Contribution ID: **26** Type: **Demonstration**

WISDOM

Tuesday, 12 February 2008 16:00 (1 minute)

The environment evolved throughout the development: It was first made of a set of scripts that generate the jobs, submit the files and check regularly their status while they are on the Grid. Through this abstract we want to present the new environement that is based on the AMGA metadata catalog for more flexibility and on a Java environment that can be used through web services. The environment is very flexible and can be used for any type of bioinformatics application.In this new environment we control the job distribution, maintaing the choice of directly submitting the executable (push mode) or implementing a two-way submission where the system submits generic wrapper jobs which request their paylod (the executable and the input data) only when they start to be executed (pull mode).

3. Impact

With WISDOM we wanted to create a robust framework to allow biologists to integrate their software with and run it on EGEE. The system was designed to deploy high-throughput experiments on the grid, and is being reengineered to offer a fully interoperable web services interface, with connections to databases to store and query, in quasi-real-time, the statistics and results.
One of the major added values of this new architecture is that the whole system can be easily integrated in workflow engines that just call the ad-hoc operations.
The developments were focused on fault-tolerance, flexibility and scalability but several issues arose during the experiments. The relatively slow information system refresh rate can also cause outdated ranking when the job submission rates are high, this is the reason why we decided to introduce an internal rank system to address these information system issues.

4. Conclusions / Future plans

As a matter of fact, the environment has proved many times that it is adapted to run hi-throughput docking experiments on the grid. During the last deployment we successfully managed up to 70000 jobs producing almost 2TB during 10 weeks corresponding to more than 400 cpu years on a single computer. The next challenge, will be to integrate WISDOM in whole workflows to apply to other bioinformatics activities.

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

Drug-Discovery, Large scale deployment, Bioinformatics

1. Short overview

During 2005,2006 and 2007 four biomedical data challenges were run on the EGEE grid: two on malaria and two on avian flu. These deployments, based on relevant biological needs, were successfully achieved using most of the available resources on the Biomed virtual organisation. As a total, almost 700 years of computations were achieved during these 4 deployments using the WISDOM production environment and some in vitro tests have been already started with really interesting results.

If demonstration is requested please explain what visual or interactive aspects of the contribution necessitate a demonstration rather than a presentation or poster?

The goal of the demo should be to start some job submissions at the beginning of the User Forum and show in realtime the number of dockings that are done. As this application goal is to make hi-throughput docking, we need to put the emphasis on the number of docking, and also on the real-time monitoring and collection of results and statistics.

Author: Mr SALZEMANN, Jean (IN2P3/CNRS)

Co-authors: Mr BLOCH, Vincent (IN2P3/CNRS); Mr KASAM, Vinod (SCAI Fraunhofer)

Presenters: Mr SALZEMANN, Jean (IN2P3/CNRS); Mr BLOCH, Vincent (IN2P3/CNRS)

Session Classification: Demonstrations

Track Classification: Scientific Results Obtained Using Grid Technology