

LINUX JOURNAL

Since 1994: The Original Magazine of the Linux Community

FEBRUARY 2008 | ISSUE 166



REVIEWED

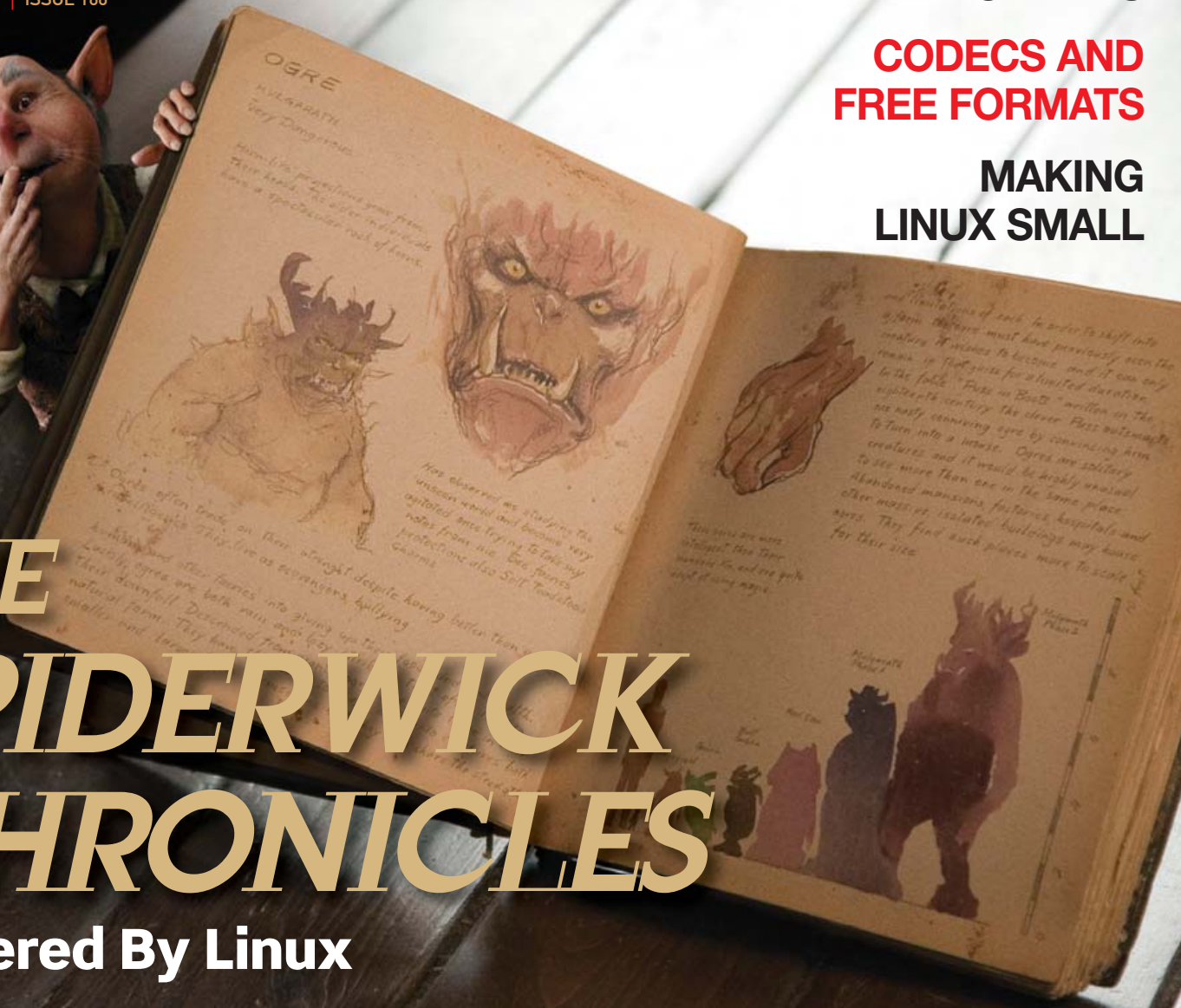
Mini Zonbu PC



SCARY DNS PROBLEMS

CODECS AND FREE FORMATS

MAKING LINUX SMALL



THE SPIDERWICK CHRONICLES

Powered By Linux

- » VirtualBox How-To
- » Virtualize Servers Painlessly
- » Distributed Hosting with OpenVPN

www.linuxjournal.com

\$5.99US \$5.99CAN





MULTIPLY ENERGY EFFICIENCY AND MAXIMIZE COOLING.

THE WORLD'S FIRST QUAD-CORE PROCESSOR FOR MAINSTREAM SERVERS. THE NEW QUAD-CORE INTEL® XEON® PROCESSOR 5300 SERIES DELIVERS UP TO 50% MORE PERFORMANCE¹ THAN PREVIOUS INTEL XEON PROCESSORS IN THE SAME POWER ENVELOPE. BASED ON THE ULTRA-EFFICIENT INTEL® CORE™ MICROARCHITECTURE, IT'S THE ULTIMATE SOLUTION FOR MANAGING RUNAWAY COOLING EXPENSES. LEARN WHY GREAT BUSINESS COMPUTING STARTS WITH INTEL INSIDE. VISIT INTEL.COM/XEON.

ACMA PERFORMA SERVERS



- Robust performance with the Quad-Core Intel® Xeon® processor
- Superior storage capacity
- Advanced RAID 6 technology for maximum data protection
- Redundant capability to minimize system downtime



Call 1-866-526-1139 or visit www.ACMA.com

ACMA's storage servers with powerful and redundant hardware are engineered to fulfill your most stringent demands for business and enterprise storage applications.



Intel is not responsible for and has not verified any statements or computer system product specific claims contained herein.

1. Performance measured using SPECjbb2005*, SPECjbb2005*/SysWatt, comparing a Quad-Core Intel® Xeon® processor E5345-based platform to a Dual-Core Intel® Xeon® processor 5160-based platform.
© 2007 Intel Corporation. All rights reserved. Intel, the Intel logo, Intel. Leap ahead., the Intel. Leap ahead. logo, Xeon, and Xeon Inside are trademarks of Intel Corporation in the U.S. and other countries.
*Other names and brands may be claimed as the property of others.

Manage Any Data Center. Anytime. Anywhere.



Avocent builds hardware and software to access, manage and control any IT asset in your data center, online or offline, keeping it, and your business, "always on".



Visit us on our Remote Control Tour. For locations near you, go to www.avocent.com/remotecontrol.



Avocent[®]
The Power of Being There[®]

Avocent, the Avocent logo and The Power of Being There, are registered trademarks of Avocent Corporation. ©2008 Avocent Corporation.

CONTENTS

FEBRUARY 2008
Issue 166



FEATURES VIRTUALIZATION

40 VIRTUALBOX: BITS AND BYTES MASQUERADING AS MACHINES

Reconfiguring your existing system to do something new is painful. Using VirtualBox is not.

Jon Watson

48 VIRTUALIZE A SERVER WITH MINIMAL DOWNTIME

If you are ready to take the plunge into virtualization, you are ready for this tried-and-true procedure to convert your existing physical machines into virtual clones.

Kyle Rankin

52 BUILDING A MULTISOURCED INFRASTRUCTURE USING OPENVPN

Explore distributed hosting options with OpenVPN.

Dmitriy Samovskiy

ON THE COVER

- Scary DNS Problems, p. 58
- Codecs and Free Formats, p. 93
- Making Linux Small, p. 76
- *The Spiderwick Chronicles* Powered by Linux, p. 64
- VirtualBox How-To, p. 40
- Virtualize Servers Painlessly, p. 48
- Distributed Hosting with OpenVPN, p. 52a

ALL SPIDERWICK IMAGES © PARAMOUNT PICTURES.

4x4

to drive your business

AMD® Quad-Core Opteron™ servers from Polywell

www.Polywell.com/LJ

1U Multi-purpose Servers

Advanced 1U Servers: starting at \$3,299
Poly 2500A16 - 2x Quad Core Opteron 8347
16GB DDR2, 4x250GB HD, Dual Gigabit LAN
Entry Level 1U starts at \$499



Polywell has been specializing in building customized computer solutions for over 20 years.

Our system specialists will work with you to get you the best solution for any project.

Polywell has a tech team fully dedicated to supporting our Linux users. We also offer a wide selection of Linux distros.

All Opteron servers will support both Dual-Core and Quad-Core.

1U to 8U up to 32 way, 128G RAM

2U 16Way Server:
Poly 8425SS - 4x Opteron 8347, 64GB DDR2
8x500GB (RAID 5 Storage) 3xGigabit LAN



Storage up to 24TB

12TB Storage driven by Opterons starting at \$7,299
Other Options: 24TB Storage
SAN/NAS, SUMA Storage also available



Blades 10 Dual or Quad Processors

PolyBlade 2500A:
10x (Dual Opteron 2210, 4GB RAM, 80G HD)
Blades servers need not be fully populated.



AMD Dual-Core & Quad-Core technology enables one platform to meet the needs of multi-tasking and multi-threaded environments; provides platform longevity.

Polywell OEM Services, Your Virtual Manufacturer

Prototype Development with Linux/FreeBSD
Support: Small Scale to Mass Production
Manufacturing Fulfillment, Shipping and RMA Repair

888.765.9686

Polywell Computers, Inc. 1461 San Mateo Ave. South San Francisco, CA 94080 650.583.7222 Fax: 650.583.1974

Opteron and ATHLON are trademarks of Advanced Micro Devices, Inc. Image of truck is a Ford Expedition, trademark of Ford Motors Corporation. All other brands, names are trademarks of their respective companies.



- 20 Years of Customer Satisfaction
- 5-Year Warranty, Industry's Longest
- First Class Customer Service

CONTENTS FEBRUARY 2008

Issue 166

COLUMNS

- 16 REUVEN M. LERNER'S AT THE FORGE**
Integrating with Facebook Data
- 22 MARCEL GAGNÉ'S COOKING WITH LINUX**
It's a Virtual World



- 26 DAVE TAYLOR'S WORK THE SHELL**
Solve: a Command-Line Calculator Redux
- 30 KYLE RANKIN'S HACK AND /**
A Little Spring Cleaning
- 96 DOC SEARLS' EOF**
Life in the Vast Lane

REVIEW

- 36 ZONBU**
James Gray



IN EVERY ISSUE

- 8 LETTERS**
12 UPFRONT
34 NEW PRODUCTS
81 ADVERTISERS INDEX

INDEPTH

- 58 DIGGING UP DIRT IN THE DNS HIERARCHY, PART II**
The techniques presented in this second article on diagnosing the DNS use simple tools to audit access to local DNS servers from the customer (and bad guy) perspective.

Ron Aitchison

- 64 LINUX POWERS THE SPIDERWICK CHRONICLES**
Read about the Linux-based production pipeline in this new movie from Paramount.

Robin Rowe

- 72 VIRTUALIZATION 2.0: WHERE THE SIDEWALK ENDS**
What happens when you put hypervisors on different subnets? How about when the underlying physical machine fails? The answers may surprise you, and not in a good way.

Kevin Epstein

- 76 SYSTEM MINIMIZATION**
Reducing the size of a Linux platform, for those new to the process, can be a mysterious task. This article takes the mystery out of making a Linux platform small.

Gene Sally

- 80 VIRTUALIZATION WITH KVM**
Introducing KVM, an open-source hypervisor.

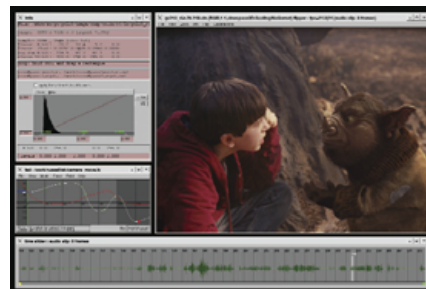
Irfan Habib

- 84 THE BEST OF BOTH WORLDS**
How to use QEMU to run Linux inside Windows.

Dashamir Hoxha

- 93 VIDEO CODECS AND THE FREE WORLD**
Learn about codecs, how to transcode, and why free formats deserve promotion.

Seth Kenlon



64 THE SPIDERWICK CHRONICLES

Next Month

DESKTOP

Next month, Doc Searls interviews Dave Liu, the 21-year-old CEO of Good OS LLC. The company's main product is gOS, an Ubuntu-based distro tweaked to run Web apps as if they were desktop ones. And, Jes Hall provides an indepth preview of the new KDE 4. Additionally, we'll cover accessibility issues on the desktop with articles on Orca and Accerciser.

As always, there's much more. We'll have reviews of the new ASUS Eee PC, the Google Android SDK and IBM's Lotus Symphony. Plus, look for an updated multi-track video editor round-up from Dan Sawyer.

USPS *LINUX JOURNAL* (ISSN 1075-3583) (USPS 12854) is published monthly by Belltown Media, Inc., 2211 Norfolk, Ste 514, Houston, TX 77098 USA. Periodicals postage paid at Houston, Texas and at additional mailing offices. Cover price is \$5.99 US. Subscription rate is \$25/year in the United States, \$32 in Canada and Mexico, \$62 elsewhere. POSTMASTER: Please send address changes to *Linux Journal*, PO Box 980985, Houston, TX 77098. Subscriptions start with the next issue. Canada Post: Publications Mail Agreement #41549519. Canada Returns to be sent to Bleuchip International, P.O. Box 25542, London, ON N6C 6B2

EmperorLinux

...where Linux & laptops converge



Portable

Since 1999, EmperorLinux has provided pre-installed Linux laptops to universities, corporations, government labs, and individual Linux enthusiasts. Our laptops range from full-featured ultra-portables to desktop replacements. All systems come with one year of Linux technical support by phone and e-mail, and full manufacturers' warranties apply.

Toucan T61/T61ws

ThinkPad T61/T61ws by Lenovo

- Up to 15.4" WUXGA w/ X@1920x1200
- NVidia Quadro FX 570M graphics
- 1.8–2.6 GHz Core 2 Duo
- 512 MB–4 GB RAM
- 80–200 GB hard drive
- CDRW/DVD or DVD±RW
- 5.2–6.0 pounds
- 10/100/1000 Mbps ethernet
- 802.11a/b/g (54Mbps) WiFi
- Starts at \$1530



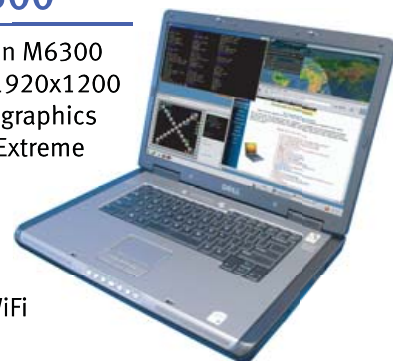
Powerful

EmperorLinux specializes in the installation of Linux on a wide range of the finest laptops made by IBM, Lenovo, Dell, Sony, and Panasonic. We customize your choice of Linux distribution to your laptop and provide support for: ethernet, wireless, X-server, ACPI power management, USB, EVDO, PCMCIA, FireWire, CD/DVD/CDRW, sound, and more.

Rhino D830/M6300

Dell Latitude D830/Precision M6300

- Up to 17" WUXGA w/ X@1920x1200
- NVidia Quadro FX 1600M graphics
- 1.8–2.8 GHz Core 2 Duo/Extreme
- 512 MB–4 GB RAM
- 60–200 GB hard drive
- DVD±RW or Blu-ray
- 6.3–8.6 pounds
- 802.11a/b/g (54Mbps) WiFi
- ExpressCard/EVDO
- Starts at \$1350



Unique

EmperorLinux offers Linux laptops with unique features. Ruggedized Panasonic laptops are designed for harsh environments: drops, vibrations, sand, rain, and other extremes. ThinkPad tablet PCs are like other laptops, with an LCD digitizer for pen-based input both as a mouse and with pressure sensitivity for writing and drawing on-screen.

Raven X61 Tablet

ThinkPad X61 Tablet by Lenovo

- 12.1" SXGA+ w/ X@1400x1050
- 1.6 GHz Core 2 Duo
- 1–4 GB RAM
- 80–200 GB hard drive
- 3.8 pounds
- Pen/stylus input to screen
- Dynamic screen rotation
- Handwriting recognition
- X61s laptops available
- Starts at \$2150



LJ '07 Ultimate
Linux Laptop

www.EmperorLinux.com 1-888-651-6686

Model prices, specifications, and availability may vary. All trademarks are the property of their respective owners.

GO SOLID.
INCREASE RELIABILITY.



solid state systems
fully x86 compatible
fanless, quiet operation



Direct-Plug
IDE Flash Modules

Intel, VIA & AMD CPUs

95% Efficiency-Rated
PicoPSU Power Supplies

DISCOVER MINI-ITX.

