Debugging Parallel Programs for Multi-Core Chip Architectures

Chris Gottbrath
Product Manager, Etnus

Euro PVM / MPI
Bonn, Germany
Sept 2006
- **Intro**
  - The Multicore Age

- **Challenge**
  - What will multicore mean for HPC?

- **Solution**
  - TotalView Parallel Debugger
  - TotalView Multithreaded Debugger
Previously

- Chip clock rates have increased roughly in pace with Moores Law

Now

- Component size continues to decrease
- Clock rates are not increasing due to thermal issues

Dual core is only the beginning

- We will have more and more cores for the near future
  - How long will that trend be maintained?
- The notion of a balanced, single processor node will be a thing of the past
No big deal!
- We've had lots of experience with shared memory machines and even clusters of shared memory machines. We know how to deal with them, right?
- They just got cheaper is all..

Do we really?
- HPC cluster architecture drives program structure. Do we have a clear idea what program structure works best with SMP nodes?
  - Threads? Which model?
  - Multiply scheduled processes? With what communication mechanism?
  - Something else?
Increasing use of threads
- Great way to get concurrency

Increasing program complexity
- It is two-level parallelism any way you slice it

More people doing 'HPC'
- Anyone who needs performance is going to be looking at this issue

Increased need for high quality tools
- Debuggers
- Performance Tuning Tools
  - Sampling, Tracing, Analysis
HPC Software Tools should be

- Highly Scalable
- Focused on concurrency
  - Multi-thread
  - Multi-process
- Portable
- HPC Aware
  - Play nicely in a typical HPC environment
- Easy to use
What is TotalView?

- **Source Code Debugger**
  - C, C++, Fortran 77, Fortran90, UPC
  - Complex language features
  - Wide compiler and platform support
  - Multi-threaded Debugging
  - Parallel Debugging
    - MPI, PVM, Others
  - Remote Debugging
  - Memory Debugging Capabilities
    - Integrated into the Debugger
  - Powerful and Easy GUI
    - Visualization
  - CLI for Scripting
HPC Software Tools should be

- Highly Scalable
- Focused on concurrency
  - Multi-thread
  - Multi-process
- Portable
- HPC Aware
  - Play nicely in a typical HPC environment
- Easy to use
About Etnus TotalView

- Highly Scalable
  - Used often at hundreds of processes
  - Scales to thousands of processes
- Used on around 90 of the top 100 supercomputers
- Customers include:
  - Major government labs, research universities and technical computing centers worldwide
  - ISV’s and computer hardware vendors
  - Commercial enterprises in the Finance, Entertainment, Telecommunications, Energy, Aerospace, Climate Modeling and Automotive sectors
- Initially developed over 20 years ago
- Partner with vendors to provide consistent C/C++ and Fortran debugging solutions for complex codes on parallel platforms used by HPC Community
HPC Software Tools should be

- Highly Scalable
- Focused on concurrency
  - Multi-thread
  - Multi-process
- Portable
- HPC Aware
  - Play nicely in a typical HPC environment
- Easy to use
Threads

- Understand which thread you are looking at
- Easily switch thread context
- See the state of all threads
- View data
  - Shared
  - Stack
  - Thread Private
- Control Individual Threads
  - Synchronize
  - Run/Step
  - Hold
**Cluster Architecture**
- Single Front End (TotalView)
  - GUI and debug engine
- Debugger Agents (tvdsvr)
  - Low overhead, 1 per node
  - Traces multiple rank processes
- TotalView communicates directly with tvdsvrs
  - Not using MPI
  - Optimized Protocol

**Provides: Robust, Scalable, Minimal Interaction**
Process & Thread Control

- Focus on just one or a few processes or threads
  - Visibility of what is happening in whole job
  - Switch focus easily
- Control at process or thread level
  - Asynchronous
  - Step, Go/Halt
  - Hold
- Control with groups of threads or processes
  - Go/Halt
  - Step
- Synchronize processes or threads
  - Breakpoints
  - Run-To
  - Barriers
Tools for the multicore age

