



E2GRIS2 REPORT

Rafael Ferreira da Silva

Universidade Federal de Campina Grande – UFCG (Brazil)

October 09th, 2009

1. Overview	2
2. Technical Issues.....	2
2.1. gLite Services Used.....	2
3. Preliminary Results.....	3
4. General Comments about the School	4
5. Attachements	5
5.1 Example of a JDL File	5
5.2 Script File	5

1. Overview

This report summarizes the porting of the BioMD application in the Second EELA-2 Grid School held at Applied Physics and Advanced Technology Centre (CFATA), Autonomous National University of Mexico, Queretaro, Mexico.

Application:	BioMD - Molecular Dynamics Using Grid Environment
Contact Person:	João Herminio Martins Da Silva Jeferson Ricardo Rodrigues da Silva
E-mail:	jhms@biof.ufrj.br eousoujeferson@gmail.com
Project URL:	http://applications.eu-eela.eu/application_details.php?l=20&ID=50
Domain:	Biomedicine
Description:	<p>Molecular dynamics is a powerful tool to investigate stability, conformational changes and atomic-level structures of bio-molecular complexes. In the last years, the rational drug design approach, based on exploring the structural complementary concept to generate specific antagonists for target molecules, has been endorsed. Knowing in detail the spatial structure of enzymes, by means of developed computational tools and help of a structural data bank, we can attempt to identify molecules whose specificity can match the active site of such enzyme, thus, leading them to become its inhibitors. The dynamical analysis of an enzyme is required in order to understand its function and interaction with inhibitors at atomic scale. Thus, theoretical models need long MD simulations in order to extract valuable information. Through those simulations, atoms and molecules motion is integrated using Newtons's equations. The GROMACS package is used as a fast-algorithm for MD simulations, running in parallel and using MPI Library. Grid computing, organized in a scientific workflow helps to minimize the computational time of the simulation.</p>

2. Technical Issues

2.1. gLite Services Used

The BioMD application uses MPI jobs to process their simulations. The applications was compiled with MPICH2 and stored in the Storage Element (SE). Each job executes a bash script which retrieves the application from the SE and performs the simulation. The WatchDog service is used to monitor the result of the simulation, since a real job takes more than 15 days to perform. Due to the duration of the jobs, the application was ported using the MyProxy service.

Table 1. Summary of the gLite services used by BioMD

Service	Usage
CE, WMS, WNs	Perform simulation of the moleculars
SE, LFC	Store the executable files
MyProxy	Long-term jobs
WatchDog	Monitoring

The students also developed a portal to submit and monitoring jobs as well a mechanism to retrieve the outputs. Figure 1 show a screenshot of the portal.

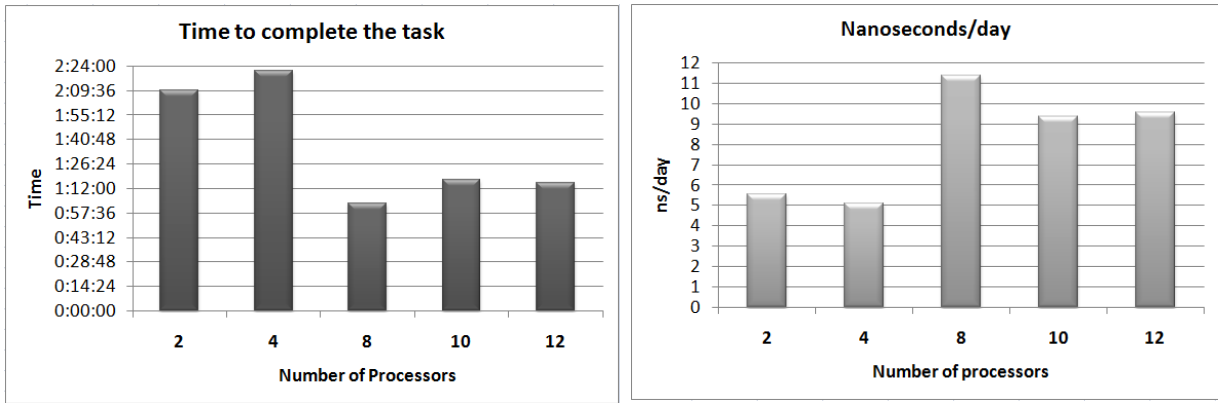


Figure 1. BioMD Portal screenshot

The application was also ported to the DIRAC middleware, which is able to perform MPI jobs on sites that do not support MPI.

3. Preliminary Results

The students performed a testbed experiment in order to measure the performance of their application on the grid. They submitted jobs which requests 2, 4, 8 and 12 processors. Figure 2.a exploits the accomplished results for a job which produces a motion with 1 ns. Figure 2.b exploits the estimated nanoseconds production per day for the application. Figure 3 presents the result of the testbed simulation which is possible to see the interaction of a protein with other elements.