LOGIC
SUPPLY

www.logicsupply.com

LINUX JOURNAL

Executive Editor Jill Franklin
jill@linuxjournal.com

Senior Editor Doc Searls
doc@linuxjournal.com

Art Director Garrick Antikajian
garrick@linuxjournal.com

Products Editor James Gray
newproducts@linuxjournal.com

Editor Emeritus Don Marti
dmarti@linuxjournal.com

Technical Editor Michael Baxter
mab@cruzio.com

Senior Columnist Reuven Lerner
reuven@lerner.co.il

Chef Français Marcel Gagné
maggagne@salmar.com

Security Editor Mick Bauer
mick@visi.com

Contributing Editors

David A. Bandel • Ibrahim Haddad • Robert Love • Zack Brown • Dave Phillips
Marco Fioretti • Ludovic Marcotte • Paul Barry • Paul McKenney • Dave Taylor

Proofreader Geri Gale

Publisher Carlie Fairchild
publisher@linuxjournal.com

General Manager Rebecca Cassity
rebecca@linuxjournal.com

Director of Sales Laura Whiteman
laura@linuxjournal.com

Regional Sales Manager Joseph Krack
joseph@linuxjournal.com

Regional Sales Manager Kathleen Boyle
kathleen@linuxjournal.com

Circulation Director Mark Irgang
mark@linuxjournal.com

System Administrator Mitch Frazier
sysadm@linuxjournal.com

Webmaster Katherine Druckman
webmaster@linuxjournal.com

Accountant Candy Beauchamp
acct@linuxjournal.com

Linux Journal is published by, and is a registered trade name of, Belltown Media, Inc.
PO Box 980985, Houston, TX 77098 USA

Reader Advisory Panel

Brad Abram Baillio • Nick Baronian • Hari Boukis • Caleb S. Cullen • Steve Case
Kalyana Krishna Chadalavada • Keir Davis • Adam M. Dutko • Michael Eager • Nick Faltys • Ken Firestone
Dennis Franklin Frey • Victor Gregorio • Kristian Erik • Hermansen • Philip Jacob • Jay Kruiuzenga
David A. Lane • Steve Marquez • Dave McAllister • Craig Oda • Rob Orsini • Jeffrey D. Parent
Wayne D. Povel • Shawn Powers • Mike Roberts • Draciron Smith • Chris D. Stark • Patrick Swartz

Editorial Advisory Board

Daniel Frye, Director, IBM Linux Technology Center
Jon "maddog" Hall, President, Linux International
Lawrence Lessig, Professor of Law, Stanford University
Ransom Love, Director of Strategic Relationships, Family and Church History Department,
Church of Jesus Christ of Latter-day Saints
Sam Ockman
Bruce Perens
Bdale Garbee, Linux CTO, HP
Danese Cooper, Open Source Diva, Intel Corporation

Advertising

E-MAIL: ads@linuxjournal.com
URL: www.linuxjournal.com/advertising
PHONE: +1 713-344-1956 ext. 2

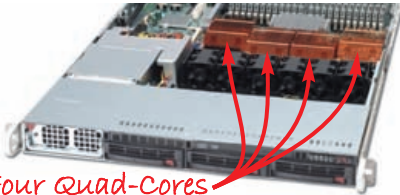
Subscriptions

E-MAIL: subs@linuxjournal.com
URL: www.linuxjournal.com/subscribe
PHONE: +1 713-589-3503
FAX: +1 713-589-2677
TOLL-FREE: 1-888-66-LINUX
MAIL: PO Box 980985, Houston, TX 77098 USA
Please allow 4-6 weeks for processing address changes and orders
PRINTED IN USA

LINUX is a registered trademark of Linus Torvalds.

QUAD-CORE. SAY NO MORE!

ABERDEEN STIRLING 144



Four Quad-Cores

1U 3TB Quad Quad-Core HPC Server

- Up to four Quad-Core Intel® Xeon® MP processors
- Intel 7300 chipset with 64-Bit Support
- Up to 192GB 667MHz ECC Fully Buffered FBDIMM Memory
- Up to 3 x 1TB (3TB Total) Hot-Swap SATA Hard Drives
- 1000W AC Power Supply w/PFC
- 5-Year Warranty

Starting at **\$4,975**

ABERDEEN STIRLING 128



1U 4TB Dual Quad-Core Server

- Up to two Quad-Core Intel Xeon processors
- Intel 5000P Chipset with 64-Bit Support
- Up to 64GB 667MHz ECC Fully Buffered FBDIMM Memory
- Up to 4 x 1TB (4TB Total) Hot-Swap SATA Hard Drives
- 560W AC Power Supply w/PFC
- 5-Year Warranty

Starting at **\$1,899**

ABERDEEN STIRLING 228



2U 8TB Dual Quad-Core Server

- Up to two Quad-Core Intel Xeon processors
- Intel 5000P Chipset with 64-Bit Support
- Up to 64GB 667MHz ECC Fully Buffered FBDIMM Memory
- Up to 8 x 1TB (8TB Total) Hot-Swap SATA Hard Drives
- 700W Redundant Hot-Swap Power Supply
- 5-Year Warranty

Starting at **\$2,279**

ABERDEEN STIRLING X314



3U 12TB Dual Quad-Core Storage Server

- Up to two Quad-Core Intel Xeon processors
- Intel 5000P Chipset with 64-Bit Support
- Up to 32GB 667MHz ECC Fully Buffered FBDIMM Memory
- Up to 12 x 1TB (12TB Total) Hot-Swap SATA Hard Drives
- Areca ARC-1231ML PCI Express 800MB/sec RAID Controller
- 650W 2+1 Redundant Hot-Swap Power Supply
- 5-Year Warranty

Starting at **\$3,389**

ABERDEEN STIRLING X418



4U 16TB Dual Quad-Core Storage Server

- Up to two Quad-Core Intel Xeon processors
- Intel 5000P Chipset with 64-Bit Support
- Up to 32GB 667MHz ECC Fully Buffered FBDIMM Memory
- Up to 16 x 1TB (16TB Total) Hot-Swap SATA Hard Drives
- Areca ARC-1261ML PCI Express 800MB/sec RAID Controller
- 650W 2+1 Redundant Hot-Swap Power Supply
- 5-Year Warranty

Starting at **\$3,849**

ABERDEEN STIRLING X526



5U 24TB Dual Quad-Core Storage Server

- Up to two Quad-Core Intel Xeon processors
- Intel 5000P Chipset with 64-Bit Support
- Up to 32GB 667MHz ECC Fully Buffered FBDIMM Memory
- Up to two Internal SATA Hard Drives for OS
- Areca ARC-1280ML PCI Express 800MB/sec RAID Controller
- 950W 3+1 Triple Redundant Hot-Swap Power Supply
- 5-Year Warranty

Starting at **\$4,989**

ABERDEEN STIRLING X633



6U 32TB Dual Quad-Core Storage Server

- Up to two Quad-Core Intel Xeon processors
- Up to 32GB 667MHz ECC Fully Buffered FBDIMM Memory
- Up to 32 x 1TB (32TB Total) Hot-Swap SATA Hard Drives
- Up to two Rear Hot Swap SATA Hard Drives for OS
- Dual Areca PCI Express 800MB/sec RAID Controllers
- 1350W 3+1 Triple Redundant Power Supply
- 5-Year Warranty

Starting at **\$6,439**

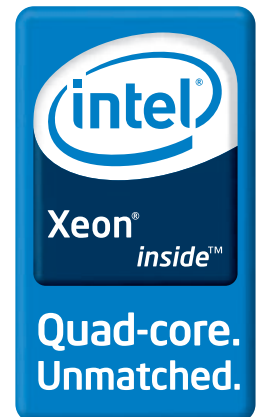
ABERDEEN STIRLING X840



8U 40TB Dual Quad-Core Storage Server

- Up to two Quad-Core Intel Xeon processors
- Up to 32GB 667MHz ECC Fully Buffered FBDIMM Memory
- Up to 40 x 1TB (40TB Total) Hot-Swap SATA Hard Drives
- Up to two Rear Hot Swap SATA Hard Drives for OS
- Dual Areca PCI Express 800MB/sec RAID Controllers
- 1350W 3+1 Triple Redundant Power Supply
- 5-Year Warranty

Starting at **\$8,029**



letters



Have a Beer, for Free

In Shawn Powers' December 2007 *Linux Journal* article "Quake, Meet GPL; GPL, Meet Quake", he states, "As Linux users, we're familiar with terms like, 'Free as in speech and free as in beer'. For the record, I have never understood the latter part of that motto. Beer is rarely free."

The reason he does not understand the second part is because he is misquoting. The true quote is: "Free as in free speech, not as in free beer."

I am sadly surprised that the editors of a magazine dedicated to the Linux community did not catch this.

--
Eric

Shawn Powers replies: in the Quake 3 article, I wrote that I didn't fully understand the notion of "Free as in beer." In the next paragraph, I go on to explain what it means, and left the "I don't get it" part as a silly view of the free beer concept. For the record, I get it. The responses I received via e-mail, however, make me wonder if we, as a community, really do "get it".

The most common response I got was that I misquoted the free software

definition, and that it's supposed to be, "Free as in speech, not as in beer." The problem is that I wasn't trying to quote the definition of "free software" but rather talk about what is meant by free beer. Trust me, I didn't wrongly coin the idea of software being free as in beer. Just ask Google.

Although I won't ramble on regarding the definitions and social implications of wordsmithing semantics, I will point out something that pains me as a Linux user. If we become an exclusive community that listens only to those versed in the doctrine of the FSF, and we don't remember our grass roots, we're doomed to be an elitist group of snobs.

So, although I understand (mostly) the ideas of free software, and I understand that free beer is referring to getting something for no money, I still say it's hard to find beer for free. So come on in, sit down, and I'll buy you a beer. We can talk about how to get one for free together—and maybe frag each other in OpenArena, because that I know where to get without paying a dime.

Why Dual-Boot?

Regarding James Gray's "The State of the Market: a Linux Laptop Buying Guide" in the December 2007 issue of *LJ*: if the Linux desktop is so great, why would a Linux laptop vendor not offering dual boot with Windows be considered a negative? Why would Linux desktop users *want* their Linux laptop to be dual-booted with Windows?

The same could be said of the "My Triple Boot Laptop" article in the same issue. Why is multiple booting with Windows such an important thing?

The real question is, why is the Linux desktop not usable enough to throw Windows away? And, what can we do about it?

Don't get me wrong, I'm a Linux guy and love your magazine!

--
Mike

James Gray replies: Thanks for your message. I get your point and the underlying frustration that comes with it. Although I cannot speak for every Linux user, our community has feistily developed a culture of choice, openness and heterogeneous environments. In other words, we choose the right tool for the job, and make sure every part of the system can interact.

In my case, I am a devoted Linux user 99% of the time, where nearly all of my needs are met. However, for instance, my wife signed us up for the Yahoo! Music service, which I've come to enjoy a great deal. Unfortunately, Yahoo! Music doesn't give a rip (yet) about Linux, and CrossOver Linux won't run the Yahoo! Music Engine (yet).

Actions from companies like Dell, which is now expanding its Linux offerings, prove that the sluggish, mainstream computing world is finally learning what we've known for years—that Linux is here to stay. I and many of our fellow Linuxers keep Windows around as an interim solution until the day of its forthcoming irrelevance.

Floating-Point Simplicity

I've read Dave Taylor's columns since the first one, and it's nice to see someone writing about shell programming and command lines. In the December 2007 issue of *LJ*, Dave spoke about the need for doing floating-point operations at the command-line level. You can perform those actions by using a single command line involving `bc`, as you can see in the following sequence:

```
$ echo 'scale=4^J11/7' | bc
1.5714
```

The `^J` is obtained by issuing `Crtl-V+Crtl-J`.

--
Joao Macedo

KEEP YOUR BUSINESS RUNNING SMOOTHLY

PROTECT YOUR SMALL BUSINESS WITH THE BUILT-IN SECURITY ENHANCEMENTS OF THE DUAL-CORE INTEL® XEON® PROCESSOR IN YOUR SERVERSDIRECT SYSTEM.

SDR-1105T 1U ENTRY LEVEL SERVER



\$1,129

EXCELLENT GENERAL PURPOSE SERVER FOR ORGANIZATIONS WITH THE NEED FOR A LOW, ENTRY LEVEL PRICE

- * 1U Rackmount Chassis with 520W power supply
- * Supermicro X7DVL-L Server Board with Intel® 5000V (Blackford VS) Chipset
- * Intel Quad-Core Xeon Processor E5310 1.6GHZ
- * Total 512MB, 2pcs x 256MB Kingston DDR2 533Mhz FB-DIMM ECC
- * Seagate SATAII 160GB 7200 RPM 8MB Cache SATA 3.0Gb/s Hard Drive
- * 4 x 1" Hot-swap SATA Drive Bays
- * Two Intel® 82563EB Dual-port Gigabit Ethernet Controller
- * Intel® ESB2 SATA 3.0Gbps Controller RAID 0, 1, 5, 10 support



STARTING PRICE \$1,259

SDR-2503T 2U APPLICATION SERVER

Highest performing with Dual Core/ Quad Core Xeon CPU based. Excellent with general purpose applications and provide the most power.

- * 2U Rackmount Chassis with 650W power supply
- * Supermicro X7DVL-E Server Board with Intel® 5000V (Blackford VS) Chipset
- * Intel Quad-Core Xeon Processor E5310 1.6GHZ
- * Total 512MB, 2pcs x 256MB Kingston DDR2 533Mhz FB-DIMM ECC
- * Seagate SATAII 250GB 7200 RPM 16MB Cache SATA 3.0Gb/s Hard Drive
- * 6 x 1" Hot-swap SATA Drive Bays
- * Intel® (ESB2/Gilgal) 82563EB Dual-port Gigabit Ethernet Controller
- * Intel® ESB2 SATA 3.0Gbps Controller RAID 0, 1, 5, 10 support



STARTING PRICE \$1,999

SDR-3500T 3U DATABASE SERVER

Easily Scalable storage solution with hot-swap functionality for growing businesses

- * 3U Rackmount chassis with Redundant 800W power supply
- * Supermicro X7DBE+ Server Board with Intel® 5000P (Blackford) Chipset
- * Intel Quad-Core Xeon Processor E5310 1.6GHZ
- * Total 1024MB, 2pcs x 512MB Kingston DDR2 533MHz FB-DIMM ECC
- * Seagate SATAII 500GB 7200 RPM 16MB Cache SATA 3.0Gb/s Hard Drive
- * 16 x 1" Hot-swap SATA Drive Bays
- * Dual-port Gigabit Ethernet Controller
- * Intel ESB2 SATA 3.0Gbps Controller RAID 0, 1, 10 support



STARTING PRICE \$1,199

SDR-7045B-TB 4U FILE SERVER

4U Quad-Core Xeon Server offers excellent value and expandability

- * 4U Rackmountable / Tower with 650W power supply
- * Supermicro Super X7DBE Server Board with Intel® 5000P (Blackford) Chipset
- * Intel Quad-Core Xeon Processor E5310 1.6GHZ
- * Total 1024MB, 2pcs x 512MB Kingston DDR2 667MHz FB-DIMM ECC
- * Seagate SATAII 750GB 7200 RPM 16MB Cache SATA 3.0Gb/s Hard Drive
- * 6 x 3.5" Hot-swap SAS/SATA Drive Bays
- * Dual-port Gigabit Ethernet Controller
- * Intel ESB2 SATA 3.0Gbps Controller RAID 0, 1, 5, 10 support



STARTING PRICE \$3,099

SDR-5111T 5U ADVANCED STORAGE SERVER

Quad Core dual Xeon CPU based, with 24 hot-swap hard disk bays suitable for 18TB of pure data Storage capacity

- * 5U Rackmount chassis with Redundant 1350W power supply
- * Supermicro X7DBE Server Board with Intel® 5000P (Blackford) Chipset
- * Intel Quad-Core Xeon Processor E5310 1.6GHZ
- * Total 1024MB, 2pcs x 512MB Kingston DDR2 667MHz FB-DIMM ECC
- * Seagate 750GB 7200 RPM 16MB Cache SATA 3.0Gb/s Hard Drive
- * 24 x 1" Hot-swap Drive Bays
- * Intel® (ESB2/Gilgal) 82563EB Dual-port Gigabit Ethernet Controller
- * Intel ESB2 SATA 3.0Gbps Controller RAID 0, 1, 5, 10 support

SERVERS DIRECT CAN HELP YOU CONFIGURE YOUR NEXT HIGH PERFORMANCE SERVER SYSTEM - CALL US TODAY!

Our flexible on-line products configurator allows you to source a custom solution, or call and our product experts are standing by to help you assemble systems that require a little extra. Servers Direct - your direct source for scalable, cost effective server solutions.

1.877.727.7887 | www.ServersDirect.com

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, Pentium, and Pentium III Xeon are trademarks of Intel Corporation or it's subsidiaries in the United States and other countries.



mkdir Errors Even More Trivial

Upon reading Steeve's letter to the editor in the November 2007 issue of *Linux Journal*, I was glad to see advocacy of error handling in bash scripts, but I think the "trivial" example provided is elaborate compared to a simple trick that has been available since the days of the original Bourne shell.

The `-e` flag tells bash to stop executing the script upon encountering any error outside of a conditional expression. Error checking is often omitted simply because it is cumbersome to write. Not only is the `-e` flag more succinct, but using it makes error checking effortless.

Comparing the two methods is impressive. The longer version produces two error messages:

```
ecashin@cat ecashin$ cat err-baroque.sh
mkdir -p /foo
if [ $? -ne 0 ]; then
    echo "Error: could not create\
    directory /foo"
    exit 1
```

```
fi
echo /foo
ecashin@cat ecashin$ sh err-baroque.sh
mkdir: cannot create directory `/foo': Permission denied
Error: could not create directory /foo
ecashin@cat ecashin$
```

The shorter version leverages the fact that the Bourne shell design has always made error handling trivial:

```
ecashin@cat ecashin$ cat err-e.sh
set -e
mkdir /foo
echo /foo
ecashin@cat ecashin$ sh err-e.sh
mkdir: cannot create directory `/foo': Permission denied
ecashin@cat ecashin$
```

Error handling is a balance between doing too much and doing too little. The `-e` flag allows us to get the job done without cluttering code or tempting us not to bother.

