PDC Newsletter
No 1 - 2004

The New Itanium2 Cluster

This summer, PDC installed a new computer system, an HP Intel Itanium2 cluster composed of 90 dual-processor nodes for a total of 180 processors as was briefly mentioned in PDC Newsletter 1 – 2003. This new cluster follows the tradition at PDC to provide systems with large memories and high memory bandwidth. The system runs the Linux operating system and PDC has installed the utilities and applications familiar to PDC users, making this an easy-to-use environment for scientific research. See readme on page 5. The new computer resource will replace the Power2SC and PowerPC part of the IBM SP. Allocations granted on the previous resources (IBM SP) has been transferred to the new improved system. In this issue of the PDC Newsletter the HP Intel Itanium2 cluster is presented in more detail.

Photo: Harald Barth

The Itanium2 Cluster at PDC

Intel's EPIC Technology

by Tommy Rydendahl, Intel:

The need for faster, more affordable computing solutions has never been greater. In both technical computing and enterprise IT environments, demands are rising dramatically as individuals and organizations seek to understand and control increasingly complex processes.

For over a decade, Intel-based platforms have delivered higher levels of performance and compatibility at more affordable prices than competing RISC-based products. Explicitly Parallel Instruction Computing (EPIC) was designed to break through the limitations of traditional architectures and to enable many years of cost-effective performance scaling for high-end applications. With the release of the Intel Itanium2 processor – the second generation in the Intel Itanium processor family – Intel estimates that businesses can purchase platforms that deliver 50% higher transaction processing performance than comparable platforms and at lower costs.

Companies that are already running applications compiled for Itanium architecture will experience a 1.5 – 2 times performance boost on Itanium2-based platforms. The Itanium2 processor is designed to be hardware and software compatible with future Intel Itanium processors. Organizations will be able to upgrade their platforms with the next-generation processors.

These performance improvements demonstrate the scalability of the EPIC model. By establishing an environment conducive to massive parallel instruction throughput, EPIC overcomes the limitations of traditional architectures, and paves the way for faster and more cost-effective performance scaling in the future.

Continued on page 3.

Lennart Johnsson
Director of PDC

In This Issue:

PDC Newsletter No 1 – 2004

Published by PDC at KTH.

PDC operates leading-edge, high-performance computing systems as easily accessible national resources and carries out research in software and tools for use and administration of such systems. The hardware resources and the costs of their operations are largely covered by funding from the Swedish Research Council and KTH. The research is funded from a variety of sources with the majority of research funds coming from the European Commission.

Publisher: Lennart Johnsson

Editors: Mats Höjeberg, Gert Svensson

Layout: Maria Engström

E-mail: pdc-newsletter@pdc.kth.se

ISSN 1401-9671

To Contact PDC

Visiting address: Teknikringen 14, plan 4, KTH, Stockholm

Mail: PDC, KTH, SE-100 44, Stockholm, Sweden

E-mail: pdc-staff@pdc.kth.se

WWW: http://www.pdc.kth.se/

Phone: +46 8 790 78 00

Fax: +46 8 24 77 84

The Evolving HPC Infrastructures

We are very pleased that our Itanium2 Linux cluster now is in production. We are confident that this new resource will further enhance our ability to serve our users. Further, in December, the Swedish Research Council made available funds for additional enrichment of PDC's computational resources. We expect these to be in service by the summer.

Itanium based clusters are now cornerstones of high performance computing infrastructures around the world. They are at the core of the US Distributed Terascale Facility (DTF) funded by the National Science Foundation. Itaniums are used in the US Department of Energy's largest non-classified system, the 11.8 TFlop cluster at the Pacific Northwest National Laboratory, and are at center stage at CERN's Open Lab that carries out research in next generation platforms and software for high-energy physics applications. In addition to the Itanium's processing advantages (up to 4 times faster than IA-32 systems for some applications at PDC) the I/O capabilities of our Itanium cluster also compares favorably with IA-32 systems. The memory and I/O capabilities of HP's Itanium platforms were important factors in our decision to invest in this architecture.

