Computational Biology Software Overview

Le Yan
HPC Consultant
User Services @ LONI
Outline

- Overview of available software packages for computational biologists
- Demo: run NAMD on a LONI Linux cluster
What are available on LONI

- Bioinformatics packages
  - Sequence analysis
  - Computational evolutionary biology
  - Protein-protein docking

- Molecular dynamic packages
  - Classical molecular dynamics
  - Quantum mechanical calculation

- Analysis tools
# Bioinformatics Packages

<table>
<thead>
<tr>
<th>Package Name</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linux clusters</td>
</tr>
<tr>
<td></td>
<td>Queen Bee</td>
</tr>
<tr>
<td>Phylip</td>
<td>x</td>
</tr>
<tr>
<td>MrBayes</td>
<td>x</td>
</tr>
<tr>
<td>Clustalw</td>
<td>x</td>
</tr>
<tr>
<td>MPIBlast</td>
<td>x</td>
</tr>
<tr>
<td>Tree Puzzle</td>
<td>x</td>
</tr>
<tr>
<td>Autodock</td>
<td>x</td>
</tr>
</tbody>
</table>
## Classical MD packages

<table>
<thead>
<tr>
<th>Package Name</th>
<th>Linux clusters</th>
<th>AIX clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Queen Bee</td>
<td>Blue Dawg</td>
</tr>
<tr>
<td></td>
<td>Eric</td>
<td>Ducky</td>
</tr>
<tr>
<td></td>
<td>Painter</td>
<td>Zeke</td>
</tr>
<tr>
<td></td>
<td>Oliver</td>
<td>Neptune</td>
</tr>
<tr>
<td></td>
<td>Louie</td>
<td>Lacumba</td>
</tr>
<tr>
<td>NAMD</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Amber</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>LAMMPS</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>GROMACS</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>
# Quantum Mechanical Packages

<table>
<thead>
<tr>
<th>Package Name</th>
<th>Availability</th>
<th>Linux clusters</th>
<th>AIX clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Queen Bee</td>
<td>Bluedawg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Eric</td>
<td>Ducky</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Painter</td>
<td>Zeke</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oliver</td>
<td>Neptune</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Louie</td>
<td>Lacumba</td>
</tr>
<tr>
<td>Gaussian</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>NWChem</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>CPMD</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>
Analysis Tools

- Visual Molecular Dynamics (VMD)
  - View and analyze the results of MD simulations
  - Organize, display and analyze both sequence and structure data for proteins and nucleic acids
  - Support over 60 molecular file formats
- AmberTools
  - Trajectory analysis, nucleic acid builder etc.
- GROMACS
  - Trajectory analysis etc.
Software Request Policy

• Users can install software packages in their own user space

• Users can request a software package to be installed on LONI clusters
  • The software will be installed in the user's home directory, or
  • The software will be installed in the public domain (/usr/local/packages) ONLY IF it can be proved to us that multiple users are using it
  • It is users' responsibility to provide the (site) license for commercial software packages
NAMD

- A parallel MD code designed for high-performance simulation of large biomolecular systems developed by Theoretical Biology Group at University of Illinois

- Features
  - Force field compatibility
  - Efficient full electrostatics algorithms
  - Multiple time stepping
  - Input and output compatibility
  - Dynamic simulation options
  - Dynamic load balancing
Run NAMD – Prepare Input Files

- Prepare input files
  - PDB (Protein Data Bank) files
  - PSF (Protein Structure File) files
  - Parameter files
  - DCD trajectory files

- Tutorials can be found at
  - http://www.ks.uiuc.edu/Training/Tutorials

- A sample set of input files can be found at:
  - http://www.ks.uiuc.edu/Research/namd/utilities/apoa1.tar.gz
Run NAMD – Write a Job Script

#!/bin/sh
#SBATCH -A <your_allocation>
#SBATCH -q checkpt
#SBATCH -M lyan1@cct.lsu.edu
#SBATCH -l nodes=4:ppn=4
#SBATCH -l walltime=06:00:00
#SBATCH -V
#SBATCH -o NAMD_test.out
#SBATCH -e NAMD_test.err
#SBATCH -N namdtest

export EXEC=namd2
export EXEC_DIR=/usr/local/packages/namd-2.6-mvapich-1.0-intel10.1
export WORKDIR=$PBS_O_WORKDIR
export NPROCS=`wc -l $PBS_NODEFILE`
export CONV_RSH=ssh
cd $WORKDIR
mpirun -machinefile $PBS_NODEFILE -np $NPROCS $EXEC_DIR/$EXEC apoa1.namd
Run NAMD – Submit and Manage Jobs

- `qfreeloni` – show the number of free nodes on all Linux clusters
  - Help to decide which cluster to run on
- `qsub <job script>` – submit a job
- `qdel <job id>` – cancel a job
- `qalter <job id>` – alter a job
Run NAMD – Check Job Status

**qstat -u <username>**

```
[lyan1@louie2 1.62_cart]$ qstat -u huiwu
louie2:
```

<table>
<thead>
<tr>
<th>Job ID</th>
<th>Username</th>
<th>Queue</th>
<th>Jobname</th>
<th>SessID</th>
<th>NDS</th>
<th>TSK</th>
<th>Memory</th>
<th>Time</th>
<th>S</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>140647.louie2</td>
<td>huiwu</td>
<td>workq</td>
<td>amb_par</td>
<td>15134</td>
<td>4</td>
<td>1</td>
<td>--</td>
<td>48:00</td>
<td>R</td>
<td>07:35</td>
</tr>
<tr>
<td>140648.louie2</td>
<td>huiwu</td>
<td>workq</td>
<td>amb_par</td>
<td>--</td>
<td>4</td>
<td>1</td>
<td>--</td>
<td>48:00</td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>140649.louie2</td>
<td>huiwu</td>
<td>workq</td>
<td>amb_par</td>
<td>--</td>
<td>4</td>
<td>1</td>
<td>--</td>
<td>48:00</td>
<td>Q</td>
<td></td>
</tr>
</tbody>
</table>

**qshow <job id>**

- Check the output to make sure that the cpus are fully loaded
- Check the time stamp, size and content of the output files