--
Ed L. Cashin

We Have Not Won the Battle

I picked up the June 2007 copy of *LJ* today, looking for one of Dave Taylor's articles on bash programming (I'll have to go further back than that issue to find an answer to my query) and opened it at Doc Searls' article "Picking New Fights".

I'm not too sure that the battle is won, Doc, and here are a couple of cases in point—admittedly both are Aussie experiences; maybe things are better in the good-old US of A?

We had a segment on prime-time TV a couple of days ago on Negroponte's \$100 laptop, which was very interesting, with an in-depth discussion of the thought that has been put into the project to enable its use in areas without electricity and to provide a unit rugged enough to withstand its likely treatment—all in an endeavor to bring the world and its consequent education to those of us who are much less fortunate. A full half-hour presentation of a great community service, and there was not one word about Linux. NOT ONE WORD on such an important factor in keeping the price as low as possible, apart from its obvious absolute superiority

in a case like this, where I reckon those young minds will revel in the variety and flexibility that Linux can offer.

Here's another example. When an ad from Mindscape Asia Pacific Pty Ltd for "Family Tree Maker Platinum 2008" appeared in my inbox, I e-mailed them, asking what OSes were supported. There was no information in the ad about operating systems—just the software.

This was their reply:

Hi David, the operating systems supported for the new FTM 17 are Windows XP and Vista.

Kind Regards,

Ann, Mindscape

To which I e-mailed: "No Linux support, eh?"

That message triggered the following:

Hi David, no, I'm afraid no Linux support. That would be something you will need to take up with Family Tree

Maker in the US who develops the program. In fact, currently all our programs are Windows- or MAC-based with single-user licenses in most cases.

Kind Regards,

Ann

My response, to which I have had no reply, was perhaps a bit terse: "That's interesting, Ann. One wonders if it wouldn't be in Mindscape's own interests to take this matter up with FTM directly rather ask than a disgruntled prospective client of yours."

--
Dave Dartnall

And, the Winner Is...

In the December 2007 issue of *LJ*, Dave Taylor writes "The challenge with bc is to revamp how you interact with it...to write a shell script wrapper that allows us not only to do simple calculations from the command line, but also have them solved as floating-point calculations."

This isn't really much of a challenge, and it doesn't require a wrapper:

```
> echo "11/7" | bc -l
1.57142857142857142857
```

What do I win?

--
Andrew Fabbro

Thank you for your letter. Your prize is seeing your name in print!—Ed.

LinuxCertified Rocks

I bought my Linux laptop before reading James Gray's "The State of the Market: a Linux Laptop Buying Guide" article in the December 2007 issue of *LJ*. In view of the cons in that article about LinuxCertified laptops, I would like to report that I bought an LC2100DC laptop from LinuxCertified last August (not reviewed in your article), and that all of the features I use worked flawlessly out of the box. During the build of my machine, I had a phone call or two from LinuxCertified suggesting sensible changes to my original configuration, which I accepted. My laptop came with a CD containing various utilities and drivers, and a complete manual for the laptop. I also received a Feisty Fawn install disk, a Feisty Fawn restore disk (containing additional LC drivers), an XP install disk (not a restore disk), and other disks for the CD/DVD burner software and so on. I was, thus, rather surprised to see LinuxCertified put down for inadequate documentation. I have had quick responses to one or two support issues via e-mail. In short, my experience with LinuxCertified has been completely satisfactory.

--
Bob Ackermann

Additions to the December 2007 "LJ Index"

I thought I'd add a couple more entries to December's "LJ Index" that, although they don't pertain to the US Constitution, pertain to the founding of our country nonetheless and should be noted so as to not sound biased. (A media, and especially a technology-related, entity would never sound biased against religion though,

would they?)

Number of times the word "God" appears in the Declaration of Independence: 1

Number of times the word "Creator" appears in the Declaration of Independence: 1

Number of times the word "liberty" appears in the Declaration of Independence: 1

Number of times the word "freedom" appears in the Declaration of Independence: 0

In addition to these, I believe a useful poll question that should be asked is, how many people believe the US was founded as a religious nation (not necessarily Christian)? I'm sure many more people (more than the 55% who believe the Constitution founded a Christian nation) would agree that this country was founded as a religious country as opposed to a secular one. Neither term is mentioned in the Constitution; however, the Declaration of Independence unquestionably leans toward the religious.

--
Brandon McCombs

Get Out of Town!

I had to smile at the name of the CTO of RapidMind! [See "Picking the RapidMind", *LJ*, November 2007.] If he were living in Ireland, everyone would call him "Mick", which would be unfortunate for Michael, but a constant source of amusement down in the local pub!

PS. Of course, is that the McCool's from Cork or the McCool's from Derry?

--
Paul

LJ pays \$100 for tech tips we publish.
Send your tip and contact information to techtips@linuxjournal.com.

LINUX JOURNAL

At Your Service

MAGAZINE

PRINT SUBSCRIPTIONS: Renewing your subscription, changing your address, paying your invoice, viewing your account details or other subscription inquiries can instantly be done on-line, www.linuxjournal.com/subs. Alternatively, within the U.S. and Canada, you may call us toll-free 1-888-66-LINUX (54689), or internationally +1-713-589-2677. E-mail us at subs@linuxjournal.com or reach us via postal mail, Linux Journal, PO Box 980985, Houston, TX 77098-0985 USA. Please remember to include your complete name and address when contacting us.

DIGITAL SUBSCRIPTIONS: Digital subscriptions of *Linux Journal* are now available and delivered as PDFs anywhere in the world for one low cost. Visit www.linuxjournal.com/digital for more information or use the contact information above for any digital magazine customer service inquiries.

LETTERS TO THE EDITOR: We welcome your letters and encourage you to submit them to ljeditor@linuxjournal.com or mail them to Linux Journal, 1752 NW Market Street, #200, Seattle, WA 98107 USA. Letters may be edited for space and clarity.

WRITING FOR US: We always are looking for contributed articles, tutorials and real-world stories for the magazine. An author's guide, a list of topics and due dates can be found on-line, www.linuxjournal.com/author.

ADVERTISING: *Linux Journal* is a great resource for readers and advertisers alike. Request a media kit, view our current editorial calendar and advertising due dates, or learn more about other advertising and marketing opportunities by visiting us on-line, www.linuxjournal.com/advertising. Contact us directly for further information, ads@linuxjournal.com or +1 713-344-1956 ext. 2.

ON-LINE

WEB SITE: Read exclusive on-line-only content on *Linux Journal's* Web site, www.linuxjournal.com. Also, select articles from the print magazine are available on-line. Magazine subscribers, digital or print, receive full access to issue archives; please contact Customer Service for further information, subs@linuxjournal.com.

FREE e-NEWSLETTERS: Each week, *Linux Journal* editors will tell you what's hot in the world of Linux. Receive late-breaking news, technical tips and tricks, and links to in-depth stories featured on www.linuxjournal.com. Subscribe for free today, www.linuxjournal.com/enewsletters.

diff -u

WHAT'S NEW
IN KERNEL
DEVELOPMENT

for translators to keep up with them.

Vince Kim submitted a patch to add support for **LZO compression to CramFS**, using **Richard Purdie's LZO kernel library**. The result was a performance gain, at a cost of a 10% larger driver binary.

Zhang Wei posted a driver for the **Freescale MPC8540 DMA controller**, commonly used in routers, switches, printers and similar devices. He also added a corresponding "Freescale DMA Driver" entry to the MAINTAINERS file, listing himself as the official maintainer.

Stephen Hemminger posted a driver, called **apanel**, to control the panel lights on some **Fujitsu LifeBook laptops**. He based his work on an earlier effort by **Jochen Eisenger**, but Stephen's work uses no `ioctl`s or user-space `dæmons` as Jochen's did. **Andrew Morton** replied with some minor criticism and praised Stephen's mastery of operator precedence in C. It looks like this driver will be accepted fairly quickly.

Samuel Ortiz posted a driver for the **Compaq ASIC multifunction chip** found in many handheld devices. Andrew Morton ran the `checkpatch` script on Samuel's patch and found many stylistic problems, which he asked Samuel to take care of and resubmit. Andrew also had technical issues and questions, and Samuel posted a new patch in response.

As he has done many times before, **Adrian Bunk** made another abortive effort to remove the **eeepro100 driver** that the **e100 driver** is supposed to replace. He submitted his patch, and **Jeff Garzik** and **Auke Kok** pointed out that there were still known problems with e100 that made it not yet a suitable replacement. **David Acker**, who's been working on these issues, said he would step up his efforts, but he also said that the difficulty of testing specific bugs had generally made the project a lower priority for him at the moment.

Adrian also posted a patch removing three **I2C drivers**: `i2c-ixp2000.c`, `i2c-ixp4xx.c` and `scx200_i2c.c`; there was no dissent on the list, so this probably is the end of those drivers in Linux.

Adrian also posted a patch to remove **legacy I2C RTC drivers** that already have replacement drivers in the kernel.

Keiichi Kii translated the **SubmittingPatches** file into Japanese, and **Greg Kroah-Hartman** passed it along for inclusion in the kernel tree. Greg initially had suggested the translation of this and other kernel docs himself, pointing out that the originals rarely changed, so it should be fairly easy

But, **Jean Delvare** said that some platforms still relied on the legacy drivers, and that they should be updated to use the replacements before the old drivers were removed. He alerted the **rtc-linux mailing list** that the **PowerPC platform code** should be updated as soon as possible.

Robert P. J. Day posted a patch removing the remaining bits of **APUS support** from the PowerPC architecture. Some APUS code already had been removed in 2.6.23, and the rest had been listed as broken for more than two years, so it was time to go. No one voiced any opposition to this patch, but Adrian said he also had a similar patch he'd been planning to release soon. It turns out Robert has written a script to find dead code in the kernel, which he runs every once in a while to locate things that can be removed safely.

Gabriel Craciunescu discovered that the **TLAN network driver mailing list** would accept posts only from subscribers, so he posted a patch to note that in the MAINTAINERS file.

Sam Ravnborg announced the creation of a new **linux-kbuild mailing list** on the **vger servers** to replace the old `kbuild-devel` list on **SourceForge** and posted a patch updating the MAINTAINERS file to show the new list. The old list could be posted to only by subscribers, and it also was moderated. Sam decided to make the new list after he'd seen too many e-mail messages dropped from the old one. There was not much discussion on `linux-kernel` about this, but it's doubtful any serious objections will be made. The old list was subscriber-only primarily because of a spam problem that had started when the list first came to SourceForge. Presumably, with the relatively new antispam measures that have been adopted on `vger`, that problem should be a lot more tolerable.

Having obtained the relevant hardware, **Maciej W. Rozycki** posted a patch to the MAINTAINERS file, listing the **DZ DECStation DZ11 serial driver** and listing himself as the official maintainer.

Larry Finger has stepped down from maintaining the **b43legacy code** and is seeking a new maintainer. He's also offered to give a **Linksys WPC54G networking card** that has the relevant BCM4306/2 chip to anyone who'll take over the code. Apparently, most of the maintenance requirements involve porting **Michael Beusch's b43 patches** into `b43legacy`. Michael also is stepping down from maintaining the `bcm43xx` code, but a replacement will not be needed, as that code is no longer needed and will be coming out of the kernel at some point.

—ZACK BROWN

Join Us at LinuxJournal.com

We are pleased to invite you to come see what all the excitement is about at LinuxJournal.com. New features and exclusive on-line content help keep our on-line home a pretty lively place.

You can't miss the Web-only

content featured at the top of our home page. Here, our authors share the latest information, in-depth tutorials and product reviews. These articles are available only on-line, so be sure to check back often and subscribe to our RSS feed to see

what's new.

We plan to bring you new features and relevant stories on-line regularly, and hope you will join us and take part as we continue to build the LinuxJournal.com community.

—KATHERINE DRUCKMAN

LJ Index, February 2008

1. Number of x86 processors required to perform the same amount of work as one IBM System z mainframe: **250**
2. zSeries mainframe energy consumption as a percentage of that required by 250 x86 processors: **2**
3. Percentage of all physical servers that will be virtualized by 2011: **50**
4. Number of partners in Google's Open Handset Alliance for its Linux-based Android phone platform: **30**
5. Number of Google employees working on the Android effort: **100**
6. Millions of mobile phones sold worldwide in Q3 2007: **289**
7. Minimum billions of dollars Google will offer in the US 700MHz spectrum auction: **4.6**
8. Number of steam engine locomotive makers who succeeded in the diesel engine business: **0**
9. Billions of phone lines in the world: **4**
10. Billions of mobile phone accounts: **2.68**
11. Millions of Bluetooth-enabled device shipments reached in 2007: **800**
12. Billions of Nokia phones in use: **900**
13. Age of Nokia as a company in years: **142**
14. Billions Nokia is spending to become a "consumer Web media company": **9**
15. Billions of mobile phones that will be sold in 2008: **1.3**
16. Percentage of 2008 mobile phones that will be sold in Asia/Pacific: **82**
17. Linux's percentage of Netcraft's top ten most-reliable hosting companies for September 2007: **50**
18. Linux's percentage of Netcraft's top three most-reliable hosting companies for September 2007: **100**
19. Linux's percentage of Netcraft's top 48 most-reliable hosting companies for September 2007: **43.75**
20. Percentage of the top 48 most-reliable hosting companies for September 2007 that are Linux, FreeBSD, Solaris or F5 Big-IP (BSD-based): **66.7**

Sources: 1, 2: IBM and its Power Estimator Tool, CNN | 3: IDC, via Guardian.co.uk | 4: *The Register* | 5-7, 12-16: *Forbes* | 8: Bob Frankston | 9, 10: Trends in Telecommunication Reform 2007, from the ITU, via Dilanchian.com.au | 11: *Laptop Magazine* | 17-20: Netcraft.com

The gPhone That Isn't

Google Keeps the Linux-Based Handset Market Open

What stood out most in Google's announcement of Android in November 2007 was not that it was a Linux-based open phone platform (there are several of those already), but that it looked, quite literally, like something more than the "iPhone killer" many expected. It was, instead, a closed-phone killer.



Well, not really. The iPhone isn't going to die. Rather, it looked like an open alternative to the iPhone that might be exactly what the phone-makers need to get out of their death dance with carriers and start making cellular telephony (and mobile everything) Net-native.

Apple, for all its inventiveness (which is enormous—credit where due), did a deal with the devil when it launched the iPhone in partnership with AT&T. In so doing, Apple became a captive manufacturer for one carrier and crippled the iPhone's Net nativity. Google, on the other hand, put its enormous market heft behind all phone-makers with the guts to risk breaking ranks with its carrier partners and to start making truly open mobile handsets (a carefully chosen word that means "more than phones").

The platform is called Android, and the SDK invites development of all kinds of devices, with phones playing the center circle of the market's bull's-eye.

But, the target is much bigger. To explore those dimensions, Google is offering \$10 million in awards for developers building mobile apps for the platform.

Challenge I runs from January 2, 2008 through March 3, 2008 (right now, if you're reading this fresh off the newsstand or out of the mailbox). Fifty winners will receive \$25,000 toward additional development and will be eligible for ten awards of \$275,000

each, plus ten others at \$100,000 each. Those are due May 1, 2008, and will be announced at the end of that month. The \$5 million Challenge II will reward development on Android-based handsets that will start shipping later in the year. Details for that have not yet been revealed at the time of this writing.

Winners will "leverage all that the Android platform has to offer in order to provide consumers with their most compelling experiences". If you win (or even if you don't), the intellectual property you create (even if you don't wish to call it that) will be yours to keep.

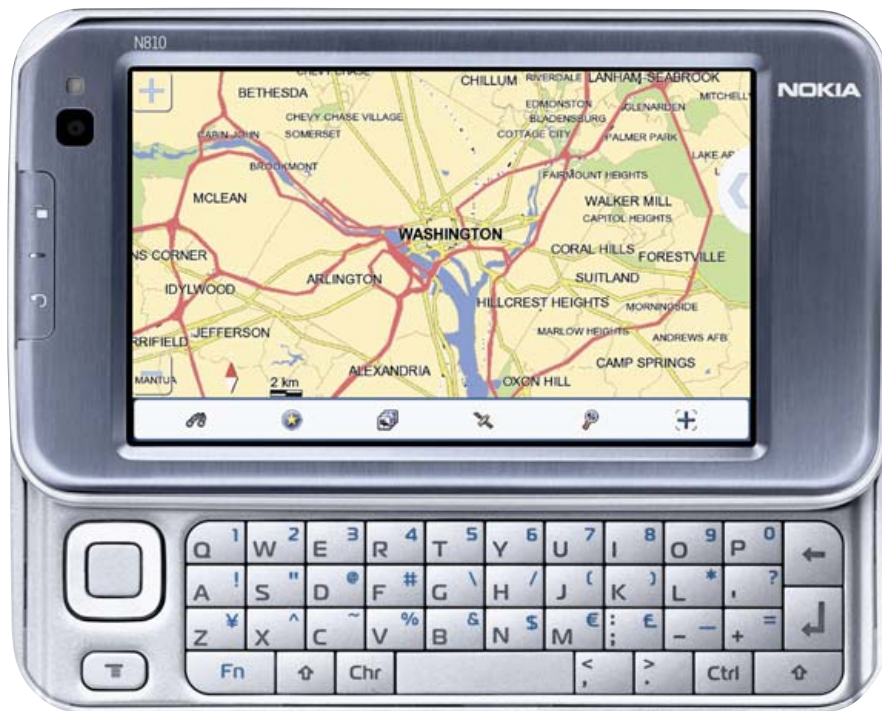
What's especially cool about the "gPhone" is that it isn't for, or by, Google. That was the defaulted expectation of many, based on expectations set by Apple with the iPhone. Instead, the field remains, as it already was, wide open.