(a)

(b)

Figure 2. (a) Time to complete a task to produce 1 nanoseconds. (b) Estimated nanoseconds production per day

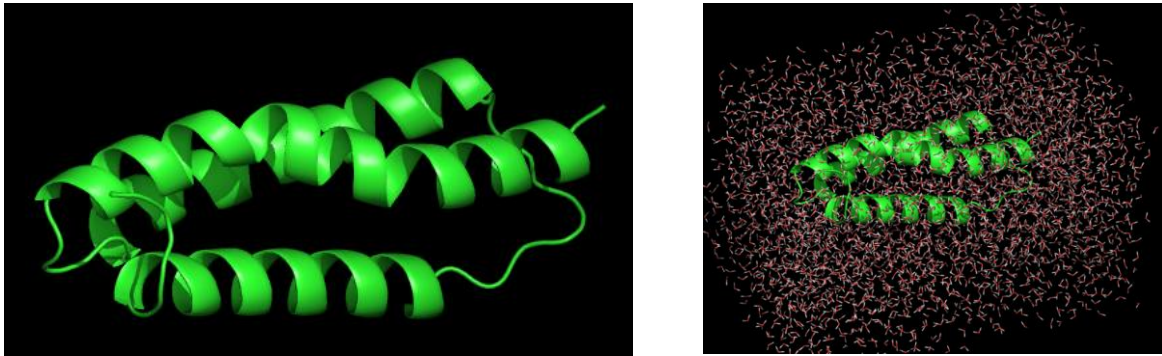


Figure 3. Result of the testbed simulation

A standalone experiment was made with a real molecular which takes about 42 days to generate a motion with 30 ns in their machines with 4 processors. In the grid context, they submitted a job that requests 8 processors to generate 1 ns of this motion. By estimation, the generation of a motion with 30 ns in the grid with this configuration will take about 18 days.

4. General Comments about the School

The application was successfully deployed in the Grid School running its jobs in the production infrastructure (prod.vo.eu-eela.eu). The only problem found by the students was the low number of sites which is able to run MPICH2 jobs.

5. Attachements

5.1 Example of a JDL File

```
Type="Job";
JobType="Normal";
CpuNumber = 10;
Executable= "/bin/bash";
Arguments = "mdrun_mpi.sh";
StdOutput="april.out";
StdError="april.err";
InputSandbox={"mdrun_mpi.sh", "april.tgz"};
OutputSandbox={"april_full.log", "april.out", "april.err", "april_full.trr",
               "april_full.edr", "april_full.tpr", "april_full.pdb"};
ShallowRetryCount=5;
Requirements = Member("MPICH2",
                     Other.GlueHostApplicationSoftwareRunTimeEnvironment);
MyProxyServer = "px.eela.ufrj.br";
```

5.2 Script File

```
#!/bin/bash

export LCG_CATALOG_TYPE=lfc
export LFC_HOST=lfc.eela.ufrj.br
export LCG_GFAL_INFOSYS=bdii.eela.ufrj.br
export PATH=$MPI_MPICH2_PATH/bin/:$PATH
export LD_LIBRARY_PATH=$MPI_MPICH2_PATH/lib/:$LD_LIBRARY_PATH

echo "MPD_SECRETWORD=BIOMD$RANDOM" >> ~/.mpd.conf
chmod 0600 ~/.mpd.conf

if [ "x$PBS_NODEFILE" != "x" ] ; then
    echo "PBS Nodefile: $PBS_NODEFILE"
    HOST_NODEFILE=$PBS_NODEFILE
fi
if [ "x$LSB_HOSTS" != "x" ] ; then
    echo "LSF Hosts: $LSB_HOSTS"
    HOST_NODEFILE=.pwd./lsf_nodefile.$$
    for host in ${LSB_HOSTS}
    do
        echo $host >> ${HOST_NODEFILE}
    done
fi
if [ "x${HOST_NODEFILE}" = "x" ] ; then
    echo "No hosts file defined. Exiting..."
    exit
fi
```

```

if [ "x${MPI_PATH_TO_SHARED_HOME}" != "x" ] then
  if [ ! -d ${MPI_PATH_TO_SHARED_HOME} ] then
    echo "Shared home is set, but it is not accessible"
    exit 1
  else
    echo "Shared home exists and is ${MPI_PATH_TO_SHARED_HOME}"
  fi
else
  echo "No shared home is accessible"
  if [ "x${MPI_SSH_HOST_BASED_AUTH}" == "xyes" ] then
    echo "Using SSH to copy the files to all nodes"
    for i in `cat $HOST_NODEFILE | grep -v ${HOSTNAME} | sort -u`; do
      echo "Copying files to $i"
      ssh ${i} mkdir -p `pwd`
      scp -rp . "${i}:\`pwd\`"
      # Copy the proxy certificate
      scp -p ${X509_USER_PROXY} ${i}:${X509_USER_PROXY}
      # Copy the MPD secret file
      scp -p ~/.mpd.conf ${i}:
    done
  else
    echo "Uh oh, cannot use SSH to copy the files to all WN.s"
    exit 1
  fi
fi

mpdboot -f ${HOST_NODEFILE}

lcg-cp --vo prod.vo.eu-eela.eu
  lfn:/grid/prod.vo.eu-eela.eu/BIOMD/gromacs4.0.5.tar.gz
  file://$PWD/gromacs4.0.5.tar.gz
tar -zxf gromacs4.0.5.tar.gz
tar xvzf april.tgz
bin/grompp -f em.mdp -c b4md_water_ion.pdb -p april.top -n index.ndx -o
  april_em.tpr -maxwarn 10
mpirun -np 10 bin/mdrun -s april_em.tpr -o april_em.trr -c april_em.pdb -e
  april_em.edr -g april_em.log
bin/grompp -f april_full.mdp -c april_em.pdb -p april.top -n index.ndx -o
  april_full.tpr -maxwarn 10
mpirun -np 10 bin/mdrun -s april_full.tpr -o april_full.trr -c
  april_full.pdb -e april_full.edr -g april_full.log

mpdallexit

```