



hepsysmanix

HEPiX report for hepsysman

RAL, 10 May 2006

Jens G Jensen
e-Science Centre

- SRM is a control protocol
 - Version 1.1 and version 2.1
- One head node, many pool nodes
- Data transfer
 - WAN protocol: GridFTP
 - LAN protocol: RFIO or DCAP
- Implementations:
 - dCache from DESY, its SRM from FNAL
 - DPM from CERN/LCG
 - StoRM from INFN (SRM 2 only)

- XFS improves SE performance
 - See Greig’s talk
 - Support issues
- Tape vs disk
 - “Adapt to customer demand more easily with tape” (FNAL)
 - Easy to expand – cost of media
 - HSMs can facilitate utilisation
 - Performs differently, mechanical

- 32 bit vs 64 bit (FNAL)
 - 32 bit OS, 32 bit apps, on 64 bit proc.
 - 64 bits OS/drivers, 32 bit apps, on 64 bit proc.
 - 4-15 % *slower* than 32/32
 - 64 bit OS, 64 bit apps, on 64 bit proc.
 - 20-40% faster than 32/32
- Extra features un(der)used – CERN
 - Security, SSE, power saving,...

- Dual core
 - Good performance/price – FZK
 - Almost same power consumption – FNAL
- More, more
 - Quad cores early '07?
 - Octo cores '09?
- Fast parallelism, IPC
 - Debugging, algorithms, paradigms
 - <http://www.multicore-association.org/>

- Maximise computing power, minimise electrical power
- IN2P3 tests
 - Opteron > Xeon
 - Dual cores > hyperthreading
 - Big names better
 - Blade systems better

- Optimised by compiler or programmer?
- Platform and compiler (and language and versions) dependencies
 - How to set the right options (easy-ish),
 - Improve algorithms,
 - Performance Analysis (correctness, regression test)
 - Understand machine code, cache, execution units,...
- Inlining: speedup, but code dup; cache misses can slow down
- Tools to diagnose – premature opt'n: 80-20 rule; diminishing returns: 90-90 rule

- SL: FNAL-repackaged RedHat
 - <https://www.scientificlinux.org/>
- SLC: SL CERN
 - With added CERN bits (e.g. kernel)
 - XFS support
- Version 3 vs version 4
 - 3.0.x widely used in LCG...
 - ...but its days are numbered
 - ...there should be a 3.0.8, at least

- XFS support
 - Not upstream, but *maybe(?)* in SLC
 - Cannot use XFS for pools now, need to upgrade to SL4 later...
- 64 bit processors
 - Likely to support x86_64 but not IA64
 - SLC might support IA64?

- MonAMI – Glasgow
 - Overview, Alerts, Gfx, Logs, Details
 - Target → Sensor → Display
 - MonAMI universal sensor framework

- ext3, XFS, JFS, ReiserFS
- Blocks and inodes
 - Can be clumsy for large files
 - Extents can prevent fragmentation
- Performance
 - Much depends on controller and disk
 - PCs bad at mixing R/W IO, and concurrent writes
 - See also Greig's talk

Journaling Filesystems in RHEL4

	SLOC	Incl. since	Extents	Status
EXT3	10K	2000	No	Widely used
Reiser v3	30K	2001	No	Small files. Fragile?
XFS	100K	2001	Yes	Large files. Widely used
JFS	30K	2002	No	Not widely used, slow devel.

- NFS
 - Slow for large files
 - Cluster only for non-security
- AFS
 - Global
- GPFS
 - Commercial
 - Fast, striping data across disks
 - Defective or malicious client can disrupt the system...

- SAN filesystems...
 - Fast, but infrastructure expensive
 - Cluster only (local uid)
 - Client can corrupt system
- Object filesystems – LUSTRE, Panasas
 - Clients can not corrupt the server
 - Only few in linux
 - Metadata server req'd (also for access ctrl)

Backup at HEP sites

- 9 Home made (INFN, RAL-HPC)
- 7 TSM (CERN, DESY-H, FZK, IN2P3, 3xINFN)
 - “Functionality, wide range of clients, good support”
- 3 Legato (DESY-Z, 2xINFN)
 - “Good offer for our Solaris requirements when started”
- 2 Time Navigator (2xINFN)
- 3 HP Data Protector (3xINFN)
- 1 TiBS(FNAL)
 - “Scalable, easily partitioned, met requirements”
- 1 AMANDA (TRIUMF)
 - “works and is cheap”
- 1 Oracle Reliaty (JLAB)
 - “Best fit to requirements in 2001”

- ssh, grid certs, Kerberos, passwords
- OTP deployed at LBNL
 - Easy to integrate
 - Slightly expensive hw tokens
- SSO project at CCLRC
 - Grid access via MyProxy, integrated with site AD and site user mgmt
 - Deployed for NGS

- CERN uses BOINC
 - (<http://boinc.berkeley.edu/>) to utilise extra resources (see LHC@home)
- Data Transfer
 - FTS seems more robust than RFT atm (see Graeme's talk)
- Procure by specint (FZK)
 - Then economy (number, power, ...)
- Helpdesk
 - SLAC (also) switched from Remedy to RT