RESOURCES

- Industry Leaders Announce Open Platform for Mobile Devices: www.google.com/intl/en/press/pressrel/20071105_mobile_open.html
- Google Announces \$10 Million Android Developer Challenge: www.google.com/intl/en/press/pressrel/20071112_android_challenge.html

—DOC SEARLS

Third-Generation Nokia Tablet Gains a Keyboard and GPS



We can't wait to get our hands on the new Nokia N810, announced last October and released in November 2007 (when we're writing this). Unlike its predecessors, the N770 and the N800, the N810 has one big stand-out (actually, slide-out) difference: a qwerty keyboard. That alone makes it far more desirable, and far more like the departed Sharp Zaurus—a unit some of us still miss.

The other big difference is built-in GPS. The N800, which currently holds the title of Ultimate Linux Handheld (September 2007), required an external

GPS receiver connected to the unit by Bluetooth.

On the downside (at least for this writer) is that Nokia has dropped the built-in FM radio featured in the N800. We liked that feature and were looking forward to improvements in it, such as RDS support. Perhaps, if enough of us care, a worthy FM radio will return in a future version.

Meanwhile, we look forward to reviewing the N810 in depth for an upcoming issue of *Linux Journal*.

—DOC SEARLS



They Said It

Ask not if your company is ready for open source, ask if open source is ready for your company.

—Laurent Lachal, Ovum, said at a conference in the UK, October 31, 2007

Many enterprises are managing quite well with both version 2 and version 3 of the GPL.

—Black Duck Software, said at a conference in the UK, October 31, 2007

The OSS is a meritocracy. If you are the chairman of IBM and you submit a patch to the kernel or to KDE that is rubbish, they will tell you. They don't care who you are, how much experience you have or how nice a guy you are. If you are short, tall, fat, thin, man, woman, OAP or teenager, your code is equally judged on its merit rather than on you. As someone who finds sucking up in business intolerable, this is very refreshing.

—Mike Arthur, mikearthur.co.uk/?p=162#comments

I still don't know of a single example of an exclusive platform that worked. Yet companies still try to launch them, ignoring history, and hoping that they can control who gets to make their platform a winner.

—Dave Winer, www.scripting.com/stories/2007/11/11/makingAHappyDeveloperHouse.html

The best way to predict the future is to prevent it.

—Alan Kay, confusedofcalcutta.com/2007/11/03/the-best-way-to-predict-the-future-is-to-prevent-it

[When] Americans can use the software and handsets of their choice, over open and competitive networks, they win.

—Eric Schmidt, www.forbes.com/intelligentinfrastructure/2007/07/20/google-wireless-fcc-tech-infrastructure-cx_bc_0720google.html?partner=links

Mobile Linux Groundswell

The rumble you hear in the ground is embedded Linux moving into the vast mobile space—filling it not just with more closed devices built on open platforms, but with truly open devices that can connect anybody with anybody or anything, any way they like—and to write and use whatever programs they like, without having to limit usage to the insides of carriers' and equipment-makers' walled gardens.

That's where several harbingers are pointed.

First, there's Nokia's Linux-based N series tablets, now in their third generation with the N810. What matters isn't just that the N line keeps improving (and the Maemo development community right along with it), but that Nokia will sell its billionth phone sometime soon—yet it still needs to cripple most of those devices to fit the customer-containing purposes of its carrier partners' walled gardens. The goal here is to make the Net mobile, and it won't happen until

Linux-modeled development methods and values prevail.

Next, there's Linux mobile phone work. In November 2007, Google announced both the Linux-based Android mobile phone platform and the Open Handset Alliance; both add momentum to established open Linux handset development efforts by MontaVista, OpenMoko, Trolltech and others. (Of course, there also are plenty of closed handsets with Linux inside, but those play a lesser role in this movement.)

Next, there's the XO Linux laptop from One Laptop Per Child (OLPC), which finally has begun shipping. The inventive little device is coming in at higher than the originally projected \$100 price, but it still has plenty of promise and breaks new technical and cultural ground.

Then, there are efforts, such as the Xandros-based ASUS Eee PC (3EPC) 701 "ultra-mobile" laptop and development platforms from Intel and Via for Mobile Internet Devices (MIDs) and Ultra-Mobile



Devices (UMDs), respectively, all of which paved the path toward wanton mobile Linux device development.

Together these suggest that Linux will win in the palm and the ear before it wins in the lap. But, those wins will be bigger anyway, as handheld mobile devices outnumber desktops and laptops by a wide margin. (See this month's *LJ* Index for some of the latest numbers.)

—DOC SEARLS

Searching for Consistencies

We ran across a list of Google search results from June 29, 2004, and thought it might be interesting to compare them with searches on two days in November 2007.

Worth noting here is that the November 8, 2007, searches were done in London, but at google.us, to avoid the google.uk site (though tested results were essentially the same when I tried both). The November 13, 2007, searches were done in the US—Boston, to be exact.

Still, the widely varying results make one wonder why the largest deployer of Linux in the world (as well as the world's leading search engine) can't yield more consistent, if not useful, numbers.

Google Search Comparison

	June 29, 2004	November 8, 2007	November 13, 2007
Linux	108,000,000	234,000,000	36,300,000
Free Software	7,230,000	72,800,000	99,000,000
GPL	14,000,000	120,000,000	13,100,000
GCC	11,800,000	30,700,000	3,640,000
Open Source	9,130,000	412,000,000	74,000,000
SUSE	10,800,000	32,400,000	3,520,000
Debian	24,300,000	64,000,000	7,150,000
Ubuntu	N/A	72,800,000	9,300,000
Red Hat	8,710,000	97,900,000	12,800,000

—DOC SEARLS



REUVEN M. LERNER

Integrating with Facebook Data

Writing a Facebook application means integrating your database with information kept on Facebook. Here's how you can combine the two quickly and easily.

For the past few months, we've been looking at the Facebook API, which makes it possible to integrate third-party applications into the popular social-networking site. Facebook is remarkable to users for the number of people already using it, as well as for the rapid pace at which new people are joining. But, it also is remarkable for software developers, who suddenly have been given access to a large number of users, into whose day-to-day Web experience they can add their own applications.

The nature of Facebook means that most developers are writing applications that are more frivolous than not. Thus, it's easy to find name-that-celebrity games, extensions to built-in Facebook functionality (such as, "SuperWall") and various applications that ask questions, match people together and so forth. I expect we eventually will see some more serious applications created with the Facebook API, but that depends on the developer community. I would argue that the continued growth of Facebook applications depends on the ability of developers to profit from their work, but that is a business issue, rather than a technical one.

Regardless of what your application does, it probably will be quite boring if you cannot keep track of information about your users. This might strike you as strange—after all, if you are writing a Facebook application, shouldn't Facebook take care of the storage for you?

The answer is no. Although Facebook handles user authentication, gives you the ability to deploy your application within the Facebook site and even provides access to certain data about the currently logged-in user, it does not store data on your behalf. This means any data you want to store must be kept on your own server, in your own database.

This month, I explain how to create a simple application on Facebook that allows us to retrieve data from a user's Facebook profile or from our local relational database seamlessly. The key to this is the user's Facebook ID, which we will integrate into our own user database. Retrieving information about our user, or about any of their friends, will require a bit of thinking about where the data is stored. However, you will soon see that mixing data from different

sources is not as difficult as it might sound at first, and it can lead to far more interesting applications.

Creating the Application

Our application is going to be simple—a Facebook version of the famous "Hello, world" program that is the first lesson in oh-so-many books and classes. However, we'll add two simple twists: first, we will display the number of times that the user has visited our application to date. (So, on your fifth visit, you will be reminded that this is your fifth visit.) Moreover, you will be told how many times each of your friends has visited the site.

In a normal Web/database application, this would be quite straightforward. First, we would define a database to keep track of users, friends and visits. Then, we would write some code to keep track of logins. Finally, we would create a page that displayed the result of a join between the various pages to show when people had last visited. For example, we could structure our database tables like this:

```
CREATE TABLE People (
    id SERIAL NOT NULL,
    email_address TEXT NOT NULL,
    encrypted_password TEXT NOT NULL,

    PRIMARY KEY(id),
    UNIQUE(email_address)
);

CREATE TABLE Visits (
    person_id INTEGER NOT NULL REFERENCES People,
    visited_at TIMESTAMP NOT NULL DEFAULT NOW(),

    UNIQUE(person_id, visited_at)
);

CREATE TABLE Friends (
    person_id INTEGER NOT NULL REFERENCES People,
    friend_id INTEGER NOT NULL REFERENCES People,

    UNIQUE(person_id, friend_id),
    CHECK(person_id <> friend_id)
);
```

Our first table, People, contains only a small number of columns, probably fewer than you would want in a real system. We keep track of the users' primary key (id), their e-mail addresses (which double as their login) and their encrypted passwords.

We keep track of each visit someone makes to our site in a separate table. We don't need to do this; it would be a bit easier and faster to have a number_of_visits column in the People table and then just increment that with each visit. But, keeping track of each visit means we have more flexibility in the future, from collecting usage statistics to stopping people from using our system too much.

Finally, we indicate friendship in our Friends table. Keeping track of friends is a slightly tricky business, because you want to assume that if A is a friend to B, then B also is a friend to A. We could do this, but it's easier in my book simply to enter two rows in the database, one for each direction. To retrieve the friends of A, whose ID is 1, we look in the Friends table for all of the values of friend_id where person_id = 1.

Integrating with Facebook

All of this seems pretty reasonable and straightforward, and it isn't hard to implement in any modern Web framework. But, if we want to implement the same functionality in a Facebook application, we have to consider that about half the database we just defined is going to be unnecessary. We don't need to worry about the Friends table, because that's something Facebook does quite well. And, we don't really need to worry about the People table either, as Facebook handles logins and authentication.

At the same time, we obviously can't use only the Friends table by itself. We need it to point to something, somewhere, so we can associate a visit with a user. How do we do that?

The answer is that instead of storing the users' information, we store their Facebook user IDs. Our People table, thus, will look like this:

```
CREATE TABLE People (  
  id SERIAL NOT NULL,  
  facebook_session_key TEXT NOT NULL,
```



Full 100 Mbps
Port

Full Duplex
\$2,000/month

Order Today!

email sales@he.net or call 510.580.4190

he.net/ip_transit

Colocation
Full Cabinet
42U
\$400/month

IP Transit
Gigabit Ethernet

BGP+IPv6+IPv4
Special
\$5/Mbps


```

    facebook_uid    TEXT    NOT NULL,

    PRIMARY KEY(id)
);

```

By storing the Facebook information in our database, we effectively hook our `id` column to what Facebook provides. But, how will we use this link?

The answer is that we don't really have to, if we use a plugin that handles the underlying details for us. I have used RFacebook for the past few months; this is a plugin for Ruby on Rails that makes it fairly easy to create a Facebook application using Rails. First, I create my models using the generate script that comes with Rails:

```

./script/generate model person
↳facebook_session_key:string facebook_uid:string

```

This creates a new model—that is, a Ruby class that represents a database table—called `person.rb`. Although this script doesn't create the model directly, it does create a migration file that defines our database table in Ruby:

```

class CreatePeople < ActiveRecord::Migration
  def self.up
    create_table :people do |t|
      t.column :facebook_session_key, :string
      t.column :facebook_uid, :string
    end
  end

  def self.down
    drop_table :people
  end
end

```

Assuming that our database is all set up, we can run the migration using the built-in rake tool (think make, but in Ruby):

```
rake db:migrate
```

The output tells us a bit of what's going on:

```

== CreatePeople: migrating =====
-- create_table(:people)
NOTICE: CREATE TABLE will create
       implicit sequence "people_id_seq" for serial column "people.id"
NOTICE: CREATE TABLE / PRIMARY KEY will create
       implicit index "people_pkey" for table "people"
-> 0.1939s
== CreatePeople: migrated (0.1944s) =====

```

The advantage of using rake and migrations is that we can modify our migrations file, change our

database definitions, and move forward and backward in time through our database design. Migrations mean that you can keep track of your changes to the database and automatically upgrade (or downgrade) the database to the latest version without losing data. And, sure enough, if we look at our database, we see that it has three columns, just as we defined.

Next, we create another model for our visits table:

```

./script/generate model visit person_id:integer
↳visited_at:timestamp

```

We migrate the database to the latest version:

```
rake db:migrate
```

And, sure enough, we have a visits table with a `person_id` column. Unfortunately, because Rails migrations are written in a generic language, there isn't any built-in support for handling foreign keys and other referential integrity checks. So, the table, as we defined it above, would have indicated that `person_id` always must point to a value.

Also note that the default model-generation script allows null values. We could go into the migration file and change this, but we will ignore it for now.

Mixing It Together

Now that we have a place for the Facebook information in our People table, we need to tell Rails to put it there. We do this by adding the line:

```
acts_as_facebook_user
```

in the model file, `person.rb`. By default, it will be almost empty, indicating that we will use ActiveRecord's defaults to work with our database table via Ruby:

```

class Person < ActiveRecord::Base
end

```

When we're done, our class will look like this:

```

class Person < ActiveRecord::Base
  acts_as_facebook_user
end

```

In our controller file (which I'm sneakily reusing from what we did last month, modifying the facebook method in the hello controller), I've modified the method to read:

```

def facebook
  render :text => "hi"
end

```

Because my application is called `rmlljtf`, I can go to the following URL: `http://apps.facebook.com/rmlljtf/`, and see my "hi" at the top of the page. After loading this page, I then look at my People table and find...that nothing has changed. After all, I told the system to create the table, but I didn't actually do anything with it! For that to happen, I need to use the built-in `fbsession` object, which gives me access to Facebook information. I then can say:

```
def facebook
  person = Person.find_or_create_by_facebook_session(fbsession)
  render :text => "hi"
end
```

And, sure enough, reloading the page creates a row in our People table.

Next, I modify my method to add a row to our visits table. I can do that with the following:

```
def facebook
  person = Person.find_or_create_by_facebook_session(fbsession)

  Visit.create(:person_id => person.id,
```

```
:visited_at => Time.now()).save!
```

```
  render :text => "hi"
end
```

Once I've modified the facebook method in this way, each visit to the site does indeed add another row to the visits table.

Now we should produce some output, indicating exactly how many visits the person has made to the site. For this, we create a view (`facebook.rhtml`), which can display things more easily:

```
<p>This is your <%= @number_of_visits.ordinalize %> visit.</p>
```

This short view displays the instance variable `@number_of_visits` and puts it into ordinal form, which is convenient. However, this means we need to set `@number_of_visits` in the facebook method. We do this by adding the following line:

```
@number_of_visits =
  Visit.count(:conditions => ["person_id = ?", person.id])
```

Expert Included.



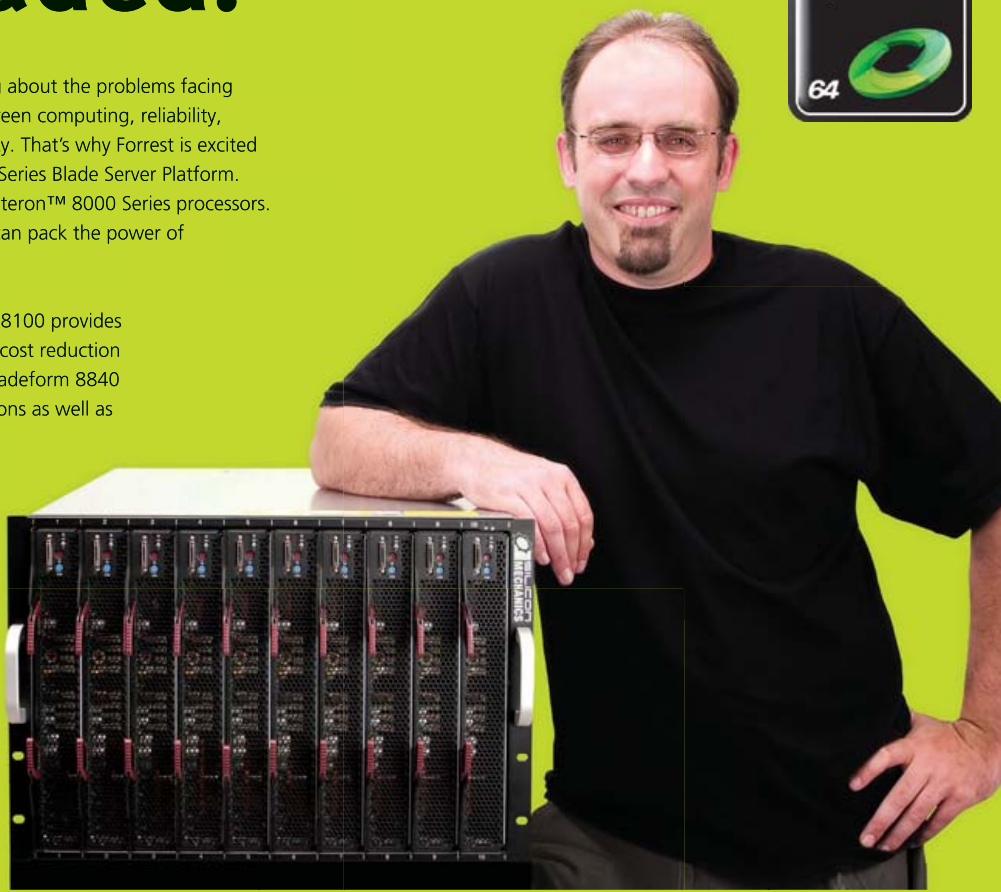
Forrest and the Product Development team never stop thinking about the problems facing today's IT administrators. Power efficiency, compute density, green computing, reliability, and serviceability are just some of the things he makes a priority. That's why Forrest is excited about the new Bladeform 8840 Blade for the Bladeform 8100 Series Blade Server Platform. Each Bladeform 8840 Blade supports four Quad-Core AMD Opteron™ 8000 Series processors. With 16 cores per blade and 10 blades per 7U enclosure, you can pack the power of 960 cores in a 42U rack.