The importance of I/O is in part driven by many sciences becoming increasingly data centric. The storage needs of the science community are expected to grow ten-fold every five years for at least 15 years. CERN's Large Hadron Collider, used by about 2,000 scientists worldwide, is expected to produce 5 – 10 PB/yr beginning late 2006. The Large Synoptic Survey Telescope is expected to produce over 10 PB/yr starting in 2008. Similar data volumes are expected in the Earth Sciences. The Life Sciences and medicine will produce even greater amounts of data. Digital mammography alone is expected to produce about 10 PB/yr.

The increased emphasis on data and data sharing is also a driving force for backbone networks. Today, GigaSunet, Nordunet, the Dutch Surfnet, UK's JANET, the European GEANT, the US Abilene network, and several others operate at 10 Gbps. Multiple 10 Gbps links interconnect these networks and enable efficient sharing of data and other resources. For instance, last October 1.1 TB was transferred 7607 km in 27 minutes from CERN (using Itaniums) to Chicago. The now emerging λ-networks offer improved quality-of-service and user control through end-to-end lightpaths. Nordunet's NorthernLight provides connectivity from KTHNOC to the global TeraLight testbed. PDC and the Karolinska Institute were the first to use this capability last August for tele-science. Future issues will cover these exciting developments and the science they enable, PDC's new facilities and SweGrid. In this issue you will find articles on the Itanium and storage systems.

...continued from page 1

Challenges in High-performance Technical Computing

In virtually every field of science and engineering, research and development teams are searching for new quantitative precision, and trying to model increasingly complex systems. Not surprisingly, this is one of the first market segments to begin embracing the power and affordability of high-performance methods, such as the Itanium-based solutions.

One example is the Distributed Terascale Facility currently under development by the National Science Foundation, where the total system will be capable of 13.6 trillion calculations per second, and able to work with 450 terabytes of data.

Another example is the high-performance computer at the Department of Energy's Pacific Northwest National Laboratory (PNNL). With 1,976 Intel Itanium2 processors, 6.8 terabytes of memory and 500 terabytes of disk space, this new system has a peak capacity of 11.8 trillion floating point operations per second and ranks as number five on the November 2003 edition of the Top 500 ranking. It achieved an efficiency in excess of 73% for the Linpack benchmark on which the ranking is based.

Challenges in Enterprise Computing

Computing needs are also mounting in the commercial sector. Companies in many industries are replacing time-consuming prototype development with computer modeling and design solutions. These computing intensive applications provide a significant payoff in accelerating time-to-market, and are steadily increasing in complexity, accuracy and sophistication.

Complex business applications, such as Large Databases (LDB), Enterprise Resource Planning (ERP), Supply Chain Management (SCM), and Business Intelligence (data mining) are also demanding more compute power, as companies link with more users, automate complex processes, and work to serve their customers more effectively. Integration is driving up workloads and bringing increasing quantities of data into the enterprises. These business demands come in addition to traditional applications.

A Bold New Approach

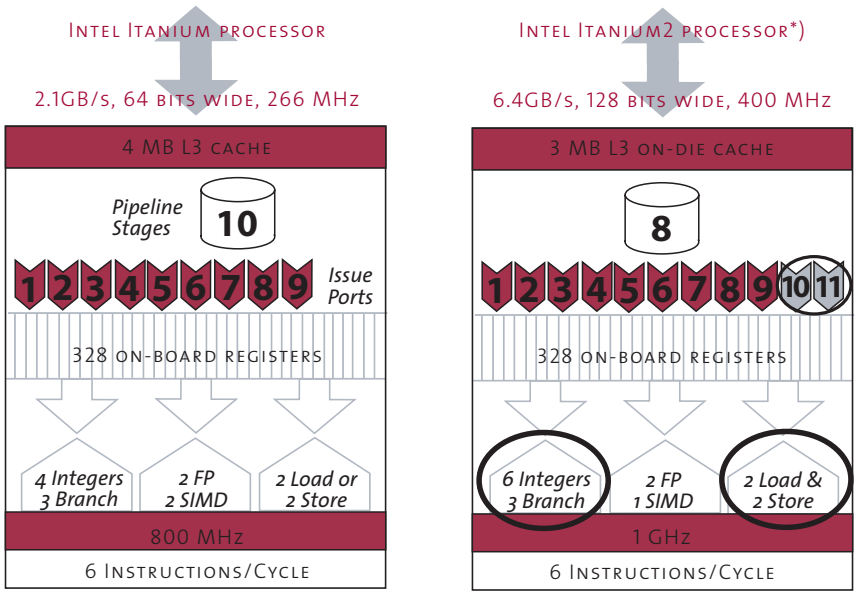
The EPIC computing model was specifically designed for highly efficient parallelism, which is the ability to process multiple instructions or processes simultaneously. Parallelism increases the amount of productive work that can be accomplished during each processor clock cycle, and can greatly accelerate application performance. By establishing a foundation for enhanced parallelism, EPIC enables Intel to scale processor performance by improving parallel instruction throughput. This approach has already achieved industry-leading performance in key application categories.

