Parallel Numerical Simulations with MPI and SGE

Unisys Scholars Project
Paul Arias

Overview

- **Lengthy Computer Simulations**
  - Engineering Computational Design
  - Scientific High Performance Computing
  - Financial Computing
- **What is Parallel Computing**
  - Message Passing Interface
  - Open Source Solutions
- **What is Sun Grid Engine**
  - Resource manager
  - Costs
- **Why use either**

Simulations That Take Time

- **Programs need resources to run**
  - Large programs require more memory than others
  - Output of executed program can exceed available memory
  - Need to organize and efficiently use resources of computational clusters
- **Large Networks of Nodes**
  - System needs organization
  - Needs Queue for managing large number of jobs

Solution

Parallel computing



Using Sun Grid Engine

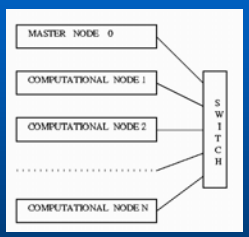



Implementation

- **Create a computational cluster for use on parallel computing**
- **Install MPI onto Linux machines**
- **Install SGE onto cluster machines**

Parallel Computing Structure

- **Parallel computing environment**
  - Work with multiple processor environment
  - Programmer's responsibility to write MPI code
- **SGE manages resources efficiently**
- **Higher productivity when system allows tight control of distributed parallel processes**



MPI Software Available

- **Commercial versions**
  - **ChaMPion/Pro**
  - **Alpha Data MPI**
  - **Hitachi MPI**
  - **HP MPI**
  - **MPI/SX**
- **Open Source versions**
  - **MPICH**
  - **DISI**
  - **LAM/MPI**
  - **MPI/FM**
  - **OPEN_MP**

Our Choices

- **Use MPICH**
  - Developed by Argonne National Laboratory
  - Readily available with instructions on how to compile
  - Open source
  - Can be used with various platforms, including RedHat Linux 9.0
  - Other newer commercial releases only worked on certain platforms

Configuring MPICH

- **Working knowledge of Linux systems administration**
  - Network Information Services
  - Network File Transfer
  - RSH
  - A bit of shell scripting

  “Rsync” for synchronizing installation

Configuring MPICH (continued)

- **Compilation of code and configuration**
  - Use Modified Configuration Script

```
#!/bin/bash
export RSHCOMMAND=rsh
./configure -comm=shared
```

Note shared memory usage

Configuring MPICH (Continued)

- **Network Information Services**
  - Required for use with MPI and for parallel environment.
  - Synchronize user information between the nodes; sharing accounts over NIS. Highly efficient when large number of users.
  - Can also use Lightweight Directory Access Protocol or Synchronize passwords files. Usually done with few number of users.

Configuring MPICH (Continued)

- **Network File System**
  - Also a requirement for MPI operation and for parallel environment
    - Executable need only be present in one file system on one computer then be mounted to the other computers

    NFS is slow, advance cluster used PVFS or Luster for sharing files

  - Using Automount
    - Shares particular system file over nodes
    - Unlike Standard mount, Automount provides on demand form of mounting file system and then un-mounts automatically

Configuring MPICH (Continued)

- **RSH – Remote SHell client**
  - **Managing nodes**
  - **Starts parallel processes**
  - **Modify /etc/hosts.equiv for rsh between computational nodes**

Problems

- **Could not kill the process easily**
  - **“ctrl+C” will end only one process, but since the program is running on multiple processor must kill each independently**
  - **Not very clean**

Solution: Use SGE

- SGE will organize resources and also make execution and termination of programs more manageable in multi-user and multi-process environment
- Add simplicity to the user. SGE will take the file, find available resources and submit them, making the process of running parallel simulation more user friendly

Sun Grid Engine

- **Software Available**
  - OpenPBS
  - MauiME
  - Sun Grid Engine 5.3
- **Used SunGrid Engine**
  - Sun is most efficient
  - Forerunner in Grid Technology

Sun Grid Engine Configuration

- **Configure master node/administrator and then configure the computational**
  - **Requires setting up queue for all of the computational nodes**
  - **Specify number of processors in each node**
  - **Need to specify parallel environment to MPI**

SGE Configuration |tight| (continued)

- **Modifying the startmpi.sh script**
  - **Must make sure that the mpi local file path is correct**
  - **Must also specify location of the SGE local files**
- **Modifying the stopmpi.sh script**
  - **Must make sure that the mpi local file path is correct**
  - **Must also specify location of the SGE local files**

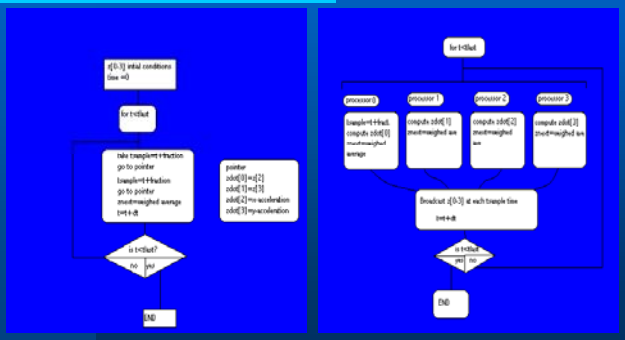
SGE Submission of Executable

- **Use mpi_submi.sh**
  - **Name if file is specified in shell script**
    - myjob=rkftwobody
  - **Single line initiates MPI and submits the job to the queue system**
    - $SGE_ROOT/mpi/watchMPI.sh &
    - $MPIR_HOME/bin/mpirun -np $NSLOTS -machinefile $TMPDIR/machines $myjob

Observations

- **Program ran slower as parallel simulation**
  - Removal of pointer
  - Addition of many lines of MPI_Bcast
  - Shows that communication between processors slows things down
- **MPI not best for System of Ordinary Differential Equations**

Comparison of Code



Observations (Continued)

- **Manual use of MPI is tedious**
  - Options to specify
    - Machine file name
    - Number of processors
  - Using submit script and SGE we can increase the efficiency of dealing with not only multiple processors, but also multiple processor JOBS
  - Queue is great way to supervise the running jobs

Conclusion

- **MPI is a powerful tool for producing scientific data from computer simulations**
- **Software that uses MPI**
  - FLUENT (CFD)
  - GASP (Gas Dynamics)
  - AMBER (Molecular Dynamics)
  - NAMD (Molecular Dynamics)
- **Areas of Research where MPI is used**
  - Automotive design
  - Financial modeling
  - Computational physics
  - Computational Fluid Dynamics
  - Materials Research

Conclusion (continued)

- **Sun Grid Engine does an exemplary job of managing the resources available in a computational cluster**
- **As more and more advances in parallel computing with multiple processor nodes and network parallel computing continues, development of Grid Engine technology will become essential in managing these resources**

Special Thanks

Dr. David Gardiner
Vice President, Architecture and Technology, Unisys Corporation
Dr. Doyle Knight
Professor, Department of Mechanical and Aerospace Engineering
Dr. Alexei Kotelnikov
Systems Programmer and Administrator, Engineering Computing Services
Mr. Amit Freeman
Department of Electrical and Computer Engineering
My Fellow 2004-2005 Unisys Scholars
Ms. Diane Palla
Mr. Malik Khan
Mr. Lucas Machado