At the same time, Forrest is very impressed that the Bladeform 8100 provides 90%+ high-efficiency redundant power supplies for operating cost reduction and earth-friendly computing. The Bladeform 8100 with the Bladeform 8840 Blade is a perfect choice for mission-critical enterprise applications as well as scale-out and high performance computing environments.

When you partner with Silicon Mechanics, you get more than a high-efficiency AMD solution—you get an expert like Forrest.



Silicon Mechanics and the Silicon Mechanics logo are registered trademarks of Silicon Mechanics, Inc. AMD, the AMD Arrow logo, AMD Opteron, and combinations thereof, are trademarks of Advanced Micro Devices, Inc.



In other words, we grab the current user's ID. We then use that ID, along with a built-in ActiveRecord value, to sum up the number of visits the user has made to the site.

Finally, it's time to introduce the Facebook magic. We know, from last month, that we can display the current user's Facebook friends without too much trouble; we use `fbsession` to grab a list of friends (and specific pieces of information about those friends), and then iterate over them, displaying them however we like.

Now, we do the same thing, but we also create a hash, `@friends_visits`, in which the key will be the Facebook user ID (`uid`), and the value will be the number of visits by that person. We give our hash a default value of 0, in case we try to retrieve a key that does not exist. We also use a bit of exception handling to ensure that we can handle null results. The final version of the facebook method looks like this:

```
def facebook
  person = Person.find_or_create_by_facebook_session(fbsession)
  Visit.create(:person_id => person.id,
              :visited_at => Time.now())

  # Count the number of visits
  @number_of_visits =
    Visit.count(:conditions => ["person_id = ?", person.id])

  @friend_uids = fbsession.friends_get.uid_list
```

Resources

Facebook developer information is at developers.facebook.com. This includes documentation, a wiki and many code examples. One article on the wiki specifically addresses Ruby development, at wiki.developers.facebook.com/index.php/Using_Ruby_on_Rails_with_Facebook_Platform.

Ruby on Rails can be downloaded from rubyonrails.com. Of course, Rails is written in the Ruby language, which almost certainly is included in your distribution, but it also can be downloaded from www.ruby-lang.org.

The RFacebook gem for Ruby, and the companion RFacebook plugin for Rails developers, can be retrieved from rfacebook.rubyforge.org.

Hpricot, written by the prolific Ruby programmer "why the lucky stiff", is at code.whytheluckystiff.net/hpricot. I have found it to be useful in many Ruby programs I've written, but it is especially useful in the context of RFacebook, given the central role of XML and the Facepricot extension.

Finally, Chad Fowler, a well-known Ruby developer, has developed a different Rails plugin (Facebooker) for working with Facebook. You can download the code, as well as learn more about the design principles behind his plugin, at www.chadfowler.com/2007/9/5/writing-apis-to-wrap-apis.

```
# Get info about friends from Facebook
@friends_info =
  fbsession.users_getInfo(:uids => @friend_uids,
                        :fields => ["first_name", "last_name"])

# Keep track of friend visits to the site
@friends_visits = Hash.new(0)
@friends_info.user_list.each do |userInfo|
  begin
    friend = Person.find_by_facebook_uid(userInfo.uid)

    @friends_visits[userInfo.uid] =
      Visit.count(:conditions => ["person_id = ?", friend.id])
  rescue
    next
  end
end
```

In other words, we grab friend information via `fbsession`. We then iterate over each friend, getting its Facebook uid. With that UID—which we have in our `People` table, in the `facebook_uid` column—we can get the person's database ID, and then use that to find the person's number of visits.

With this in place, we can rewrite the view as follows to include friend information:

```
<p>This is your <%= @number_of_visits.ordinalize %> visit.</p>

<% @friends_info.user_list.each do |userInfo| %>
<ul>
  <li><fb:name uid="<%= userInfo.uid -%>" target="_blank" />
    <fb:profile-pic uid="<%= userInfo.uid -%>" linked="true" />
    <%= @friends_visits[userInfo.uid] %> visit(s)</li>
</ul>
<% end %>
```

Sure enough, when you visit the page, it tells you how many times you have visited, as well as how many times each friend has visited.

Conclusion

Facebook's API gives us the opportunity to think about how we can structure an application that doesn't have access to some of the data. This application doesn't have any authentication information about the users, and it can get only particular pieces of data about them. But, because we have an `id` column, we can use it to store data on our local server and then join that data with what comes from Facebook seamlessly. ■

Reuven M. Lerner, a longtime Web/database developer and consultant, is a PhD candidate in learning sciences at Northwestern University, studying on-line learning communities. He recently returned (with his wife and three children) to their home in Modi'in, Israel, after four years in the Chicago area.



Are you Shocked

by the
high cost
of iSCSI &
Fibre Channel
SAN storage?

AoE is the answer!

ATA-over-Ethernet = **Fast, Reliable, Simple** storage.

www.coraid.com



EtherDrive® SRxxxx

- Fast & Flexible RAID appliances with slots for hot swap SATA disks
- Check out our full line of EtherDrive® Storage and VirtualStorage Appliances and NAS Gateways



1. Fast 10 Gigabit Ethernet Storage without the TCP/IP overhead!
2. Unlimited expandability, at the lowest possible price point!!
3. You want more storage...you just buy more disks – it's that simple!!!

Visit us at www.coraid.com



1.706.548.7200

The Linux Storage People

www.coraid.com



MARCEL GAGNÉ

It's a Virtual World

Virtualization is such a strange world, particularly when applied to an otherwise, sometimes equally strange world.

Don't tell me, François. Let me guess. Okay, I give up. Tell me. Why do you have all those wires coming out of your head? What do you mean you are trying to create a virtual François? You think the transhumanists are right, so you've decided to speed up the process by re-creating yourself in your Linux system? Aside from the obvious jokes about one of you being enough, I'm not sure this is really doable. *Quoi?* Not just you? You're trying to digitize the entire planet? *Mon Dieu!* Even if that were technically possible, we don't have anywhere near the computing power in this restaurant to do such a thing, at least not on the scale you are contemplating. At best, we can do a little global desktop virtualization. There's no more time to talk about this now, *mon ami*; our guests will be here any moment. Look, they are approaching as we speak. Quickly, François, prepare the tables.

Hello, everyone, and welcome to *Chez Marcel*. It is here that you will find the best in Linux and open-source software, guaranteed to enhance your Linux experience, served with exceptional wines to please and titillate your palate. My faithful waiter will provide us with the wine momentarily. François! To the wine cellar. Over in the south wing, you'll find three cases of 2005 Cloudy Bay Chardonnay from New Zealand. *Vite!*

While he is off fetching the wine, I should tell you that François has some aspirations that are somewhat loosely reflected in the software you'll see today. He wants to create machine copies of all of us; however, I think he'll have to settle for the world on his desktop.

Good to have you back, François. Please, pour for our guests. Enjoy, *mes amis*.

By now, you know that the new KDE 4 desktop has been released and with it, there are some cool new applications. One of these, Torsten Rahn and Inge Wallin's very cool Marble, is part of the KDE educational suite. Some of you, like myself, may remember a kid's show that ran from 1974 to 1983 called *Big Blue Marble* and the theme song that accompanied it—the opening line was “The Earth's a Big Blue Marble when you see it from out there.” Well, Marble is described as “a generic geographical map widget and framework” in the sense that it can be used by different KDE 4 applications, but to you and me, and for today's menu, it's an interactive globe of the planet we live on and a 3-D atlas. As with a physical globe, the Earth is represented as a sphere, and you can spin it around, explore the surface and learn about our planet. It's a fast, fun application (Figure 1), and you can find the latest on Marble at edu.kde.org/marble.

Some people invariably will compare this to Google Earth, but unlike Google Earth, Marble is extremely lightweight and doesn't require 3-D acceleration. Consequently, it runs beautifully on more modest hardware. When you start Marble the first time, it creates the map and prepares the data. While this process takes place, a progress bar appears on the screen (Figure 2). This only happens once.

Marble's interface is easy to use. On the right, in

Figure 1. Marble puts a 3-D virtual globe on your desktop.

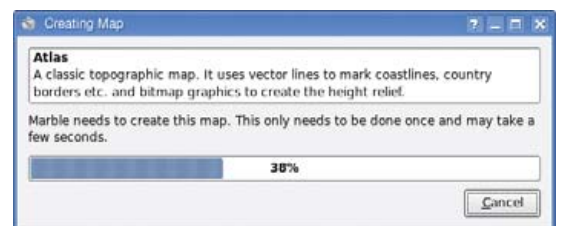
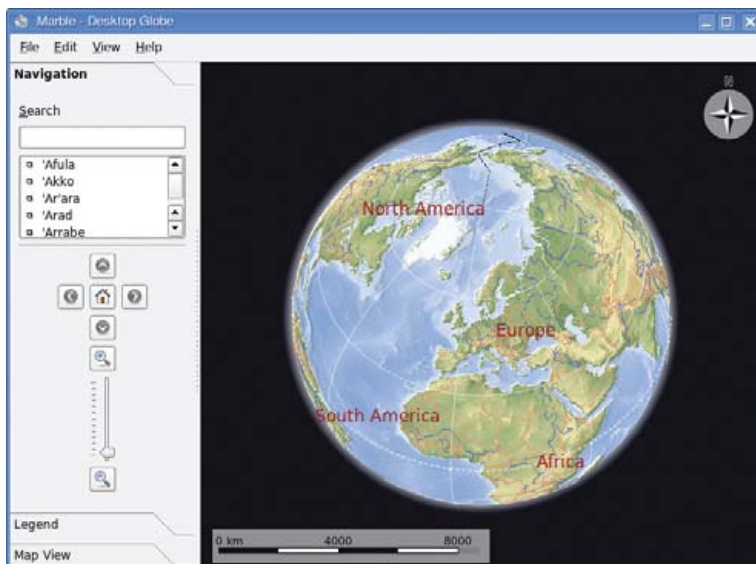


Figure 2. Marble starts up very quickly, except the first time when it builds its map.



Figure 3. Navigating to a faraway place is as easy as typing its name in the search field.

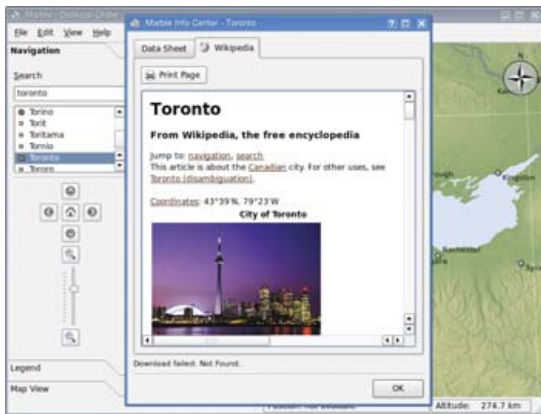


Figure 4. Marble links to Wikipedia to provide details on the city you're exploring.

the main part of the program window, your globe beckons you. In the left-hand sidebar, the main tab (one of three) is labeled Navigation with a list of cities right below the search field. Marble will take you to a city in a flash (Figure 3). You can scroll around the globe simply by clicking and dragging from any point on the surface, or you can type the name of a city in the search field. If you have a scroll wheel, you can zoom in on the surface view by rolling your scroll wheel up and zoom out by scrolling down. Alternatively, you also can click the arrow keys in the navigation pane and use the zoom controls there.

There are, as I mentioned, three tabs. Below the Navigation tab is another labeled Legend and one more labeled Map View. The symbols used to identify various features on the map are all listed under the Legend tab. The Map View tab gives you access to other surface views. The default is the atlas view you see in the screen-

shots here, but Marble also can use a plain map, a realistic satellite image and the famous "earth at night" view. Switching from one to the other takes only a second.

Exploring the world with Marble is great fun and educational as well. To find out more about a city, left-click on that city, then select its name from the pop-up window. When you click the name, Marble opens a link to the Wikipedia entry for that city. It's a great way to pay a virtual visit to some exotic places like Toronto, Canada (Figure 4); London, England; Paris, France; Milan, Italy; or even Snellville, Georgia.

Some great additional features also are fun to explore. For instance, right-click on any point and select Add Measure Point. A small cross appears. Move to another location on the globe and add another measuring point. Doing this, I discovered that it's 6,024.89 kilometers from Madrid to Toronto (as the proverbial crow flies). I was also happy to know that it's roughly 760 kilometers from the North Pole to the North Magnetic Pole.

You also can define a home location by right-clicking on a city and selecting Set Home Location. Marble remembers the view, including the scale you were using when you set the home point.

Closing time approaches, *mes amis*, and although it saddens me that time passes so quickly, this gives me an opportunity to introduce one final virtual view of the world—one that specifically deals with time—Matthias Hoelzer-Kluepfel's KWorldClock. Essentially, KWorldClock is a graphical application that shows what parts of the world are in sunlight and what parts are in darkness, in real time. It's also a clock that shows the time in a number of locations around the world (Figure 5).

The default map has several points visible across the surface of the planet. The terminator moves across the planet as the minutes tick by, so you can

Figure 5. KWorldClock shows the passing of time in a satellite's view.

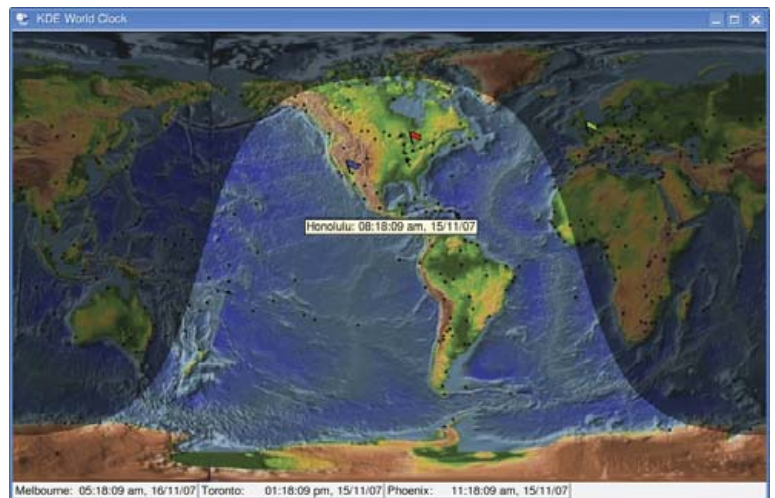




Figure 6. Add markers, flags, city clocks or change the look of the map.

see who not to call long distance in the evening. As you move across the map, you'll see pop-ups showing the date and time for that location. Right-click on a location, select Clocks, then Add, and a clock for that location appears at the bottom of the display window. In this way, you can keep track of the date and time from many different world locations. You also can flag locations of interest (so you can find them again) by right-clicking on a location and selecting Flags from the menu (Figure 6). This pins a colored flag wherever you choose.

Notice as well the Map Theme option on that pop-up menu. By default, KWorldClock comes with two maps: a flat atlas-style map and a colored map simulating elevation. Other map styles are available from the kdeartwork-misc package. These include a rainfall map, a relative altitude map and more. To see available themes, run `kworldclock --themes` from the command line. Passing the `--help` option shows other parameters available to the program.

Here's a cool trick. KDE desktop users can use a full-screen version of KWorldClock as their desktop wallpaper. Right-click on your desktop

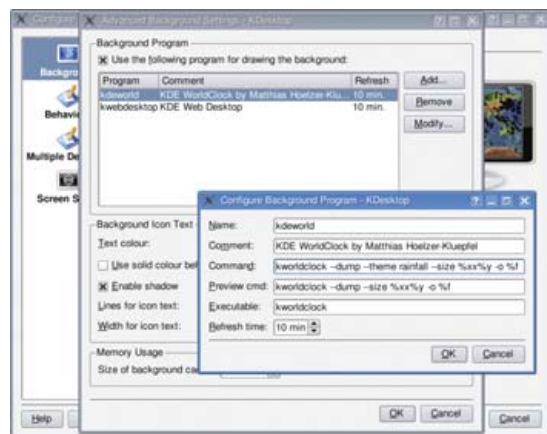


Figure 7. The KWorldClock program makes for a great dynamic wallpaper.