In other computing models, parallel-processing opportunities must be identified by the processor itself. EPIC includes an enhanced instruction set that allows parallel processing opportunities to be explicitly identified by the compiler before the software code reaches the processor. The compiler can view and analyze the entire code to determine the most efficient strategy for parallel processing. It re-shapes the program to optimize efficiency. The processor simply processes the instructions in parallel as rapidly as possible. This division of labor not only delivers

The Intel Itanium2 processor builds on the EPIC foundation to provide significant performance benefits over the first-generation Intel Itanium processor. These performance improvements include:

- 3 times increase systembus bandwidth
- Large integrated cache with reduced latency
- Additional issue ports
- Additional execution units
- Increased core frequency
- Compatible with Intel Itanium processor software

*) The 900 MHz processors in Lucidor have 1.5 MB L3 cache



immediate performance benefits, it also opens up considerable opportunities for future performance scaling.

On the software side, compilers will become increasingly advanced in optimizing code for parallel execution. On the hardware side, development efforts will continue to focus on increasing the number of instructions that can be processed during each clock cycle.

A Comprehensive Solution for High-end Computing

Explicit parallelism is just one of the advantages of the EPIC computing model. EPIC was designed to address the most pressing challenges of high-end computing, including performance, scalability, reliability and manageability. The Intel Itanium processor was built on this foundation to provide a powerful, flexible and open architecture. The Intel Itanium2 processor built on the same foundation provides significant performance benefits over the "first generation" Intel Itanium processor (Figure on page 3). Future Itanium processor generations will continue this trend, providing regular upgrade opportunities for existing Itanium architecture-based solutions.

Explicit Parallelism

The ability to extract a high degree of parallelism is the key to EPIC's computational efficiency. Current Intel Itanium processors can process up to 6 simultaneous instructions and EPIC has the flexibility for increased parallelism in future processor generations. The number of simultaneous instructions is not the only measure of application performance. The processor must also be able to sustain high levels of parallelism to optimize total throughput. A variety of features make the EPIC computing model well suited for this task, including extensive computational resources, and enhanced predication and speculation (see below).

The Intel Itanium2 processor can sustain higher parallel instruction throughput. It can process more 6-instruction combinations than the first generation Itanium processor, and keep those instructions moving more rapidly through the computing pipe-line.

Predication

Most software code contains many conditional branches such as if/then, in which the result of the operation determines what the processor should do next. Conditional branches can prevent the processor from moving forward until the conditional statement is processed.This can become a major limitation to overall throughput. In the EPIC computing model, predication allows the compiler to explicitly identify instruction streams that can be processed in parallel. It also allows the processor to pre-load instructions and data and begin processing for both branches simultaneously. Once the conditional statement is processed, the information gathered for the incorrect path is simply discarded. This enables the processor to move beyond a conditional branch without waiting for it to be resolved, which can significantly improve parallel processing efficiency.

The Itanium2 processor discards incorrect results and resumes productive compute cycles more efficiently than the first-generation Itanium processor. This improves total throughput.

Speculation

Fast processing speeds are of limited value if computational units sit idle while the processor retrieves required data from memory. Speculation allows the compiler to identify future data needs, so essential data can be pre-loaded into the processor. This technique can significantly reduce or eliminate processor wait times.

The Itanium2 processor discards incorrect results and resumes productive compute cycles more efficiently than the first-generation Itanium processor. This enables the predication feature to be used more aggressively and efficiently to accelerate total throughput.

