jobs to be generated from one JDL file. This is exactly the case for our multiple sub-jobs. They are all the same except for the random sequence that must be different from one sub-job to another and a few other parameters.

Our application needs large input data files that cannot be passed in the input sandbox of a job. Therefore, we exploited the possibility to specify input data stored on the storage elements of the grid. Moreover, this functionality ensures that the WMS will schedule the job to a computing element close to one of the storage elements where the data is.

URL for further information:

<http://www.creatis.insa-lyon.fr/rio/ThIS/>

4. Conclusions / Future plans

The process of porting ThIS on the EGEE Grid is currently in progress. Our first results show that the grid can bring an important amelioration in computation time. However, work still needs to be done in order to cope with delayed and failed jobs among the jobs belonging to the same simulation. As a second stage in the porting and deployment of ThIS we consider implementing a web-based grid portal that would make ThIS available for physicians and researchers who could benefit from it.

Provide a set of generic keywords that define your contribution (e.g. Data Management, Workflows, High Energy Physics)

Therapeutic Irradiation Simulator, grid, application porting, parametric jobs, data management

1. Short overview

We present ongoing developments and first results on porting ThIS (a Therapeutic Irradiation Simulator) on the EGEE Grid. ThIS is a Geant4 based software dedicated to the Monte-Carlo simulation of irradiations of living tissues with photons, protons or light ions beams for cancer therapy. The large number (~100000000)

of simulated particles needed for only one simulation requires a very high computation time that can be considerably diminished if the application runs on the grid.

Primary authors: Mr SARRUT, David (CNRS - CREATIS LRMN, Léon Bérard Cancer Center); Mr BENOIT-CATTIN, Hugues (CNRS - CREATIS LRMN); Mr GUIGUES, Laurent (CNRS - CREATIS LRMN); Mrs CAMARASU, Sorina (CNRS - CREATIS LRMN)

Presenter: Mrs CAMARASU, Sorina (CNRS - CREATIS LRMN)

Session Classification: Life Sciences

Track Classification: Application Porting and Deployment