Achieving Order Through CHAOS: The LLNL HPC Linux Cluster Experience

Ryan L. Braby, Jim E. Garlick, Robin J. Goldstone

Production Linux Group

Lawrence Livermore National Laboratory

Science in the National Interest

Lawrence Livermore National Laboratory

Department of Energy
University of California

Lawrence Livermore National Laboratory ensures national security and applies science and technology to important problems of our time.

Background

Livermore Computing (LC) is one of the largest supercomputing facilities in the world, with over 43TF of production computing capacity.

System	Manufacturer Model	Interconnect	Nodes	CPUs	Memory (GB)	Peak GFLOP/s	Total GFLOP/s
Unclassified Network							
ASCI Linux Cluster	IBM	Elan3	960	1,920	3,840	9,216	
MCR	Linux NetworX	Elan3	1152	2,304	4,608	11,059	
ASCI Frost	IBM SP	Colony DS	68	1,088	1,088	1,632	
ASCI Blue-Pacific CTR	IBM SP	TB3	280	1,120	410	744	
Tera Cluster 2000 (TC2K)	Compaq SC ES40	Elan3	128	512	280	683	
iLX	RAND Federal	N/A	45	90	180	432	
GPS Cluster	Compaq GS320/ES45/DS20E	N/A	49	160	344	277	
Viz Engine (PVC)	Acme Micro	Elan3	64	128	128	614	
Viz Engine (Riptide)	SGI Onyx2	8 IR2 Pipes	1	48	16	24	
Qbert	Digital 8400	MC 1.5	2	20	24	25	**24,706**
Classified Network							
ASCI White	IBM SP	Colony DS	512	8,192	8,192	12,288	
ASCI Ice	IBM SP	Colony DS	28	448	448	672	
ASCI Blue-Pacific SST	IBM SP	TB3/HPGN	1464	5,856	2,628	3,888	
Sector S		TB3	488	1,952	1,164	1,296	
Sector K		TB3	488	1,952	732	1,296	
Sector Y		TB3	488	1,952	732	1,296	
PCRA (Adelie)	Linux NetworX	Elan3	128	256	256	870	
PCRB (Emperor)	Linux NetworX	Elan3	88	176	176	598	
Furnace Cluster	API CS20	N/A	64	128	128	213	
SC Cluster	Compaq ES40/ES45	N/A	40	160	384	235	
ICF Cluster	Compaq ES40/DS10L	N/A	12	36	12	48	
Tri-Lab Viz Engine (Whitecap)	SGI Onyx3800	4 IR3 Pipes	1	96	96	77	
Viz Engine (Tidalwave)	SGI Onyx2	16 IR2 Pipes	1	64	24	38	
Viz Engine (Edgewater)	SGI Onyx2	10 IR2 Pipes	1	40	18	24	**18,951**

Background (continued)

- LC's scalable systems strategy is known as the Livermore Model.
  - Provides a common programming and execution framework that:
    - Allows applications to be portable across multiple platforms
    - Remains stable over multiple generations of platforms.
  - Abstracts parallelism and IO services to a high level that evolves slowly.
  - Based on SMP compute nodes attached to a high-speed, low-latency interconnect.
  - Uses OpenMP to exploit SMP parallelism within a node and MPI to exploit parallelism between nodes.
  - Provides a POSIX interface parallel filesystem.
  - Application toolset: C, C++ and Fortran compilers, scalable MPI/OpenMP GUI debugger, performance analysis tools.
  - System management toolset: parallel cluster management tools, resource management, job scheduling, near-real-time accounting.

Measure of Success

- Our goal has been to demonstrate that we can design and deploy HPC Linux clusters using COTS hardware and Open Source software that:
  - Implement the Livermore Model
  - Deliver required performance
  - Have outstanding RAS characteristics
  - Are cost effective
- By having effective hardware and software strategies, leveraging off our existing HPC expertise and paying careful attention to detail, we have successfully met this goal.