Conclusion

With the release of the Intel Itanium2 processor, the Intel Itanium architecture and the EPIC computing model are moving into the forefront of high-end computing. Customers will have increasing access to a variety of compatible, best-of-breed solutions at competitive prices. Just as Intel's 32-bit computing architecture set the standard in the entry-level and midrange server market segments, the Itanium architecture will drive increased performance, value and choice into the high-end computing arena.

Further reading:
http://www.intel.com/eBusiness/products/itanium or
http://www.intel.com/itanium2

Lasse Lucidor

Lasse Lucidor is the most common pseudonym used by the 17th century Swedish poet Lars Johansson. He sometimes also signed his work Blumino. He was a vagabond but settled in Stockholm in 1669 and soon became the most called for contemporary poet. His poems were only occasionally published during his lifetime. He was killed in a scuffle in 1674.

Illustration from:
KB's, The Royal Library, collections.

Lucidor

The new Linux cluster installed at PDC during this summer, as mentioned in Newsletter 1 – 2003, is named Lucidor. Lucidor is a distributed memory computer (a cluster) from HP. It consists of 74 HP rx2600 servers and 16 HP zx6000 workstations, each with two 900MHz Itanium2 "McKinley" processors and 6 GB of main memory. The rx2600 and zx6000 nodes differ only in the configuration of the memory banks (although they both have the same total amount of main memory) and in that the zx6000 has a PCI-X slot replaced with an AGP slot with a graphics card.

The interconnect is Myrinet. All nodes have a Myricom M3F-PCIXD-2 card (64-bit wide 133 MHz PCI-X). All cards are connected to a Myrinet-2000 M3-E128 switch populated with 96 ports. Each card/port has a data rate of 2+2 Gbit/s, all through 50/125 multi-mode fiber.

You can read more about the file systems and disk usage policy in http://www.pdc.kth.se/compresc/.

Some node performance numbers

Processors per node	2
Number of nodes	90
Peak floating point capacity	7.2 Gflop/s
Memory Bandwidth	6.4 GByte/s
Local Disk I/O Bandwidth	58 MByte/s (measured writes)
Network Adapter Peak	
Bandwidth (unidirectional)	248 MByte/s
Bandwidth (bidirectional)	489 MByte/s
Network Latency	6.3 microseconds (measured)

Introduction to using the new PDC Cluster Lucidor

The new Linux cluster at PDC, Lucidor, is now fully operational. Lucidor consists of 90 dual-CPU nodes with 6 Gbytes memory/node. 74 nodes are HP rx2600 servers and 16 nodes are HP zx6000 workstations. The CPUs are Intel Itanium2 ("McKinley") running at 900 MHz.

In order to be able to logon to Lucidor you need Kerberos v5 enabled software, for example Heimdal. New Kerberos v5 enabled travelkits are available at the PDC homepage. If your desktop operating environment is not available, contact PDC staff with information about your system so that we can provide the necessary software for you. The login node of Lucidor is called blumino.pdc.kth.se. As always, the login node is intended for compilation and submission of jobs, not for execution of jobs.

Available compilers are the Intel compiler suite and of course gcc. Version 3.2 of gcc is available through the module gcc. The Intel compilers are available through the module i-compilers. Note that you need to have tickets before loading this module with `module add i-compilers`.

For parallel programs, the module mpich is needed. Compile and link with the commands mpif90 or mpicc. Execution of code is demonstrated in an example session below:

Running an MPI code on interactive nodes:

```
> module add heimdal
> kinit -f
> module add i-compilers mpich easy
> spattach -i -p3
> mpif90 -o example -O2 example.f90
> mpirun -np $SP_PROCS -machinefile $SP_HOSTFILE ./example
```

Running an MPI code in batch mode (dedicated nodes):

```
> cp /afs/pdc.kth.se/misc/pdc/mpich/mpich.lxl
> esubmit -n3 -t15 -c MyUserCAC ./mpich.lxl ./example
> esubmit -n3 -t15 -c MyUserCAC ./mpich.lxl -p 2 ./example
```

where the latter esubmit gives two processes on each node, i.e. a total of six processes.

The scheduler, EASY, is similar to the ones on PDC's other systems. There are some minor differences, most notably that the submit command is called esubmit instead of spsubmit. EASY is accessed through the module easy. All EASY commands have a -h option for help. The zx6000 workstations are referred to as A-nodes and the rx2600 servers are referred to as B-nodes. The A-nodes have a slightly different memory configuration (although they have the same amount of memory as the others) which may translate into a slight performance penalty (a few percent) for some codes.

