
Quantian: A single-system image scientific cluster computing environment

Dirk Eddelbuettel

edd@debian.org

Presentation at the



conference, Vienna, May 20-22, 2004

Introduction and overview

- Quantian is a directly bootable and self-configuring Linux system that runs from a compressed cdrom image.
- Quantian contains several hundred megabytes of scientific software with a focus on scientific computing:
 - math: gap, giac, ginac, maxima, pari, yacas (+ axiom)
 - stats: **R** & CRAN packages, ESS, Ggobi; gretl, xlipstat
 - viz: opendx, mayavi, gnuplot, grace, gri, plotutils, xfig
 - matrix: octave, perl pdl, scientific / numeric python, yorick
 - typesetting: \LaTeX with xemacs/auctex/preview, lyx, kile
 - libraries: gsl, quantlib

Adding cluster computing

- Only the first two releases of Quantian were based directly on Knoppix. Since release 0.3, clusterKnoppix has been used as a basis.
- This has added a Kernel with the openMosix patch as well as a set of openMosix utilities such as openmosixview.
- As a result, “instant cluster computing” is possible based on a single cdrom:
 1. boot one master instance from the cdrom,
 2. enable 'openmosixterminalserver' from the menu after answering a few simple configuration questions,
 3. boot one, two, three, . . . 'slave' nodes via the PXE protocol (enabled in most recent desktop or laptop computers) from the master, and
 4. enjoy openMosix on the cluster.

Current issues I: Size

- Size is a constant problem:
 - Users like to see more additional software added. Some of these packages are truly huge (e.g. the parts of the ROOT physics libraries from CERN that have now been added to Debian, or the entire BioConductor project).
 - We would also like to add more if not all of the available documentation for packages already present in Quantian.
 - Lastly, Open Source software has a tendency to grow rapidly in scope and size so that newer almost always exceed the size requirements of the prior releases.
- There are two direct solutions to the size constraint imposed by cdrom technology:
 - Ignore cdroms and boot directly from the hard disk
 - Switch to dvds

Current issues II: Beowulf extensions

- Beowulf-style computing with explicit parallelism
 - An easy-to-use environment is provided by the R package snow (Tierney et al., 2003) which wraps around LAM/MPI (rmpi), PVM (rpvm) or direct socket connections
 - Currently held back by pre-historic build procedure for the sprng libraries needed for the rsprng RNG for parallel computing
 - LAM, PVM and other libraries also installed with headers for building / running custom applications

Current issues III: openMosix extensions

- Better security in open clusters based on tyd (Ian Latter, Macquarie U) which layers an OpenSWAN-based VPN over the (possibly public) tcp/ip communications layer used by openMosix.
- Better openMosix tools based on ongoing openMosix work.