KDE desktop users can use a full-screen version of KWorldClock as their desktop wallpaper.

and select Configure Desktop. When the configuration window opens, click on Background in the left-hand sidebar. Click the No picture radio button for your background image, then click the Advanced Options button. Next, click the Use the following program for drawing the background box, select `kdeworld` from the list, and click the Modify button (Figure 7).

This also is a good time to change various settings, such as the display theme (for example, `--theme rainfall`). The default refresh time is every ten minutes, which probably is a sane setting, but you certainly can change it if you like. When you are happy with your settings, click OK to close the various dialogs, and you'll have a nice, dynamic background.

As the great shadow of night moves across the planet, we are reminded that it is indeed closing time here at *Chez Marcel*. Before we all head home, I'm sure we can convince François to refill everyone's glass one more time. And, perhaps, someday, if he manages to digitize himself, an electronic François will do the job for us. Personally, despite his many quirks, I prefer the real François, as I'm sure you all do. Raise your glasses, *mes amis*, and let us all drink to one another's health. *A votre santé! Bon appétit!* ■

Marcel Gagné is an award-winning writer living in Waterloo, Ontario. He is the author of the all-new *Moving to Free Software*, his sixth book from Addison-Wesley. He also makes regular television appearances as Call for Help's Linux guy. Marcel is also a pilot, a past Top-40 disc jockey, writes science fiction and fantasy, and folds a mean Origami T-Rex. He can be reached via e-mail at mggagne@salmar.com. You can discover lots of other things (including great Wine links) from his Web site at www.marcelgagne.com.

Resources

KDE World Clock (KWorldClock): docs.kde.org/stable/en/kde toys/kworldclock/index.html

Marble: edu.kde.org/marble

Marcel's Web Site: www.marcelgagne.com

The WFTL-LUG, Marcel's Online Linux User Group: www.marcelgagne.com/wftllugform.html

USENIX

Upcoming Conferences

2008 LINUX STORAGE & FILESYSTEM WORKSHOP

Co-located with FAST '08

FEBRUARY 25–26, 2008, SAN JOSE, CA, USA

<http://www.usenix.org/lsf08>

6TH USENIX CONFERENCE ON FILE AND STORAGE TECHNOLOGIES (FAST '08)

Sponsored by USENIX in cooperation with ACM SIGOPS, IEEE Mass Storage Systems Technical Committee (MSSTC), and IEEE TCOS

FEBRUARY 26–29, 2008, SAN JOSE, CA, USA

<http://www.usenix.org/fast08>

2008 ACM INTERNATIONAL CONFERENCE ON VIRTUAL EXECUTION ENVIRONMENTS (VEE '08)

Sponsored by ACM SIGPLAN in cooperation with USENIX

MARCH 5–7, 2008, SEATTLE, WA, USA

<http://vee08.cs.tcd.ie>

USABILITY, PSYCHOLOGY, AND SECURITY 2008

Co-located with NSDI '08

APRIL 14, 2008, SAN FRANCISCO, CA, USA

<http://www.usenix.org/upsec08>

1ST USENIX WORKSHOP ON LARGE-SCALE EXPLOITS AND EMERGENT THREATS (LEET '08)

Botnets, Spyware, Worms, and More

Co-located with NSDI '08

APRIL 15, 2008, SAN FRANCISCO, CA, USA

<http://www.usenix.org/leet08>

Paper submissions due: February 11, 2008

5TH USENIX SYMPOSIUM ON NETWORKED SYSTEMS DESIGN AND IMPLEMENTATION (NSDI '08)

Sponsored by USENIX in cooperation with ACM SIGCOMM and ACM SIGOPS

APRIL 16–18, 2008, SAN FRANCISCO, CA, USA

<http://www.usenix.org/nsdi08>

THE SIXTH INTERNATIONAL CONFERENCE ON MOBILE SYSTEMS, APPLICATIONS, AND SERVICES (MOBISYS 2008)

Jointly sponsored by ACM SIGMOBILE and USENIX

JUNE 10–13, 2008, BRECKENRIDGE, CO, USA

<http://www.sigmobile.org/mobisys/2008/>

2008 USENIX ANNUAL TECHNICAL CONFERENCE

JUNE 22–27, 2008, BOSTON, MA, USA

<http://www.usenix.org/usenix08>

2008 USENIX/ACCURATE ELECTRONIC VOTING TECHNOLOGY WORKSHOP (EVT '08)

Co-located with USENIX Security '08

JULY 28–29, 2008, SAN JOSE, CA, USA

3RD USENIX WORKSHOP ON HOT TOPICS IN SECURITY (HOTSEC '08)

Co-located with USENIX Security '08

JULY 29, 2008, SAN JOSE, CA, USA

17TH USENIX SECURITY SYMPOSIUM

JULY 28–AUGUST 1, 2008, SAN JOSE, CA, USA

<http://www.usenix.org/sec08>

22ND LARGE INSTALLATION SYSTEM ADMINISTRATION CONFERENCE (LISA '08)

Sponsored by USENIX and SAGE

NOVEMBER 9–14, 2008, SAN DIEGO, CA, USA

8TH USENIX SYMPOSIUM ON OPERATING SYSTEMS DESIGN AND IMPLEMENTATION (OSDI '08)

DECEMBER 8–10, 2008, SAN DIEGO, CA, USA

<http://www.usenix.org/osdi08>

Paper submissions due: May 8, 2008

USENIX: THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

TECHNICAL SESSIONS AND TRAINING PROGRAM INFORMATION AND HOW TO REGISTER ARE AVAILABLE ONLINE AND FROM THE USENIX OFFICE:

<http://www.usenix.org/events> | Email: conference@usenix.org | Tel: +1.510.528.8649 | Fax: +1.510.548.5738



DAVE TAYLOR

Solve: a Command-Line Calculator Redux

Dave completes his explanation of writing a helpful interactive command-line calculator as a shell script.

Ooops! Two months ago, I started exploring how you can write a simple but quite helpful interactive command-line calculator as a shell script and ended the column with “Next month, we’ll dig into useful refinements and make it a full-blown addition to our Linux toolkit. See you then!”

Unfortunately, last month, I got sidetracked with the movie *The Number 23* and started another script looking at how to do numerology within the shell scripting environment. You’d think I was a typical programmer or something, being sidetracked and losing a thread by picking up another one. It reminds me of those glorious startup days from the late 1990s too, but that’s an entirely different story.

Anyway, numerology can wait another month. This column, I’d like to complete the command-line calculator because, well, because it’s so darn useful and simultaneously astonishing that there isn’t a decent command-line calculator in Linux after all these years. I mean, really!

When Last We Met

It was a while back, so let me remind you that the wicked short script to give you the rudimentary calculator is this:

```
#!/bin/sh

bc << EOF
scale=4
$@
quit
EOF
```

That’s it. Name it `solve.sh`, for example, and you can test it, as shown here:

```
$ sh solve.sh 1+3
4
$ sh solve.sh 11/7
1.5714
```

It’s easy enough to alias `solve` to the shell command too:

```
alias solve="sh solve.sh"
```

Or, better:

```
alias solve="sh ~/bin/solve.sh"
```

As that’ll work regardless of where you are in the filesystem (location-dependent commands are a typical shell gaffe).

What I’d really like, however, is to be able to go into a “solve” mode where anything I type automatically is assumed to be a mathematical equation, rather than have to type `solve` each time.

Rapping about Wrappers

We’ve talked about shell script wrappers in the past, so you should recall this basic structure:

```
while read userInput
do
    echo "you entered $userInput"
done
```

That’s too crude to use as of yet, but we easily can add a prompt so that it looks like a real program:

```
echo -n "solve"
while read expression
do
    echo "you entered $expression"
    echo -n "solve: "
done
```

Look good? Actually, it’s not. There’s a subtle error here, one that’s another common scripting mistake. The problem is that there are two `echo` commands in Linux: one that’s the built-in capability of the shell itself, and one that’s a separate command located in `/bin`. This is important because the built-in `echo` doesn’t know what the `-n` flag does, but the `/bin/echo` command does. A tiny tweak, and we’re ready to test it:

```
/bin/echo -n "solve: "

while read expression
do
    echo "you entered $expression"
    /bin/echo -n "solve: "
```


done

Let's see what happens:

```
solve: 1+1
you entered 1+1
solve: ^D
```

That's more like it.

What we really want though, is a script that's smart enough to recognize whether you've specified parameters on the command line. If you have, it solves that equation, and if you haven't, it drops you into the interactive mode.

That's surprisingly easy to accomplish by testing the \$# variable, which indicates how many arguments are given to the script. Want to see if it's greater than zero? Do this:

```
if [ $# -gt 0 ] ; then
```

One more refinement before I show you the script in its entirety: I want to have it quit if users type in

quit or exit, rather than force them to type ^D to indicate end of file on standard input (which causes the read statement to return false and the loop to end).

This is done with a simple string comparison test, which you'll recall is done with = (the -eq test is for numeric values). So, testing \$expression to see whether it is "quit" is easy:

```
if [ $expression = "quit" ] ; then
```

To make it a bit more bulletproof, it's actually better here to quote the variable name, so that if users enter a null string (simply press Return), the conditional test won't fail with an ugly error message:

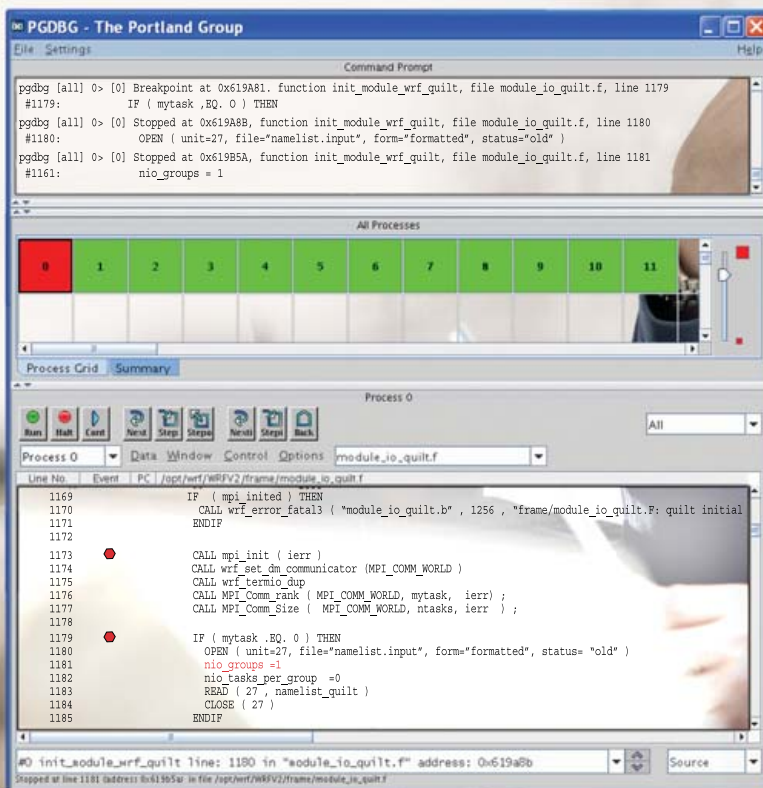
```
if [ "$expression" = "quit" ] ; then
```

Because I like to make my scripts flexible, I've also added exit as an alternative to quit, which easily is done with a slightly more complicated conditional test:

```
if [ "$expression" = "quit" -o
    "$expression" = "exit" ] ; then
```

PGI[®] CDK[™]

MPI Debugger/MPI Profiler

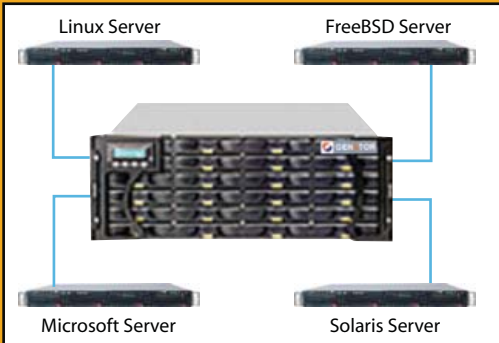


www.pgroup.com/cdk

The Portland Group, Inc. is an STMicroelectronics company. PGI and CDK are trademarks or registered trademarks of STMicroelectronics. Other brands and names are the property of their respective owners.



Linux - FreeBSD - x86 Solaris - MS etc.



GENSTOR STORAGE SOLUTIONS:

- Storage options - FC to SATA/SAS, FC to FC SAS to SAS/SATA, SCSI to SATA, SCSI to SCSI
- Exceptional Performance with Proven Reliability
- 24 TB in 4U with easy upgrade path
- Host Servers and Storage comes Pre-Configured with heterogeneous OS- Linux, * BSD, Solaris Microsoft etc.
- Fully redundant Storage solutions

Proven technology. Proven reliability

When you can't afford to take chances with your business data or productivity, rely on a GS-1245 Server powered by the Intel® Xeon® Processors.

Quad Core Woodcrest



2 Nodes & Up to 16 Cores - in 1U

Ideal for high density clustering in standard 1U form factor. Upto 16 Cores for high CPU needs. Easy to configure failover nodes.

Features:

- 1U rack-optimized chassis (1.75in.)
- Up to 2 Quad Core Intel® Xeon® Woodcrest per Node with 1333 MHz system bus
- Up to 16 Woodcrest Cores Per 1U rackspace

Servers :: Storage :: Appliances

Genstor Systems, Inc.

780 Montague Express. #604
San Jose, CA 95131

www.genstor.com

Email: sales@genstor.com

Phone: 1-877-25 SERVER
1-408-383-0120



The -o is the logical OR statement in a shell conditional test, but I have a feeling you've already figured that out.

The Full Script

Here's where the script stands at this point, in its entirety:

```
#!/bin/sh

if [ $# -gt 0 ] ; then
bc << EOF
scale=4
$@
quit
EOF
else
/bin/echo -n "solve: "

while read expression
do
if [ "$expression" = "quit" -o
"$expression" = "exit" ] ; then
exit 0
fi
bc << EOF
scale=4
$expression
quit
EOF
/bin/echo -n "solve: "
done

echo ""
echo "solved."
fi

exit 0
```

Neat and darn useful, I'd say. If I were to continue hacking on it, the next thing I would do is write a simple help page that I'd store in some library folder and display on entry of ? or help. It simply would explain the syntax of the expressions understood by bc (though as we're invoking bc iteratively, we can't have persistent variables and so forth, so unfortunately, this approach won't let us access the full power of the binary calculator).

To learn what type of sophisticated expressions you can enter, simply type man bc. Then, let that be your inspiration for further tweaks and mods to this script!

Next month, I'll go back to the numerology script and see what strange things we can ascertain about the apparently benign world around us. Remember, just because I got out of sequence, doesn't mean we're not out to get you, dear reader! ■

Dave Taylor is a 26-year veteran of UNIX, creator of The Elm Mail System, and most recently author of both the best-selling *Wicked Cool Shell Scripts* and *Teach Yourself Unix in 24 Hours*, among his 16 technical books. His main Web site is at www.intuitive.com, and he also offers up tech support at AskDaveTaylor.com.

LINUX JOURNAL



1994-2006 ARCHIVE

ISSUES 1-152 of *Linux Journal*

www.LinuxJournal.com/ArchiveCD

The 1994–2006 Archive CD,
back issues, and more!



KYLE RANKIN

A Little Spring Cleaning

If your filesystem has dust bunnies and clutter filling up your free space, check out these simple tips to track down and reclaim space from some common offenders.

No matter how big your hard drives are, at some point you're going to look at your storage and wonder where all the space went. Your /home directory is probably a good example. If you are like me, you don't always clean up after yourself or organize immediately after you download a file. Sure, I have directories for organizing my ISOs, my documents and my videos, but more often than not, my home directory becomes the digital equivalent of a junk drawer with a few tarballs here, an old distribution ISO there and PDF specs for hardware I no longer own. Although some of these files don't really take up space on the disk—it's more a matter of clutter—when I'm running out of storage, I'd like to find the files that take up the most space and decide how to deal with them quickly. This

The duck command works great to discover how the space is being used in your home directory, but if you are like me, your home directory is actually on a different partition from the root filesystem.

month, I introduce some of my favorite commands for locating space-wasting files on my system and follow up with common ways to clear some space.