Available tools include the debugger Totalview and the PDC developed profiling tool i2.prof, which is a perl script based on the pfmon tool. Both pfmon and i2.prof are available through the module perftools. The profiling tool qprof is also available in this module. Furthermore the performance counter API library (PAPI) is available. We also have the GNU profiler gprof. The present version (version 7) of the Intel compilers does not support generation of a call-graph profile with gprof. However, a flat profile may be generated.

BLAS and LAPACK routines are available in the Intel Math Kernel Library (MKL). It is accessed through the module mkl. A high performance and highly accurate vector math library is also installed, namely the HP Vector Math Library (VML). It is accessed through the module vml. The ScaLAPACK parallel linear algebra library and ARPACK are also installed.

The following applications are installed: Gaussian, NWChem, Gamess, Dalton, Jaguar, CHARMM and ABAQUS. There are also modules for Metis, parMETIS, Foresys, FFTW and netCDF.

Detailed information on Lucidor can be found on the PDC webpages, http://www.pdc.kth.se/compresc/hardware.html

More Students Dive into HPC

by Mike Hammill

More students than ever took the "Introduction to High-Performance Computing" course offered at PDC this summer. 69 students, which is 23% more than last year and a full 50% more than the year before, participated. The course has been given at PDC every summer since 1996.

The course was special this year not just because of the record attendance. It was also the first time a group of students used PDC's new HP Intel Itanium2 cluster. In fact, prior to the class, the only users who had access to the system were a relatively small set of test users. Being at the leading edge of high-performance computing is what the class is all about.

The students get both a strong conceptual foundation in state-of-the-art HPC as well as experience with practical aspects through lab work and projects. They gain valuable knowledge that will help them with their research and PDC benefits by encouraging more efficient use of its resources by the next generation of its users.

A seasoned cadre of professors and professionals from around Sweden are joined by world-famous practitioners to cover the topics in the course. The topics include parallel programming (OpenMP, MPI), modern computer architectures, parallel algorithms, efficient programming, and case studies. These topics are covered by such well-known leaders in the field as Björn Engquist (Princeton/KTH), Lennart Johnsson (University of Houston/KTH), Mats Brorsson (KTH), Thomas Ericsson (Chalmers), and Steve Tuecke (Argonne National Laboratory).

Each student is expected to complete a project, which is often related to their research area. This year's projects range from "FDTD code for solving the electromagnetic field in the inner ear of a mobile phone user's head" to "Molecular dynamics method for docking ligands to receptors" with everything in between.

To help them complete this work, a tutor is assigned. In addition to helping answer questions the students have, the tutor reviews the work in conjunction with the course examiner, Jesper Oppelstrup, a numerical analysis professor at KTH.

What do the students think? A couple of quotes from this year's course evaluation: "The course is very important for my work. I am happy that this course exists." And "Well organized and structured. Good with a basic introduction and then more special fields!"

The course, which is given the last two weeks in August, is within the National Graduate School in Scientific Computing (NGSSC), funded by the Swedish Foundation for Strategic Research. It is also open to KTH masters students as well as Ph.D. students world-wide.

For more information, see the course Web page at: http://www.pdc.kth.se/training/2003/SummerSchool/.

Photo: Mike Hammill

More students than ever took the "Introduction to High-Performance Computing" course offered at PDC this summer.

SBC Cluster Upgrade

by Daniel Ahlin

In April the SBC (Stockholm Bioinformatics Center) Calculation Cluster that is managed and operated by PDC was upgraded with 112 new computation nodes from South Pole AB. Each new node is equipped with one AMD Athlon XP 2700+ CPU, 1 Gbyte RAM and 40 Gbyte HDD.

The nodes were delivered to PDC in three batches. Within six hours after delivery 111 of the 112 nodes were available for production use. After the upgrade the SBC cluster consists of 211 nodes of which 204 are used for computation.

The co-operation between PDC and SBC was a major theme in the previous PDC Newsletter.

