Introduction to the Message Passing Interface: Jazz Intro

MPI Tutorial

Rusty Lusk

LCRC Staff

Mike Dvorak, App. Engineer
Katherine Riley, App. Engineer
Susan Coghlan, Systems

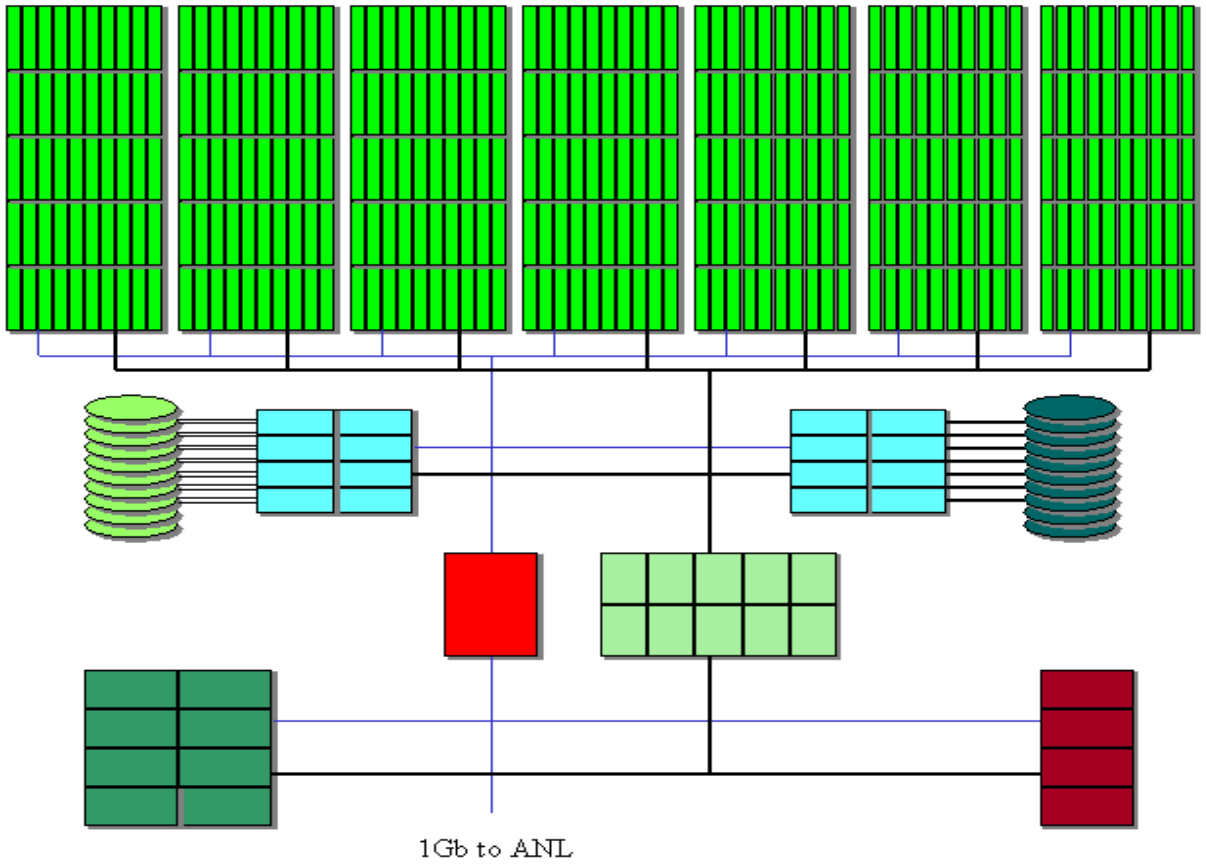
September 24, 2003

Tutorial Schedule

- 8:30-9:30 Intro to Jazz
  - Mike Dvorak
- 9:30-10:45 Intro to MPI
  - Rusty Lusk
- 10:45-11:00 Break
- 11:00-12:30 Intermediate MPI
  - Rusty Lusk