Think Locally

First, let's start with file clutter in your main home directory. Although all major GUI file managers these days make it easy to sort a directory by size, because I'm focusing on command-line tips, let's cover how to find the largest files in the current directory via the old standby, ls. If you type:

```
$ ls -lSh
```

you'll get a list of all the files in your current directory sorted by size. Of course, if you have a lot of files in the directory, the files you most want to see are probably somewhere along the top of the list, so I

typically like to type:

```
$ ls -lSh | less
```

to view the list slowly starting at the top. Or, if I'm in a hurry, I type:

```
$ ls -lSh | head
```

to see only the top ten largest files. Now, this is pretty basic, but it's worth reviewing, as you'll use these commands over and over again to track down space-wasting files. Depending on how you structure your home directory, you probably won't find all the large files together. It's more likely that they are scattered into different subdirectories, so you then need to scan through your directory structure recursively, tally up the disk space used in each directory, and sort the output. Luckily, you don't have to resort to ls for this; du does the job quite nicely. For instance, one common use for du that I see referenced a lot is the following:

```
$ du -sh *
```

This scans through all the subdirectories you list as arguments (in this case, all the subdirectories within my current directory) and then lists them one by one with human-readable file sizes (the -h option converts the file sizes into megabytes, gigabytes and so forth, so it's easier to read). Here's some example output from that command:

```
456K  bin
28K   Default-Compiz
16K   h14070cdwcups-1.0.0-7.i386.deb
344K  h14070cdwlpri-1.0.0-7.i386.deb
27M   images
60K   LexmarkC750.ppd
850M  mail
```

Although you certainly could work with this information, it would be much easier if it were sorted. To do that, replace the -h argument with -k, and then pipe the output to sort:

```
$ du -sk * | sort -n
```

```

16      hl4070cdwcups-1.0.0-7.i386.deb
28      Default-Compiz
60      LexmarkC750.ppd
344     hl4070cdwlp-1.0.0-7.i386.deb
456     bin
10224   writing
26948   images
869588  mail

119236  ./mail/example.net/sent-mail-2004
119236  ./mail/example.net/sent-mail-2004/cur
869852  ./mail
869852  ./mail/example.net
1064100 .
1064100 total

```

This works better, because now I can see that my local e-mail cache is taking up the bulk of the storage; however, next I would need to change to the mail directory and run the command again, over and over, until I narrow it down to the subdirectory that has the large files. That's why I normally skip the above commands and go straight for what I affectionately call the duck command:

```

$ du -ck | sort -n
. . .
87704  ./mozilla
87704  ./mozilla/firefox

```

The `-c` option essentially recurses into each sub-directory like before, except it keeps a running tally of the space used by each subdirectory down the tree, not just the first level of directories. When it reports its findings, it might list the same top-level directory multiple times. This makes it easy to drill down to the actual directory that consumes the most space, which in this example seems to be `./mail/example.net/sent-mail-2004/cur`. If I wanted to clean up files there, I could `cd` to that directory and then run the `ls` commands I used above to see which files used the most space.

Act Globally

The duck command works great to discover how the space is being used in your home directory, but

Rugged Mobile Linux Ready for Everything



Introducing the TDS Nomad® 800 Series

The TDS Nomad begins where other handhelds end. With integrated GPS, 802.11, camera and bar code scanner, the Nomad packs in plenty of features.

Standard Features:

- Ultra-Rugged Design (IP67, MIL-STD-810F testing)
- Linux® 2.6 with Qtopia® or X11/GTK+
- Marvell® PXA320 at 806 MHz
- 128 MB RAM
- 1 GB NAND Flash
- Integrated Bluetooth®, 802.11g, GPS
- Sunlight-Visible VGA Screen (640x480)
- CompactFlash®, SDIO® Slots
- Available in Yellow, Olive Drab, and Gray

Optional Features:

- Integrated 2 Megapixel Camera
- Integrated Bar Code Scanning



Equipping the Mobile User

1:724.452.9366 www.sdgsystems.com/rugged

if you are like me, your home directory is actually on a different partition from the root filesystem. If root is filling up, you still can use the duck command (with a slight tweak) to see which directories consume the most space. You need root privileges to scan all the directories in your root filesystem, so use either `su` or `sudo -s` (depending on how you get root permissions) before the duck command:

```
# cd /
# du -ckx | sort -n
. . .
243920 ./usr/lib/openoffice
277600 ./var/cache/apt
296376 ./var/cache
475144 ./var
952096 ./usr/share
1099264 ./usr/lib
2259332 ./usr
2908804 .
2908804 total
```

The extra `-x` argument I added above tells `du` to stay on one filesystem—in this case, the root filesystem. Otherwise, if you don't specify `-x` and you have /home or other directories on different filesystems,

You would be amazed how far you can compress incredibly large log files if you haven't tried it before.

`du` will scan through those partitions as well, so you ultimately will have to skip them out as you scan through your results. As you can see from this output, the `/usr` directory takes up the bulk of the space on my system, with `/usr/lib` using almost half the space inside `/usr`. Also note that `/var/cache/apt` is listed here—more on how to deal with that below.

Free as in Space

Now that you know how your storage is being used, here are a few common-sense ways to manage those files and free some space. If you do Linux programming, build software from source or regularly download tarballs, you probably have these tarballs lying around along with their extracted directories. One easy way to free up space is to delete either the tarball or the extracted directory. If you build your own kernels, you probably have a number of old kernel source trees in `/usr/src` that you won't ever use again and could stand to delete.

Another common space-waster is old ISO files. Do you really still need that Red Hat 7.2 ISO? If so, burn an archive copy or two to CD and then delete the image. Along those same lines, audio files

always end up with either an extra copy in a directory for a mix CD, or if you play with video conversion tools like `me`, you have video files in different phases of being transcoded. If you are done with a project, why not delete them and save the space?

On desktops, but especially on servers, one of the most common places you will find wasted space is in log directories. Logs definitely can be useful, but some logs and some levels of debugging are useful only immediately after a bug is found; the rest of the time they can be truncated or archived safely. Take a look in `/var/log/`, and see how many large uncompressed log files you have. If the file is no longer being used, you should gzip it. You would be amazed how far you can compress incredibly large log files if you haven't tried it before. If you aren't sure whether a log file is still being written to, use `lsuf` to check:

```
# lsuf | grep "/path/to/filename"
```

If you regularly find yourself cleaning up or gzipping the rotated log files in `/var/log` (they append `.0`, `.1` and so on as they are being rotated), then edit `/etc/logrotate.conf` and enable compression. Usually, this simply requires finding the commented line labeled `#compress` and uncommenting it.

Another great place to free up space is in your package manager's local package cache. For instance, in the case of Debian-based systems, the packages apt downloads are cached in `/var/cache/apt/archives`. You could go to that directory and remove the files manually, or you simply could become root and type:

```
# apt-get autoclean
```

to remove all the cached packages you no longer need. If you have a distribution that uses `yum`, the following two commands will clear out the cached headers and packages from your system:

```
# yum clean headers
# yum clean packages
```

Finally, archiving can be a good solution when cleaning your storage space. If you have a local file server or one machine with more storage than the rest, why not make sure all your large files exist only there and then access them over the network? Alternatively, burn large files you want to keep but don't immediately need to CD or DVD. Once you are done, you'll have plenty of newly freed space—hopefully, enough to last you until next spring. ■

Kyle Rankin is a Senior Systems Administrator in the San Francisco Bay Area and the author of a number of books, including *Knoppix Hacks* and *Ubuntu Hacks for O'Reilly Media*. He is currently the president of the North Bay Linux Users' Group.

8th Annual

2008 Web Services/SOA on Wall Street

Show and Conference

Web Services, Service Oriented Architecture, SaaS and Web 2.0

February 11, 2008, Mon Roosevelt Hotel, NYC

Madison Ave and 45th St, next to Grand Central Station

"An intimate show and conference – I could talk to the speakers and see the exhibits without the usual big-show hassle."

Get on board for High Performance SOA and Web Services, SaaS and Web 2.0 for Wall Street.

Case studies for Service Oriented Architecture by leading Wall Street players.

2008 Web Services/SOA on Wall Street expands to include Web 2.0 for the Wall Street enterprise and Software as a Service (SaaS) offerings for the financial markets. Underlying technologies including ESBs, Java, .Net, XML and AJAX are also covered.

Wall Street IT directors are under the gun to reduce Total Cost of Ownership while increasing Agility. Our event will offer a unique, person-to-person networking opportunity for Wall Street IT executives and project directors.

The 8th Annual will bring together major Wall Street speakers together with leading vendor experts. See who is changing the way the financial markets employ new innovations while maintaining 24/7 operations.

This is your opportunity to attend this Live Search Engine covering:

- Web Services and Integration Technologies
- High Performance Service Oriented Architectures
- SOA Management, Security and Performance
- Software as a Service (SaaS) Delivery of Financial Applications
- SaaS Infrastructure, Virtualization and On Demand Technologies
- Web 2.0 Collaboration suites - From RSS to Wikis
- Rich User Interface delivery - AJAX, Flex, JavaFX
- Web 2.0 and SOA Mashups - A New Middleware for Wall Street
- XML, Java and .Net infrastructure

The Show is free, but you must register in advance. Full conference only \$295. Save \$100. \$395. on site. Register today.



Show: Mon, Feb 11 8-4:00
Conference: 9-4:50

2007 Sponsors



Microsoft

webMethods



2008 Media Sponsor



Register online: www.webservicesonwallstreet.com

For more information on the event, to be a sponsor or an exhibitor, contact:

Flagg Management Inc
353 Lexington Ave, NY, NY 10016
(212) 286 0333 flaggmgt@msn.com

Free Software Foundation's GNU Affero General Public License

Although not a product per se, the Free Software Foundation's (FSF's) newly minted GNU Affero General Public License Version 3 (GNU AGPLv3) will affect many forthcoming works of software artisanship. Based on version 3 of the GNU General Public License (GNU GPLv3), the new Affero "fork" includes additional terms that allow users who interact with the licensed software over a network to receive the source for that program. With Affero, FSF seeks to foster user and development communities around network-oriented free software. FSF claims that the GNU AGPL will enable the same kind of massive collaboration among developers around Web services and other networked software that the GNU GPL has fostered over the years with non-networked applications.

www.fsf.org



eZ Systems' eZ Publish

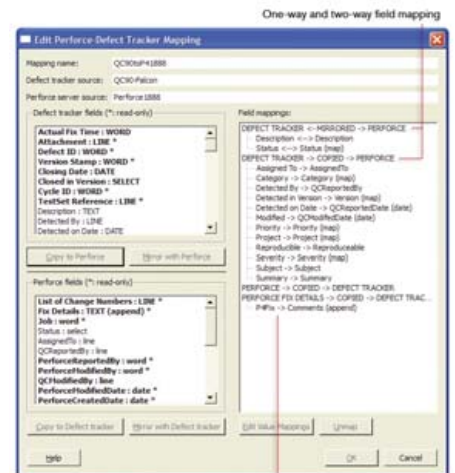
Further boosting Norway's place in global open-source development, eZ Systems recently released version 4.0 of eZ Publish, the company's enterprise content management system. eZ Publish is an application for creating Web sites, on-line stores, intranets and extranets. New features in 4.0 include full PHP 5 compatibility, full support for using eZ Components in plugins, improved internal XML handling and an updated Web site interface. The product is available as either an out-of-the-box or a tailor-made solution, depending on the varying needs of clients. GPL'd Linux and Windows versions are available for download at eZ Systems' Web site.

www.ez.no

Perforce's Fast Software Configuration Management System & SDK

Perforce wrapped up 2007 announcing a new version of its Fast Software Configuration Management (SCM) System, Perforce 2007.3. SCM is an application version lifecycle management (ALM) tool that versions and manages source code and digital assets for enterprises of all sizes. The most significant component of this release is the new SDK for the Defect Tracking Gateway, which allows customers and vendors to develop improved integrations to commercial and in-house tracking systems. Perforce also claims an advantage from its ability to integrate with other tools rather than being a one-stop shop, allowing customers to add the project management and process automation tools of their choice. A 45-day full version of Perforce with support and a free, two-user version are available from the firm's Web site.

www.perforce.com



Fidelity National Information Services' FIS Profile

Fidelity National Information Services announced new performance benchmarks on FIS Profile, its real-time technology solution for the commercial and retail banking industry, now that it runs on Linux. By running FIS Profile on Red Hat Linux Enterprise 5 and the HP ProLiant DL580 G5 server platform with four Intel Quad-Core Xeon Series 7300 processors, the solution can manage a bank with 25 million accounts, running core banking processes in real time on a single server. Fidelity claims that the solution offers a tenfold improvement in cost performance per account while maintaining the reliability and security required by the commercial-banking industry. This solution is intended to replace the mainframe-based systems for mid-tier banks that were developed in the 1980s. Both Red Hat and Intel were involved in developing the integrated platform.

www.fidelityinfoservices.com

Please send information about releases of Linux-related products to James Gray at newproducts@linuxjournal.com or New Products c/o *Linux Journal*, 1752 NW Market Street, #200, Seattle, WA 98107. Submissions are edited for length and content.

AMCC's 3ware 9690SA Serial Attached SCSI RAID Controller

New in the SAS space is AMCC's 3ware 9690SA Serial Attached SCSI (SAS) RAID Controller whose sales proposition includes the flexibility offered by its three PCI Express low-profile controller choices: eight internal ports, eight external ports or four internal/four external ports. The 3ware 9690SA provides 2–24 ports of SATA connectivity and maximized SAS expandability for up to 128 devices per controller. The SAS controllers include AMCC's unified RAID management interface and software suite, enabling a simplified configuration experience irrespective of its storage interface. The product is destined for data-center environments needing expanded connectivity and high levels of read and write performance. Targeted applications include databases, NAS storage, Web servers, cluster servers, supercomputing, near-line backup and archival, security systems and pro audio and video editing appliances.

www.amcc.com



Cray Inc.'s XT5 Supercomputer Family

Cray Inc., progeny of the storied Cray Research, recently released its XT5 family of Linux-based supercomputers. Cray says the XT5's massively parallel processor (MPP) system includes a new eight-socket compute blade that quadruples local memory capacity, doubles processor density and improves energy efficiency. Other features include single-fan vertical cooling, compute blades designed for optimal airflow and CPU configurations up to 192 processor sockets or 768 CPU cores. To improve scalability, the Cray XT5 family also includes the industry's first integrated hybrid supercomputer, the Cray XT5^h system. The XT5^h integrates multiple processor architectures—including vector processors, GPUs, accelerators and FPGAs—with a complete software development environment into a single system supporting diverse work flows.

www.cray.com



Appro's Xtreme-X Supercomputer Series

At Supercomputing 2007, Appro unveiled its forthcoming Xtreme-X Supercomputer Series, a product line based on scalable clusters that provide cost-effectiveness, energy efficiency and scalability. The series is designed to scale out data centers for medium- to large-scale enterprises and HPC deployments. The first model in the series, the Appro Xtreme-X1, will ship in the first half of 2008 and is based on dual-socket, Quad-Core Intel Xeon processors. Besides its 128 nodes/512 processors and 6TF of computing power in a single 42U equipment rack, the product has Appro's new Directed Airflow cooling configuration, which the company says will reduce data-center floor space by 30% while maximizing power and cooling efficiency. In addition, the Xtreme-X1 features redundant (Dual Rail) InfiniBand connections with low-latency Mellanox ConnectX host channel adapters and Ethernet management fabric and network switches, with all critical components being easily accessible, hot-swappable and redundant.

www.appro.com



Hacking: The Art of Exploitation by Jon Erickson (No Starch Press)

No Starch Press continues its tradition of naughty geek entertainment with the 2nd edition of *Hacking: The Art of Exploitation* by Jon Erickson. Although other books in this genre show not only how to run other people's exploits but also how to perform and write them on your own, Erickson uses examples to illustrate the most common computer security issues in three related fields: programming, networking and cryptography. Some examples include stack-based overflows, heap-based overflows, string exploits, return-into-libc, shellcode and cryptographic attacks on 802.11x. A live Linux CD also is included.

www.nostarch.com



HARDWARE

Zonbu

Not only does the mini Zonbu PC run Linux, maintain itself and store your files on-line, it's also one of the greenest machines out there. JAMES GRAY

Although you may not yet find preinstalled Linux too easily at your neighborhood computer superstore, our beloved OS is bubbling up in more scintillating ways, including in the recently released Zonbu PC. Zonbu is a mini, fanless Gentoo Linux-driven PC that, with its on-line storage (sans hard drive!), functions with a "computing as a service" ethos. In addition, Zonbu is one of the few PCs that markets its green "street cred", aiming to provide us concerned citizens with a means to reduce our energy consumption and thus our impact on the environment.

The Zonbu Concept—Does It Add Up?

Zonbu plugs itself as a "compact, totally silent, ultra-low-power mini with all of the bells and whistles". Although true to a degree, the Zonbu arrives on your doorstep in a compact box with simply the machine and a power cord. You'll have to purchase, or more likely scrounge for, the requisite monitor, keyboard and mouse. Zonbu also lacks the monstrous hard drive to which we've accustomed ourselves in PCs these days. Instead, the machine includes a 4GB CompactFlash card containing the Gentoo Linux OS and a local cache (around 3GB) for a

limited number of files. The bulk of your files likely will reside on your on-line storage space.

Here is where the "computing as a service" ethos comes in. Yes, you could purchase a Zonbu outright for \$249 and go your own way, simply taking advantage of Zonbu's free upgrade service and storing your files on a Flash drive. What Zonbu would rather you do is pay for the convenience of its subscription service, which includes secure on-line storage space of varying sizes, secure backup, 30 days live support and e-mail support thereafter, and transparent upgrades of the OS and installed applications. You can pay full price and subscribe on a month-to-month basis, or if you prepay for a subscription, Zonbu will kick in a discount of \$150 for two years or \$50 for one year.

At the time of this writing, you can subscribe to a plan with 25GB of storage for \$12.95 per month, 50GB for \$14.95 or 100GB for \$19.95.