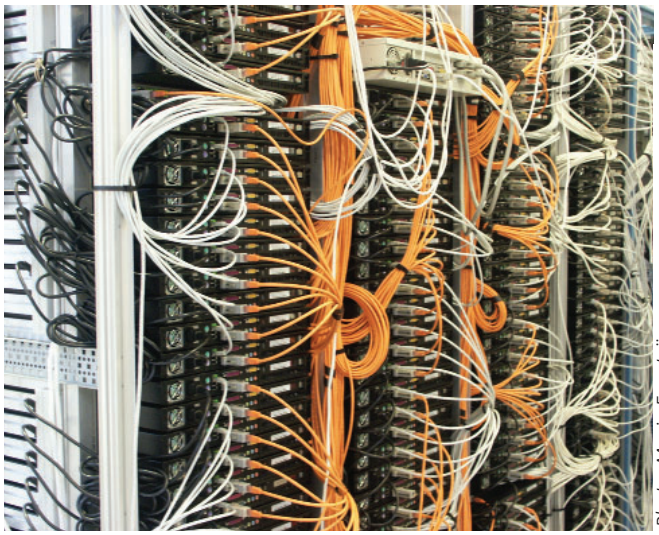


Photo: Maria Engström

Wiring for the network in the SBC Calculation Cluster.

Calendar

• February 25-27, 2004, San Francisco, CA, USA: SIAM Conference on Parallel Processing and Scientific Computing (PP04) http://www.siam.org/meetings/pp04/index.htm

• March 14-17, 2004, Nicosia, Cyprus: SAC 2004–19th ACM Symposium on Applied Computing;http://www.acm.org/conferences/sac/sac2004/

• March 14-17, 2004, Orlando, FL, USA: PerCom 2004–Second IEEE International Conference on Pervasive Computing and Communications; http://www.percom.org/

• March 18, 2004, Uppsala, Sweden: Inauguration of SweGrid http://www.swegrid.se/

• April 12-15, 2004, Chicago, IL, USA: GLOBAL and PEER-TO-PEER Computing "From Theory to Practice"; http://www.lri.fr/~fci/GP2PC-04.htm

• April 14-16, 2004, Ischia, Italy: CF '04, 2004 ACM International Conference on Computing Frontiers; http://www.computingfrontiers.org/

• April 19-22, 2004, Chicago, IL, USA: 4th International Symposium on Cluster Computing and the Grid; http://www.mcs.anl.gov/ccgrid2004/

• April 26-30, 2004, Santa Fe, NM, USA: IPDPS 2004 –International Parallel and Distributed Processing Symposium; http://www.ipdps.org/
A number of workshops will be held in conjunction with IPDPS:
- HiCOMB, 3rd International Workshop On High Performance Computational Biology (http://www.hicomb.org/)
- PACGrid-04, The First Workshop on Partitioning Applications for Computational Grids
- HIPS 2004, 9th International Workshop on High-Level Parallel Programming Models and Supportive Environments
- CAC'04, Workshop on Communication Architecture for Clusters
- PMEO-PDS'2004, 3rd International Workshop on Performance Modeling, Evaluation, and Optimization of Parallel and Distributed Systems
http://www.dcs.gla.ac.uk/people/personal/mohamed/pmeo04.html

• May 10-12, 2004, Hong Kong, SAR, China: I-SPAN 2004–The 7th International Symposium on Parallel Architectures, Algorithms and Networks; http://www.csis.hku.hk/ispan2004/

• May 14-17, 2004, Assisi, Italy: ICCSA 2004–2004 International Conference on Computational Science and Its Applications http://www.iccsa2004.unipg.it/

• May 16-19, 2004, Kufstein, Austria: PADS 2004–18th Workshop on Parallel and Distributed Simulation http://www.pads-workshop.org/pads2004/

• June 2-3, 2004, Stockholm: Swedish National Infrastructure for Computing, SNIC Interaction 2004 at KTH PDC http://www.pdc.kth.se/

• June 6-9, 2004, Kraków, Poland: ICCS 2004, International Conference on Computational Science http://www.cyfronet.krakow.pl/iccs2004/

• June 20-23, 2004, Copenhagen, Denmark: PARA04, Workshop on State-of-the-Art in Scientific Computing; http://imm.dtu.dk/~jw/para04/

