



Enabling Grids for E-science

Creatis
LRMN

THIS on the Grid

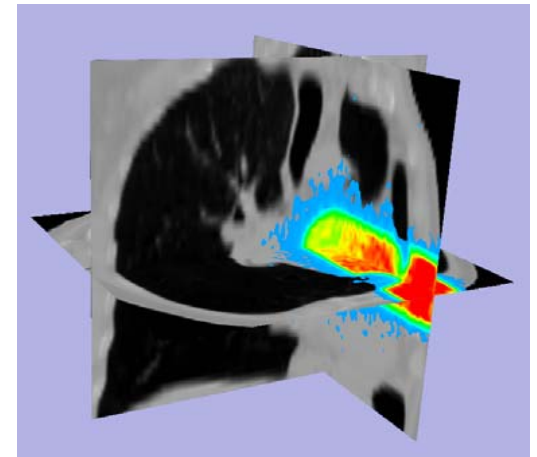
Sorina CAMARASU
3rd EGEE User Forum
Clermont-Ferrand, 13 Feb 2008



- **What is THIS ?**
- **This on the Grid – The Gridification Process**
 - Why Bring THIS on the Grid?
 - Implementation
 - Results
- **Perspectives**
- **Conclusion**

- **A Therapeutic Irradiation Simulator**

- Cancer treatment by irradiation of patient with beams of photons, protons or carbons
- Simulation of the interaction between particles and matter
 - Matter \leftrightarrow patient tumor
 - Particles \leftrightarrow photons, carbon, protons
 - Interaction \leftrightarrow dose distribution near tumor
- Based on Geant4
- Developed by: David Sarrut and Laurent Guigues
- <https://www.creatis.insa-lyon.fr/rio/THIS>



- **Aims**

- Offer an open platform to researchers for Monte Carlo simulations optimization
- Offer a fast and reliable simulation tool for researchers in medical physics (treatment planning) and medical imaging for treatment control
- Produce a reference dataset (energy deposit, positron emitters distributions, ...) for non-conventional therapies (hadrontherapy).



Enabling Grids for E-scienceE

Creatis
LRMN

This on the Grid The Gridification Process

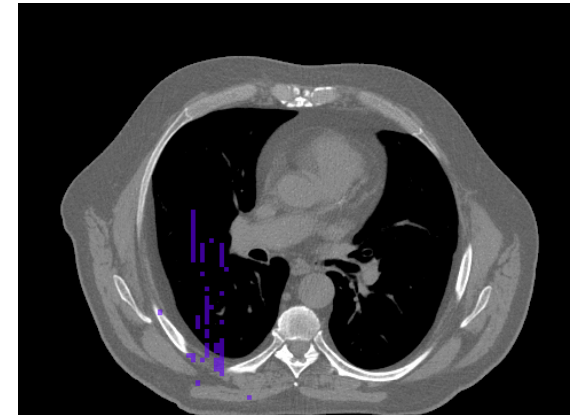


- **Intensive computing**

- Monte Carlo simulation (1 to 900 h cpu)
- Stochastic simulation algorithm
- 3D image pre-processing: segmentation

- **Data management**

- Important input and output data
- Input: patient images (20 to 50 Mb in 3D and up to 600 Mb in 4D), configuration and script files
- Output: the images with the dose distribution (up to 50 Mb) and/or the phase space (up to 1G)



3D dose distribution, 700h CPU
CT image (482x360x141)

- **Geant4, CLHEP and shared libraries**
 - Static linking and build
- **Job submission**
 - Split the simulation into independent sub-jobs
 - e.g. a 50 M particles simulation is divided into 50 sub-jobs, each with 1 M particles to simulate
- **Parametric jobs**
 - Very similar jobs generated from one JDL file – exist with glite-wms
 - These sub-jobs are parametric jobs
 - However, the current implementation does not use them
 - Problems with status and cancelation
- **The JJS Tool**
 - Developed at IN2P3 by Pascal Calvat
 - Java Job Submission – a Java application for job submission and management

- **Data management**

- The executable and G4EMLOW libraries are distributed on SE
- Configuration and input specific files are provided at submission
- Output files/results are copied from the WN on the IN2P3 SRM and then on the user's machine

- **A jdl and script example**

```
srmcp srm://dpmsrm.ciemat.es/dpm/ciemat.es/home/biomed/2008-02-07/filedbbae9d7-a3eb-4b64-da5e1420 file:///`pwd`/this_on_egee.tar.gz
tar -zxf this_on_egee.tar.gz
cd this_on_egee
srmcp srm://ccsrm02.in2p3.fr/pnfs/in2p3.fr/data/biomed/2008-02-07/filec2b4a653-7347-4a96-51628016 file:///`pwd`/G4EMLOW3.0.tar.gz
tar -zxf G4EMLOW3.0.tar.gz
export G4LEDATA=`pwd`/G4EMLOW3.0
cd Thorax/
globus-url-copy gsiftp://cclcgse101.in2p3.fr/var/jjs/dteam/biomed/archive.tar.gz file:///`pwd`/input.tar.gz
tar -zxf input.tar.gz
./this params
tar -czf output000$1.tar.gz output/*
globus-url-copy file:///`pwd`/output000$1.tar.gz gsiftp://cclcgse101.in2p3.fr/var/jjs/biomed/5NPVoutput000$1.tar.gz
cd ../..
rm this_on_egee.tar.gz
rm -r this_on_egee
```

- **Execution and waiting time vary significantly**
 - Execution type for a same type of job may vary depending on WN (computing power) from 1 to 4 hours
 - For 50 jobs submitted simultaneously, often at least one waiting job one hour after submission
 - A certain number of aborted jobs
- **Resubmission**
 - Resubmission must be taken into account
 - JJS integrates a job submission manager that resubmits aborted jobs or jobs that have been waiting for too long
 - The new WMS with glite-wms also integrates a ShallowRetryCount JDL attribute
- **THIS has already been ported on the IN2P3 cluster**
 - For few simulations, results are retrieved faster
 - For many simulations, there may be a scalability problem



- **A ‘probabilistic’ implementation**
 - Currently 2 problems
 - No exact prior knowledge of the necessary number of simulations
 - Non optimal job repartition due to significant execution and waiting time variations
 - Solution
 - Jobs write their results periodically
 - A statistical uncertainty is calculated in real time
 - Jobs are done when the uncertainty threshold (or the max default number of simulated particles) is reached
 - No re-submission, no job cancelation
- **Testing and improving the scalability of our implementation**
 - Data management is currently rather centralized
- **Web portal**
 - A friendly-user web portal would ease the execution of THIS on the grid for people who are not familiar with grid technologies

- **THIS is a typical application that can benefit from grid computing technologies**
 - Its parallelization is natural, as simulations can be run independently
- **Everything is not perfect yet, work is to be done**
 - Data management
 - Job submission optimization
 - User access – web portal
- **Running THIS on EGEE allows for a considerable speed-up in computation time**
 - Brings simulations lasting weeks (too much for a dose estimation) on one PC to only a few hours on EGEE



Enabling Grids for E-scienceE



**Thank You
for Your Attention!**
Questions?

