The HEidelberg LInux Cluster System

one of the most powerful PC−Clusters worldwide

www.helics.de

Key facts

·256 Double Athlon MP 1,4 GHz (512 Processors total), ATX Midi−Tower
·Chipset AMD 760 MPX, Motherboard Tyan Tiger MPX
·256 GByte aggregated system RAM, 10 TByte local disk space
·Myrinet 2000 data network (fiber) using 6 128−port switches in a full bisection Clos network
·Cascaded Fast−/Gigabit Ethernet network for administrative and filesystem communication
·Scalable, fully redundant central Filesystem 2.2 TByte − ?
·theoretical peak performance: 512x2.8 GFlop/s = 1.43 TFlop/s
·current measured performance: 825 GFlop/s Linpack ~ Top500 list position 24
·data network: ~ 20 my.sec latency (round trip), 128 x 2 Gbit/s full bisectional bandwidth
·1.26 Mil. EUR total costs
·shall cover all IWR applications
·price/performance ratio: 1500 EUR/Gflop/s (linpack)

MEGWARE COMPUTER
Halle 2/ Stand A47

Myricom

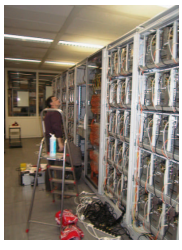
TYAN

AMD

Installation

test installation in Chemnitz at MEGWARE

delivery in Heidelberg

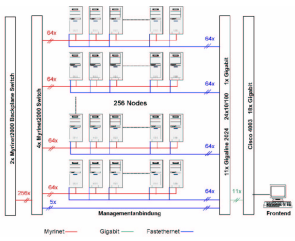


assembly through MEGWARE

Applications

·IWR (Interdisciplinary Center for Scientific Computing): founded in 1986 as an institution of the University of Heidelberg, Germany [www.iwr.uni−heidelberg.de]
·groups from different fields of research share common system resources and knowledge in the field of computer science and applied mathematics
·fields of research include:
  ·Biocomputing
  ·Bio−Medicine
  ·Discrete Optimization
  ·Reactive Flows and Combustions
  ·Technical Informatics
  ·Technical Simulations
  ·Digital Image Processing / Visualization
·−> use of models based on parallel systems
·−> need for adequate platforms
·−> need for adequate software (message−passing interfaces)
·−> need for specific development tools
·The system resides inside the HWW and therefore is accessible for commercial applications

Implementation Details



- the System is hidden behind a private network: file−, master− and frontend hosts are directly connected to the gigabit back−bone. User access is only guaranteed through a set of well firewalled access nodes that are multihomed
- each rack is equipped with an selfengineered powercycling unit; these units communicate using a redundant serial bus with the master server; this way it is possible to powercycle/reset nodes remotely
- nodes are configured to boot from network under any condition; receive their network configuration and booting instructions using dhcp
- depending on different criteria (version numbering ...) or on manual request the system gets installed automatically over network or boots up in running mode using the local system. Alternatively an image for updating firmware (BIOS..) is available.
- Debian GNU/Linux (www.debian.org) as the most stable Linux distribution together with own enhancements is used as the native OS on all x86 components.
- Using SCore 4.x (www.pccluster.org) implementing the parallel environment provides technologies only known from traditional supercomputers
  - Single System Image
  - checkpointing/restarting of jobs, multinode redundancy
  - highly performant communication library/device interface (PM)
  - efficient resource allocation using advanced scheduling procedures
- Own implementation for system monitoring (hardware + software health control)

Dr. Wolfgang Hafemann
IWR, Universität Heidelberg
Im Neuenheimer Feld 368
D−69120 Heidelberg
Tel.+49 06221 548240, Fax 545224
e−mail: Wolfgang.Hafemann@iwr.uni−heidelberg.de
helics@uni−hd.de
http://www.helics.de