The ANL LCRC Computing Cluster

Supplier: Linux Networx



350 computing nodes:
- 2.4 GHz Pentium IV
- 50% w/ 2 GB RAM
- 50% w/ 1 GB RAM
- 80 GB local scratch disk
- Linux

10 TB global working disk:
- 8 dual 2.4 GHz Pentium IV servers
- 10 TB SCSI JBOD disks
- PVFS file system

10 TB home disk:
- 8 dual 2.4 GHz Pentium IV servers
- 10 TB Fiber Channel disks
- GFS between servers
- NFS to the nodes

Network:
- Myrinet 2000 to all systems
- Fast Ethernet to the nodes
- GigE aggregation

Support:
- 4 front end nodes: 2x 2.4 GHz PIV
- 8 management systems

Allocations and Projects

- All ANL employees--> 1,000 node/hour one time allocation
  - For writing/testing code for a project
  - One time allocation
  - You need a project after your initial allocation is gone
  - Jobs will fail in queue

Requesting a Project

- Projects can apply for large amounts of time
- Queuing preference given to:
  - Users w/o other systems e.g. NERSC
  - New projects
  - Scheduler modifications forthcoming
- Do you already belong to a project?
  - In your divisional group?
  - Look on web on project guidelines

Connecting to Jazz

- Login Nodes
  - ssh to jazz.lcrc.anl.gov
  - dual processor, 2.4 GHz nodes
  - DNS round robin to jlogin*.lcrc.anl.gov
  - Normal UNIX shells (tcsh/bash/csh)
  - Code development
  - Submit jobs, monitor jobs
  - Run visualization apps e.g. Jumpshot, Totalview
  - Debugging short jobs

Intro to Softenv

- Automatically add software to your environment
- Installed packages
  - Not in usual Unix locations i.e. /usr/bin, /usr/local/bin, etc...
  - Most in /soft/apps/packages
- Dynamically adds/removes
  - PATH environment variables
  - Library environment variables
  - Other variables such as PYTHONPATH

Using Softenv

- To find what is available
  - Jazz software documentation page
  - *% softenv -k mpich*
- Only put in software you need
  - Can create conflicts e.g. two MPI compilers in PATH
  - Path that is too long
- @default at end (for safety)

Adding to your Softenv

- Two ways to add software
  - Add tokens to your .soft file
    - add keywords: +totalview
    - add macros: @all-mpich_gm-intel7
  - Use *soft add/delete*
    - For temporary use only such as debuggers
    - % soft add +totalview
    - % soft delete @climate

Adding packages to Softenv

- We can install software on Jazz via Softenv
  - Other users can use libraries and applications
  - Contact systems@lcrc.anl.gov
- We can create project appropriate macros
  - New users have a complete environment
  - Good example, @climate macro
    - Adds all relevant software for running and post processing
    - Leaves the compiler choice up to user

File Systems

- Global File System (GFS)
  - Sistina vendor
  - Your home directory i.e. /home/<username>
  - Available on all compute nodes
  - 10 TB of space (about 80% full)
  - Backed up nightly to tape
  - Moderate performance for applications
  - No parallel file support (no locking)

File Systems (cont.)

- Parallel Virtual File System (PVFS)
  - Developed in MCS
  - 10 TB (10,000 GB) of space
  - Mounted at /pvfs/scratch/<username>
  - NOT backed up, data can be lost
  - Cannot run executables from this space
  - No symbolic link support

Jazz Compilers

- Compilers
  - C/C++ : Gnu, Portland Group, Intel
  - F77: Gnu, Portland Group, Intel, Absoft
  - F90: Portland Group, Intel, Absoft
  - F95: NAG
- Default MPICH compiler is Intel 7.0 over Myrinet
- Always compile applications on Jazz
  - Never build binaries on another machine

Debuggers and Profilers

- Intel Debugger – IDB
- GNU Debugger - GDB
- Totalview
  - Allows parallel debugging of MPI programs
  - Has a nice graphical user interface
- Jumpshot
  - Developed in MCS
  - Allows to automatically visualize MPI calls

Getting Help on Jazz

- Check the Jazz FAQ
- Review the Jazz web pages
- Two application engineers available
  - Help compiling code and installing software
  - Writing job submission scripts
  - Using Jazz software libraries to solve problems
  - Performance improvement consultation
  - Can set up in person appointments
  - Email consult@lcrc.anl.gov

Getting Help on Jazz (cont.)