Hardware Considerations

- Hardware component choices will affect supportability and ease of system management.
  - Node components: motherboard, chipset, processors, memory, IDE vs. SCSI hard drives, form factor.
  - Message passing interconnect
  - Remote power and console management
- Integration considerations
  - Racking, cabling and labeling
  - BIOS/CMOS settings
  - Burn-in
- Hardware support strategy
  - Self-maintenance vs. vendor maintenance
  - Hot spare cluster
  - Spare parts cache

Software Considerations

- LLNL CHAOS strategy
- The CHAOS kernel
- Cluster System Management Tools
- Monitoring and Failure Detection
- Other software components
  - Resource Management
  - Authentication and Access Control
  - Parallel Filesystem
  - Application Development Environment
- Software Support Strategy

CHAOS Strategy

- CHAOS = Clustered High Availability Operating System
- Augment standard Linux (Red Hat) distribution with support for HPC clusters including:
  - Scalable system management and monitoring tools
  - High performance interconnect (Quadrics Elan3)
  - Lustre parallel filesystem
  - Resource Management and Control (SLURM)
- Provide the ability to integrate new hardware at the earliest opportunity.
- Leverage in-house development team to provide rapid turnaround on bug reports and requests for feature enhancements.
- Keep kernel mods to a minimum; leverage relationship with RedHat to get our patches accepted into their distribution.
- Provide a framework for release management.
- Deploy production-quality HPC Linux clusters

The CHAOS Kernel

Currently based on RedHat 2.4.18-20 kernel with the following mods.

- ecc module
  - i860, E7500, E7501 chipset support
  - Panic on uncorrectable error, log correctable errors
  - Support for bit scrubbing
- mtd driver – support for motherboard flash device
- Quadrics support: Elan device driver, IP over Elan module, RMS module, core path patch, Totalview ptrace patch
- p4therm module – panic node if Pentium4 thermal sensor is tripped
- Crash dump facility: Currently mcore, pursuing transition to RedHat netdump
- Lustre support: VFS intents, Read-only devices, Zero copy networking enhancements for portals, Lustre and portals kernel modules (not yet integrated)
- Updated qlogic HBA adapter support
- Workaround for e1000 lockup bug
- Hardware performance counters
- Raise _FD_SETSIZE from 1024 to 8192
- Nfs groups > 16
- statfs64

Cluster Tools

LLNL has developed a number of feature-rich, scalable system management tools including:

- YACI: Yet Another Cluster Installer
- pdsh: Parallel Distributed Shell
- Genders framework for configuration management
- Powerman remote power control software
- Conman remote console management software
- HM: host monitoring framework

Software Support Strategy

- Problems are reported by users to LC's User Services department and tracked through Remedy trouble ticket system.
- System administrators perform first level problem determination.
- If a software defect is suspected, problem is reported to CHAOS development team.
- Bugs are entered and tracked in CHAOS GNATS database.
- Fixes are pursued as appropriate:
  - In Open Source community
  - Through RedHat on-site analyst
  - Through vendor partners (Quadrics, CFS, etc.)
  - Locally by CHAOS developers
- Bug fix is incorporated into CHAOS CVS source tree and new RPM(s) are built.
- Changes are rolled out in next CHAOS release, or sooner depending on severity.

Conclusions

- Lessons Learned
  - Not all COTS hardware is created equal.
  - Integration: the devil is in the details.
- System Manageability
  - Linux Clusters versus vendor proprietary systems
  - Two Linux clusters compared: MCR and ALC
- Total Cost of Ownership
  - Software: Invest in local Linux development staff rather than vendor software support contracts.
  - Hardware: High MTBF of commodity hardware combined with self-maintenance saves considerable cost over vendor hardware maintenance.

Future Directions

- Parallel Filesystem (Lustre)
- Continued development of CHAOS and associated clustering tools
- New processor architectures
  - IA64
  - Opteron
- New cluster interconnects
  - Elan4
  - Infiniband
- Serial ATA RAID
- Diskless configurations

For more info…

www.llnl.gov/linux

Linux@ Livermore

[Production Clusters]
[Linux Projects]
[Project Reports]
[Software Downloads]
[News and Events]