Date Published: 08/07/2022

Student Cluster Compeititon - Tutorial 5

Table of Contents

Overview

Tutorial 5 has you compile the GROMACS scientific software. You will then run the software on data sets that are provided to you and upload the results.

In this tutorial you will:

- [ ] Download, install and compile the GROMACS benchmark.
- [ ] Run the GROMACS benchmark on sample data sets.
- [ ] Visualise the output from your GROMACS runs on your local computer.

Date Published: 08/07/2022

GROMACS Application Benchmark

GROMACS is a versatile package to perform molecular dynamics, i.e. simulate the Newtonian equations of motion for systems with hundreds to millions of particles. It is primarily designed for biochemical molecules like proteins, lipids and nucleic acids that have a lot of complicated bonded interactions, but since GROMACS is extremely fast at calculating the nonbonded interactions (that usually dominate simulations) many groups are also using it for research on non-biological systems, such as polymers.

The files required for this tutorial can be found on the ACE Lab SSH server (ssh.ace.chpc.ac.za) under the `/apps/gromacs` folder. The archive name should be `gromacs_benchmarks.tar.gz`.

Part 1: Installation

Detailed installation instructions can be found at: http://manual.gromacs.org/current/install-guide/index.html, but here's a general installation overview:

1. Ensure you have an up-to-date `cmake` available on your system.

2. You will also require a compiler such as the GNU `gcc`, Intel `icc` or other, and **MPI (OpenMPI, MPICH, Intel MPI or other)** be installed on system. Your **PATH** & **LD_LIBRARY_PATH** environment variables should be set up to reflect this.

3. Compile GROMACS **with MPI support** from source using `cmake`.

Date Published: 08/07/2022

Part 2: Benchmark

You have been provided two **GROMACS** benchmarks. The first benchmark **(adh_cubic)** should complete within a few minutes and has a small memory footprint, it is intended to demonstrate that your installation is working properly. The second benchmark **(1.5M_water)** uses more memory and takes considerably longer to complete. The metric which will be used to assess your performance is the **ns/day** (number of nanoseconds the model is simulated for per day of computation), quoted at the end of the simulation output. **Higher is better**.

Benchmark 1 (adh_cubic):

Ensure that your GROMACS /**bin** directory is exported to your **PATH**. You should be able to type `gmx_mpi --version` in your terminal and have the application information displayed correctly. The first task is to pre-process the input data into a usable format, using the `grompp` tool:

```
[...@node ~]$ gmx_mpi grompp -f pme_verlet.mdp -c conf.gro -p topol.top -o md_0_1.tpr
```

You will need to prepare a **Slurm batch script**, `gromacs_mpi.sh`. Modify the variables to appropriate values.

```
#!/bin/bash
#SBATCH --nodes=XXXX
#SBATCH --ntasks-per-node=XXXX
#SBATCH --cpus-per-task=XXXX

# !!!! Depending on your environment configuration, uncomment or add the following
# --------------------------------------------------------------------------------
#module load gromacs
#ml load gromacs
#export PATH and LD_LIBRARY_PATH
mpirun gmx_mpi mdrun -nsteps 5000 -s md_0_1.tpr -g gromacs.log
```

Then execute the script from you head node, which will in turn launch the simulation using MPI and write output to the log file `gromacs_log`.

You may modify the `mpirun` command to optimise performance (significantly) but in order to produce a valid result, the simulation must run for **5,000 steps**. Quoted in the output as:

```
"5000 steps,     10.0 ps."
```

!!! Please be able to present the instructors with the output of `gmx_mpi --version`. Also be able to present the instructors with your Slurm batch script and `gromacs_log` files for the **adh_cubic** benchmark.

Benchmark 2 (1.5M_water):

Pre-process the input data using the `grompp` command

```
[...@node ~]$ gmx_mpi grompp -f pme_verlet.mdp -c out.gro -p topol.top -o md_0_1.tpr
```

Using a batch script similar to the one above, run the benchmark. You may modify the mpirun command to optimise performance (significantly) but in order to produce a valid result, the simulation must run for 5,000 steps. Quoted in the output as:

```
"5000 steps,     10.0 ps."
```

```
Date Published: 08/07/2022
```

!!! Please be ready to present the `gromacs_log` files for the **1.5M_water** benchmark to the instructors.

Part 3: Protein Visualisation

> **! >>> You will need to work on your personal computer (or laptop) to complete this section.**

You are able to score bonus points for this tutorial by submitting a visualisation of your **adh_cubic** benchmark run. Follow the instructions below to accomplish this and upload the visualisation.

Download and install the VMD visualization tool by selecting the correct version for your operating system. For example, for a Windows machine with an Nvidia GPU select the “Windows OpenGL, CUDA” option. You may need to register on the website.

```
https://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=VMD
```

Use the `WinSCP` application for Windows, or the `scp` command for Linux to copy the output file `confout.gro` of the **adh_cubic** benchmark from your cluster to your PC. Attempting to visualise the larger "1.5M_water" simulation is not necessary and not recommended due to memory limitations of most PCs.

1. Open VMD, select **File** then **New Module…**, click **Browse…** and select your `.gro` file.
2. Ensure the filetype was detected as **Gromacs GRO** then click **Load**. In the main VMD window you will see that 134177 particles have been loaded. You should also see the display window has been populated with your simulation particle data.

   You can manipulate the data with your mouse cursor: zoom with the mouse wheel or rotate it by dragging with the left mouse button held down. This visualisation presents a naturally occurring protein (blue/green) found in the human body, suspended in a solution of water molecules (red/white).
3. From the main VMD window, select **Graphics** then **Representations..**.
4. Under **Selected Atoms**, replace **all** with **not resname SOL** and click **apply**. You will notice the water solution around your protein has been removed, allowing you to better examine the protein.
5. In the same window, select the dropdown **Drawing Method** and try out a few different options. Select **New Cartoon** before moving on.
6. From the main VMD window, once again select **Graphics** then **Colors**. Under **Categories**, select **Display**, then **Background**, followed by **8 white**.
7. Finally, you are ready to render a snapshot of your visualisation. From the main window, select **File** then **Render…**, ensure **Snapshot…** is selected and enter an appropriate filename. Click **Start Rendering**.

Simulations like this are used to to develop and prototype experimental pharmaceutical drug designs. By visualising the output, researchers are able to better interpret simulation results.

> Copy the resulting `.bmp` file(s) from yout cluster to your local computer or laptop and demonstrate this to your instructors for bonus points.