- System problems
  - Myrinet errors
  - Nodes down
  - PBS problems
  - Software installation problems
  - Reservation requests
  - Any other non-application issues
  - Email systems@lcrc.anl.gov

Mailing Lists

- Staying current with system news
  - Run % *notifyme -y* to receive critical system news
  - Useful when you are in production mode on Jazz
  - Can be turned off with *% notifyme -n* when work is finished
- Jazz Users <jazz-users@lcrc.anl.gov>
  - All users subscribed
  - Can't unsubscribe
  - Restricted posts

Managing Project Allocations

- *qbank* command tells project balances, transactions, and project info
- Several basic qbank commands
  - Set your default LCRC project on Jazz
    - *% lcrc-qbank -s default <projectName>*
  - Find your project balance
    - % lcrc-qbank -q balance
  - Find all your transactions
    - *% lcrc-qbank -q trans*

Jazz Job Submission

- Portable Batch Scheduler (PBS)
  - Handles scheduling, starting, stopping of jobs
  - Highly configurable by the user
    - Email notification when jobs finish
    - Interactive job submission
  - Works with the account system (qbank) to deduct hours from users accounts
  - Most commands start with 'q' e.g. qstat, qsub, qdel

Two Jazz Job Queues

- Batch queue
  - Default
  - Normal submission queue
  - Almost all cluster nodes
- Shared queue
  - 8 nodes always available
  - Multiple users can run on these 8 nodes
  - Useful for debugging
  - Use % *qsub -q shared* ...

Submitting Jobs to PBS

- Most useful way, write shell scripts with PBS directives (later)
- Two main way users submit jobs
  - Batch submission
    - Use % *qsub <scriptname>*
    - Check job status with *% qstat -a*
  - Interactive job
    - Use *% qsub -I [job options]*
    - When job runs, your terminal gives you interactive PBS environment

Queue Status on Jazz

- How many nodes available on Jazz?
  - Run *% nodes*
  - Many options, run % *nodes -h*
  - Returns a list of available nodes
- Monitoring job status
  - Run *% qstat -a*
  - Returns all queued jobs in submission order

Deleting queue jobs

- Get the jobs ID from qstat
  - % *qstat -a | grep <username>*
- Run *qdel* to delete the job
  - % *qdel <jobID>*
  - To force job deletion
    % qdel -W force <jobID>
- Run *qstat -a* again to make sure your job has been deleted
- Be patient with job deletion

PBS Documentation

- Review Jazz 'PBS Tutorial' in Jazz doc
- Read online man documentation
  - *% man qsub*
  - *% man qstat*

Today's PBS Reservation

- Reservation made for today's tutorial
- To use the reservation
  - Add the '*-q R82943'* argument to your *qsub* command
  - Only applies for today's tutorial
  - Don't try use this after 1:00 pm, it won't work

MPI on Jazz

- Adding MPICH to your environment
  - *mpi<compiler>* sets environment vars e.g. lib paths
  - Provides *mpirun*
  - Don't use cc, f77, f90, etc... to compile
  - Use mpicc, mpif77, mpif90, etc... to compile
  - Don't hard code MPICH paths into makefiles, etc..
  - Combined libs of compiler and mpich and networking device (gm and p4)

Interactive User Session Overview

1. Add the MPICH Intel Myrinet compiler to your Softenv.
2. Copy over the MPI C Pi program to your local directory.
3. Compile the C Pi program using mpicc.
4. Write a short PBS script and submit the job to the queue.
5. Recompile the C Pi program using the Intel Ethernet drivers.

Exercise: Setting up Softenv

- Use your handout sheet to follow along.
- Ask Katherine or Susan for help if you run into problems.

Next

Introduction to MPI

Rusty Lusk

September 24, 2003