



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

Package Name	Availability											
	Linux clusters						AIX clusters					
	Queen Bee	Eric	Painter	Oliver	Louie	Poseidon	Bluedawg	Ducky	Zeke	Neptune	Lacumba	
Phylip	x											
MrBayes	x		x	x	x	x						
Clustalw	x											
MPIBlast	x											
Tree Puzzle	x											
Autodock	x		x	x	x	x						





Classical MD packages

Package Name	Availability										
	Linux clusters						AIX clusters				
	Queen Bee	Eric	Painter	Oliver	Louie	Poseidon	Bluedawg	Ducky	Zeke	Neptune	Lacumba
NAMD	x	x	x	x	x	x					
Amber	x	x						x			
LAMMPS	x	x	x	x	x	x	x	x	x	x	x
GROMACS	x	x	x				x	x	x	x	x





Quantum Mechanical Packages

Package Name	Availability										
	Linux clusters						AIX clusters				
	Queen Bee	Eric	Painter	Oliver	Louie	Poseidon	Bluedawg	Ducky	Zeke	Neptune	Lacumba
Gaussian		x	x				x	x	x		
NWChem	x	x	x	x	x	x					
CPMD	x	x	x	x	x	x	x	x	x	x	x





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
#PBS -A <your_allocation>
#PBS -q checkpt
#PBS -M lyan1@cct.lsu.edu
#PBS -l nodes=4:ppn=4
#PBS -l walltime=06:00:00
#PBS -V
#PBS -o NAMD_test.out
#PBS -e NAMD_test.err
#PBS -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:
```

Job ID	Username	Queue	Jobname	SessID	NDS	TSK	Req'd Memory	Req'd Time	Elap S	Time
140647.louie2	huiwu	workq	amb_par	15134	4	1	--	48:00	R	07:35
140648.louie2	huiwu	workq	amb_par	--	4	1	--	48:00	Q	--
140649.louie2	huiwu	workq	amb_par	--	4	1	--	48:00	Q	--

- `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