• June 22-25, 2004, Heidelberg, Germany: ISC2004–19th International Supercomputer Conference; http://www.isc2004.org/

• June 28-30, 2004, Valencia, Spain: VECPAR'2004–6th International Meeting on HPC for Computational Science; http://vecpar.fe.up.pt/2004/

• July 11-16, 2004, Portland, OR, USA: SIAM Annual Meeting 2004 and Conference on the Life Sciences; http://www.siam.org/meetings/an04/

• July 24-28, 2004, Jyväskylä, Finland: ECCOMAS 2004–European Congress on Computational Methods in Applied Sciences and Engineering; http://www.mit.jyu.fi/eccomas2004/

• July 26-30, 2004, Leuven, Belgium: ICCAM-2004–Eleventh International Congress on Computational and Applied Mathematics http://www.cs.kuleuven.ac.be/conference/iccam2004/iccam.htm

readme

A crash course in module usage

On the new cluster Lucidor as well as most otherPDC systems, much of the software is accessed through modules. Hence, we here give a short description of the most important module commands.
To load a module do `module add <modulename>.`
To see which modules you have loaded do `module list.`
To remove a module do `module rm <modulename>`
To list the available modules do `module avail.`
To list the available mpich modules do `module avail mpich`
To see what loading a module achieves do
`module show <modulename>`

B
Porto
betalt

ROYAL INSTITUTE
OF TECHNOLOGY

PDC Newsletter
No 1 – 2004

Long-term Mass Data Storage

by Per Ekman

There is a sometimes bewildering number of data storage options available on the various computing resources at PDC. Fast local scratch disk, parallel filesystems, AFS and, lurking at the back, the HSM system.

HSM stands for Hierarchical Storage Management and is essentially a filesystem interface to tape storage. The way it works is that files are stored in the HSM filesystem. When the filesystem is full the oldest files are migrated to tape. The file is still visible in the filesystem and can be accessed as any normal file with the caveat that it may take a while for the system to retrieve the data from tape. The chief benefit of an HSM system is that it provides an easy way to use tape storage resources. Users simply manipulate files as usual and the system takes care of the rest.

The HSM system is intended for archival of large data sets. The storage is permanent, there is no automatic cleanup of old files and all data is backed up. The current usage guidelines say that a user may use up to 50GB of storage and have up to 1000 files in the HSM system. Users who need to store more than 50GB should email a request for more space to pdc-staff@pdc.kth.se.

At PDC the HSM filesystem is mounted on the server hsm.pdc.kth.se. All PDC users have a directory under /hsm/home on that machine and can login to the server and manipulate the stored files using normal UNIX commands. Files can be transferred to/from the HSM system using kerberized ftp/rcp or using the hsm* commands that are available on all major computing systems at PDC through the "hsm" module. The hsm* commands allow easy access to permanent mass data storage. They are described in the HSM Howto (http://www.pdc.kth.se/info/hsm/hsm-howto.html).

The HSM system is set up so that files smaller than 64kB are kept on disk all the time while larger files are eligible for migration to tape. For performance reasons the HSM system should only be used with large files. If many files that are smaller than 64kB are stored they will fill up the disk since they are not migrated. Recalling many small files that have been migrated can take a long time since they can be spread out over many tapes and mounting and seeking on tapes are very slow operations. If many small files need to be stored the best way is to store them in a tar-archive and then put the archive into the HSM filesystem.

Photo: Harald Barth

The HSM system

The HSM system at PDC is based on the DMF software from SGI. The system runs on a Origin300 server with a 137GB disk cache. The server is connected to an IBM 3494 tape library (or robot). The tape library has a maximum capacity of 941 tapes where each tape has room for 40GB of uncompressed data which gives a total capacity of 37TB. The tape library contains four IBM 3590E tape drives that are shared between the HSM system and backup services. A single drive has a nominal transfer rate of 13.5MB/s of uncompressed data.

The IBM 3494 tape library could in its original configuration hold 541 tapes. Each tape could hold 10GB of uncompressed data for a total capacity of 5.4TB. Since then the drives have been upgraded to the 3590 E-version, doubling the capacity of each tape by writing twice as many tracks on them. In 1999 the tape library was expanded with an extra cabinet increasing the total number of tape slots to the current 941.