- **HPC Software Tools should be**
  - Highly Scalable
  - Focused on concurrency
    - Multi-thread
    - Multi-process
  - Portable
  - HPC Aware
    - Play nicely in a typical HPC environment
  - Easy to use
Supported Compilers, Distros and Architectures

- **Platform Support**
  - Linux x86, x86-64, ia64, Power
  - Mac Power and Intel
  - Solaris Sparc and AMD64
  - AIX, Tru64, IRIX
  - Cray X1, XT3, IBM BGL

- **Languages / Compilers**
  - C/C++, Fortran, UPC, Assembly
  - Many Commercial & Open Source Compilers

- **Parallel Environments**
  - MPI (MPICH1 & 2, LAM, Open MPI, poe, MPT, Quadrics, MVAPICH, & many others)
  - UPC
HPC Software Tools should be

- Highly Scalable
- Focused on concurrency
  - Multi-thread
  - Multi-process
- Portable
- HPC Aware
  - Play nicely in a typical HPC environment
- Easy to use
- **Simple Start Up**

- **Restarts**

- **Subset attach**
  - Helps with information overload
  - Scalability
MPI defines a process table
- Global variable
- Structure type with specified fields
  - Host, PID, Executable

TotalView reads & writes data
- Directly in the target program using ptrace()

Synchronization
- TV notified target being debugged
- MPI notifies TV that the table is ready
- MPI processes pause and wait to be attached
- TV releases the process
- **Hangs & Deadlocks**
- **Pending Messages**
  - Receives
  - Sends
  - Unexpected
- **Inspect**
  - Individual entries
- **Patterns**
Message Queue Debugging

- Filtering
  - Tags
  - MPI Communicators
- Cycle detection
  - Find deadlocks

The MPI vendor implements a supporting library defined by Etnus.
TotalView does not need to be attached to the entire job

- You can be attached to different subsets at different times through the run
  - You can attach to a subset, run till you see trouble and then 'fan out' to look at more processes if necessary.
- This greatly reduces overhead
  - There is a danger of missing things
• Sometimes env has optimal method for launching tasks across cluster

• Take advantage of that tech for
  • Faster Startup
  • Security

• File-> Preferences-> Bulk Launch

• See manual for details
Graphical stack trace
- Tell structure at a glance
- Filter based on processes and threads
Sometimes it is very helpful to look at the value of a variable in all MPI processes
- Dive on the variable
- Select the View > Laminate > Process command

TotalView creates an array indexed by process

You can filter and visualize a laminated array
Parallel Evaluate Expressions

- Expressions evaluated in parallel with control groups.
- Action Point Properties window showing conditions and expressions.
- Evaluate dialog boxes with expressions and results displayed.
Almost all CLI commands take scope into account

- Scope defined by a PT Set (Process/Thread Set)
- Be aware that some commands have behavior that interprets the PT Set differently, see individual command descriptions

Use dfocus CLI command

- dfocus <PT set> <command>
  - dfocus g215.2
dfocus p list
dgo

- Use it alone to set default scope
  - reflected in prompt
  - Prepend dfocus command to other commands to temporarily alter scope

GUI

Group -> Edit Group
Tools for the multicore age

- **HPC Software Tools** should be
  - Highly Scalable
  - Focused on concurrency
    - Multi-thread
    - Multi-process
  - Portable
  - HPC Aware
    - Play nicely in a typical HPC environment
  - Easy to use
Try it yourself!

- **Kick the tires**
  - Sign up for a 15 day trial at http://www.etnus.com/

- **Get more info**
  - Talk to us here at the show
  - Full documentation available online at http://www.etnus.com/
  - Watch a webcast at http://www.etnus.com/
    - Introduction to TotalView
    - Introduction to Memory Debugging
  - Contact Etnus at info@etnus.com

- **Get started debugging**
  - Pricing starts at $669 for TotalView Individual
  - Team and Enterprise pricing for groups of users