The on-line storage system is nice, as Zonbu transparently manages the interplay between the 3GB local cache and the larger storage space. You also can go on-line anytime from any computer with a browser and access your files.

So, let's take a look at what we've



Figure 2. Zonbu Side View

got here with Zonbu. Consider that you're getting a machine loaded with a 1.2GHz low-power Via C7 processor and 512MB of RAM, but no hard drive, mouse, monitor or keyboard. Let's also say you purchase the 50GB of storage/service plan for two years, as well as the optional Wi-Fi dongle and CD-RW/DVD drive, all of which will set you back around \$500. Then, consider that you will have to renew your subscriptions after those two years, or else drag all of your files down onto your own storage device.

From another angle, consider Zonbu's conveniences, such as its diminutive and quiet presence and letting someone else worry about backup, security and updating tasks. In addition, if you're bouncing around the globe, think of how nice it would be to log on from any computer in the world and access all of your files. As always, the trade-off is time or money. Personally, I would pay for this convenience when getting less geeky friends or family running Linux and minimizing their support requests



Figure 1. Zonbu Front View

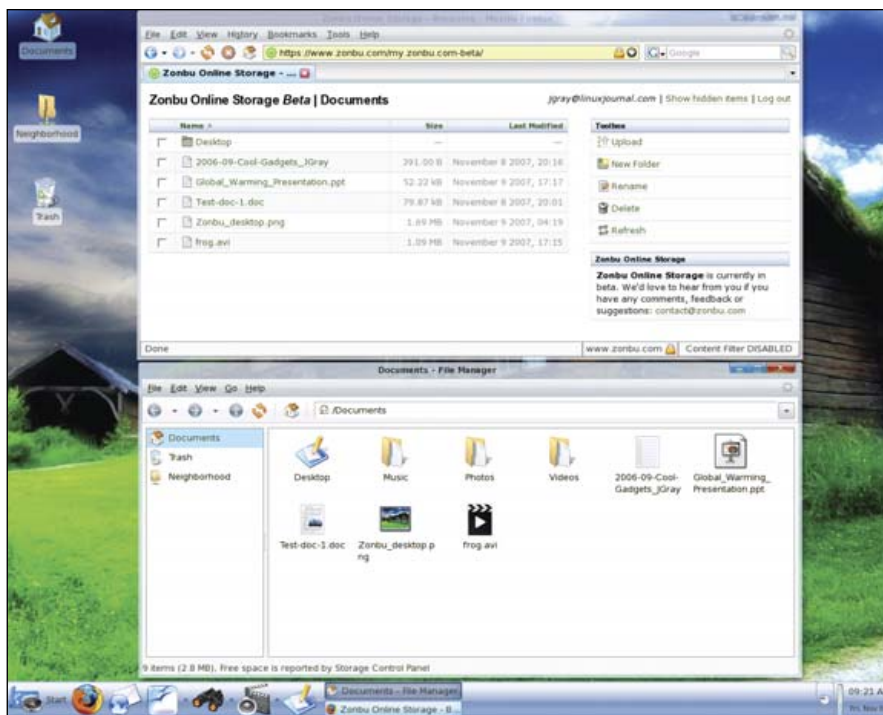


Figure 3. Zonbu keeps recent files in its 3GB local cache. Other files reside in on-line storage if you subscribe to a storage and maintenance plan.

to me.

Don't forget that Zonbu is a power-miser, consuming roughly 10–15 Watts, depending on the load, which compares well with laptops. Most PCs of similar robustness (without monitor) will gulp 60–100 Watts or more, depending on numerous factors. Zonbu's marketing folks say you'll save more than 1,200 kilowatt hours over the course of a year, which seems generous given their assumption that a typical PC averages 175 Watts, but let's be conservative and assume a savings of half that amount—that is, 600 kilowatt hours. I currently pay \$0.07 per kilowatt hour, which would save me \$42 over the course of a year.

Versions of Zonbu

If you set out to buy a Zonbu, you'll find the Standard Edition preloaded in a configuration as elaborated above, either with or without subscription plan. However, dig deeper and you'll also find three other Zonbus—Developer, Free Edition and Kiosk editions.

Because the standard Zonbu PC has a fixed OS configuration and application set, this version is ideal for Linux evangelism—that is, introducing Linux to

people or organizations with limited computing requirements.

However, we insatiable tinkerers might pop an artery with the fixed configuration and lack of control found in the Zonbu Standard Edition. Therefore, the *LJ* crowd will likely find more joy in the alternative Zonbu editions.

Let's have a look at each of these editions.

Zonbu and Newbies

If only I had a nickel for each time I've heard one of you tell me "All my [insert non-geek relatives/friends] need is Web access, an office suite and MP3s. There's no reason they shouldn't be using Linux." For situations like these, consider Zonbu your "Linux Conversion Appliance".

Though your grandmother likely won't care, Zonbu Standard runs Gentoo Linux and uses Xfce as its desktop environment. Our review machine ran Version 6.999 of its software, with Linux kernel 2.6.22.4, which Zonbu still considers betaware. Meanwhile, the hardware is not in beta.

For better or worse, neither you nor your grandmother can change a Zonbu build in the Standard Edition, not even install additional applications. On the

Specifications

Company: Zonbu, www.zonbu.com

Bandwidth requirement: 64KB/s minimum, 256KB/s or faster recommended.

Physical dimensions (height x width x depth): 6.75" x 4.75" x 2.25" (17.1cm x 12.1cm x 5.7cm).

Processor: 1.2GHz Via C7.

Graphics card: Integrated Via C7.

RAM: 512MB.

Flash memory: RiData 150x 4GB CompactFlash card included.

Hard drive: none (storage is on-line).

Memory: no HD, only a 4GB CompactFlash card, which contains the OS and applications.

Optical drive: none included; optional CD-RW/DVD drive available for \$49.

Ports: six USB, VGA out, speaker, mic.

Networking: built-in 10/100 Mbps Ethernet, 802.11 b/g Wi-Fi available via optional USB adapter for \$29.95.

Subscription options: \$249 per month + optional month-to-month subscription fee (\$12.95–\$19.95) depending on on-line storage space. \$150 discount on machine with prepaid two-year subscription, and \$50 discount with a one-year plan.

flip side, Zonbu contains a wide range of standard Linux apps and presents them to the user in a very functional, logical manner. See the Included Applications sidebar for a sample of Zonbu's applications.

Support for printers, Flash drives and digital cameras is as good as any Linux machine, and other music players and iPods (though not fully) are supported as well. Buyers beware that neither Bluetooth, scanners nor Webcams are currently supported, and the Belkin F5D7050 is the only supported Wi-Fi device.

Although media codecs are always an issue with Linux, Zonbu has much of that pretty well solved. One can play

Zonbu's Included Applications

- Firefox
- Evolution
- Pidgin (supports IRC, AIM, ICQ, MSN and Yahoo! networks)
- Skype
- Azureus
- aMule
- OpenOffice.org
- Acrobat Reader
- GnuCash
- Banshee
- MPlayer
- F-Spot
- GimpShop
- Scribus
- Nvu
- Numerous games, including *FreeCiv*, *Supertux*, *Frozen Bubble*, *Penguins* and others.

back the following: MP3s, WMA, WMV, AVI, QuickTime, MPEG/MP4, RealMedia and DVDs from around the globe (given you've got the optional CD/DVD drive).

Another issue to consider is Internet connection. Luckily, you're not completely up a creek if your Internet access is down, because most recent files will be stored in the onboard cache. You simply won't have access to your older files in online storage. In addition, Zonbu needs only 64KB/s to work, but a minimum of 256KB/s is recommended. I tested Zonbu with a slower connection, around 100KB/s, and found it to work fine under everyday working conditions with small files. However, logic tells you that pumping gigantic files

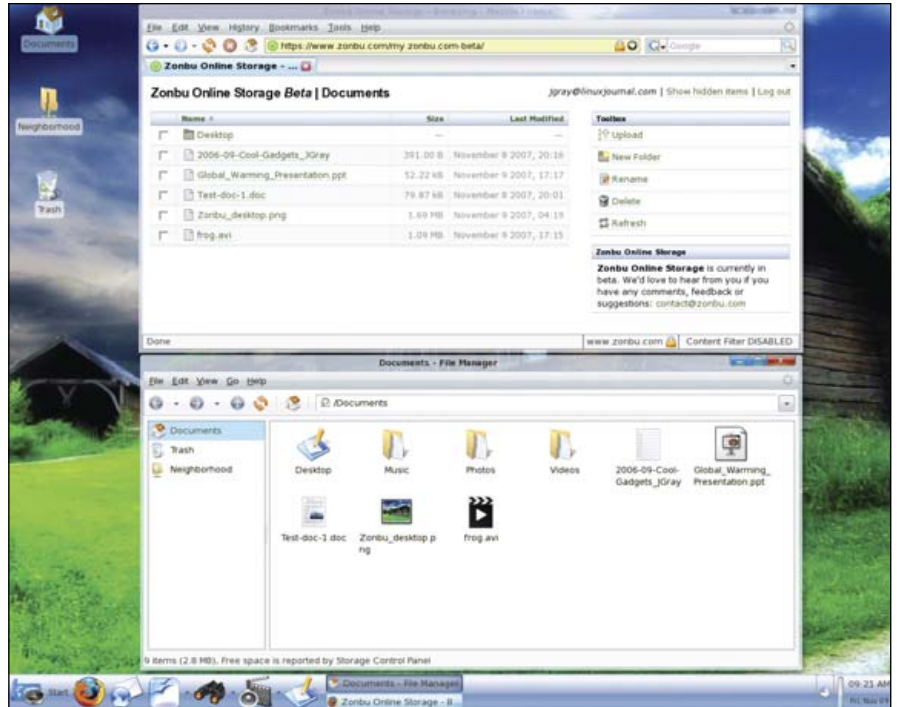


Figure 4. Zonbu runs Gentoo Linux and the Xfce desktop environment. In the Standard Edition, Zonbu maintains the OS and all applications. If you want to change anything, set up the Developer Edition instead.

though small pipes is no fun, so keep this in mind if you'll be transferring large files frequently.

Those computing in the wacky world of Windows who wish to take their existing files along into the Zonbu universe can utilize the Windows Importer Tool. This tool, which runs on Windows, allows you to select the files to transfer, including e-mail, and will upload them to the storage space. Zonbu will synchronize the e-mail files from Outlook, Outlook Express, Netscape Composer 4.0 and Eudora to work with Evolution. We were able to get a bunch of Outlook-based e-mail synchronized without a hitch.

Beyond the annoyance of the inability to change either the OS or your applications, a few other minor issues arose. In addition to making annoyingly loud beeps while starting up, Zonbu's boot time is a bit slow even for Linux—close to two minutes (20–30 seconds longer than our SUSE Linux and Ubuntu systems).

Another annoyance in my book is Zonbu's avoidance of Linux to the general public, saying that "The Zonbu OS looks and works like the

latest PC operating systems" and offering advantages like superior security. Clearly when reading the Web pages for developers, Zonbu is zealous about Linux, and yes, we do want to present Linux's modern, user-friendly face to new recruits. Nevertheless, why not use this as a teaching opportunity to plug Linux to the world not just by functional advantage but by name too?

Zonbu and Geeks

As mentioned previously, Zonbu has three additional editions that are interesting to users seeking more control and configurability: the Free, Developer and Kiosk Editions. Though not promoted heavily, you can find a wealth of information about them in Zonbu's Developer Corner on its Web site.

If you want to forego Zonbu's subscription and storage offerings, simply buy the \$249 machine and follow the instructions to set yourself free. With the Free Edition, you still can take advantage of free system and application upgrades, but as with the Standard Edition, you cannot change anything. Later, you can reactivate the Standard Edition and choose a

subscription plan if so desired.

For full control and root access, go with the Zonbu Developer Edition, which can be activated quite easily. In this edition, you can install additional system files and applications while still taking advantage of the subscription service. Luckily, Zonbu still offers support for Gentoo if you go this route, but other distributions are not supported. Regardless, Zonbu does provide tips and pitfalls about using other distros. In addition, you can find information on installing the Zonbu OS on a VMware virtual machine, as well as putting the Zonbu OS onto a CompactFlash card.

Finally, Zonbu gives you the option of activating the Kiosk Edition, which functions the same as the Standard Edition. Unfortunately, unless you're using the Developer Edition, very little customization is possible besides determination of the home page.

The Environmental Bonus

Although a number of computer companies are greening their operations and products, Zonbu appears to be one of the first to use its "environmental cred" as a core selling point. Furthermore, Zonbu is trying to cover all of the bases, which is summed up in its EPEAT Gold rating for strong overall environmental performance. Only 12 desktop machines have reached this mark to date. Specifically, Zonbu delivers, as illustrated above, significant gains in energy efficiency, achieving the US EPA Energy Star 4.0 rating. Second, Zonbu purchases carbon offsets from the firm Carbon Trust, which invests in projects that reduce net carbon emissions society-wide, such as wind energy or methane capture in landfills. Third, Zonbu builds its hardware with recycling in mind and follows the European RoHS Directive, such that no more than 25% of the hazardous substances (such as lead and mercury) that go into typical desktops are used. Fourth, when you're ready to upgrade, Zonbu will take back your old device and foot the bill for its recycling.

The Final Word on Zonbu

To answer our question from above, the Zonbu mini-PC indeed adds up

if you're leaning toward convenience over penny-pinching on your next PC purchase. You'll save a great deal of time on backups, updates and other maintenance, and you'll get excellent functionality out of the box. Furthermore, if your situation calls for a plug-and-play Linux solution with basic functionality and without a lot of esoteric Windows-only applications, Zonbu is an excellent choice. Also, remember that you can activate the Developer Edition and add applications and functionality to your heart's delight.

If greenness is part of your calculus, then Zonbu is nearly peerless. Its credentials regarding low power consumption, recycling and carbon footprint are all industry-leading.

So, should you commit to a two-year subscription with Zonbu? Two years is an eternity in our business. Given the plethora of positive press Zonbu has received, it seems that the

firm should expect success. However, fame is fleeting and users fickle. My gut says start with a one-year plan and see how things go. ■

James Gray is *Linux Journal* Products Editor and a graduate student in environmental science and management at Michigan State University. A Linux enthusiast since the mid-1990s, he currently resides in Lansing, Michigan, with his wife and cats.

Resources

Zonbu: www.zonbu.com

Electronic Product Environmental Assessment Tool (EPEAT):
www.epeat.net

US EPA Energy Star Program:
www.energystar.gov

The Climate Trust:
www.climatetrust.org

Fasten your Seatbelts!

Feel the speed of True Quad Core Power

Quad Core	Quad Core	2x Quad Core
 Kentsfield Xeon 3200 \$100	 Clovertown Xeon 5300 \$140	 Clovertown Xeon 5300 \$180

1 GB RAM
500/250 GB SATA 2
1300 GB/mo Included
100 Mbps Dedicated Port



carinet CARI.NET/LJ
888.221.5902

VirtualBox

BITS AND BYTES MASQUERADING AS MACHINES



Using virtualization to turn your ho-hum desktop into a computer cluster.

JON WATSON

As our home computers become more robust, we can do more powerful things with them. Virtualization isn't new; it's almost as old as computers themselves, but the ability to run virtualization platforms on a typical home computer is relatively new and becoming more exciting every day.

For the uninitiated, virtualization in the context of this article refers to the ability to run a full-blown operating system within an application running on an existing computer. For example, virtualization allows us to run Windows XP

in a window on a Linux desktop, a full-blown LAMP server on a Windows machine, or BSD inside a Mac OS machine. The combinations are endless, and any relatively modern personal computer is likely to be beefy enough to handle it.

So why would you bother? Well, only those who haven't been exposed to virtualization generally ask this question. Once the benefits and neat factors of desktop virtualization become apparent, it's hard to stop the ideas from flowing. In my opinion, the two major benefits of virtualization are ease

of backups and host system stability. If I want to figure out how to set up a LAMP server, I have two choices: get my hands on another physical machine or do it in a virtual machine. Either way will leave me with a fully functional LAMP server, but the virtual machine costs me nothing and is set up on my existing computer without making any system changes to it.

Once my LAMP server is happily humming away, it would behoove me to have some off-site backups of it. With a physical server, I would need to have yet another machine of some kind

perform off-site backups, and bare-metal restores can be tricky if the entire machine melts down. With a virtualized LAMP server, all I have to do is copy the files that the virtual machine is composed of, burn them to CD/DVD and chuck it in a drawer—cheap and easy.

Virtualizing requires two distinct components: a host machine and a guest. The host machine is your desktop or laptop where the virtualization software is installed. The more common virtualization applications on the market these days are VMware, Win4Lin (Windows virtualization only), VirtualBox and Parallels, with more offerings appearing every day. Although many of these products also provide enterprise-level server virtualization, in this article, I focus on the home enthusiast with a typical Linux computer to get things rolling.

Having Fun with VirtualBox

It makes the most sense to use innotek's VirtualBox, because unlike the other virtualization offerings, VirtualBox has a GPL'd Open Source Edition (OSE). The closed-source edition has dual licensing, in that it is gratis for personal and evaluation use but fee-bearing for enterprise use. The OSE is licensed under the GNU GPLv2 but is missing some well-thought-out functionality that is available in the non-GPL'd editions. The missing functionality from the OSE includes:

1. No RDP support—you cannot connect to VirtualBox virtual machines from a remote location.
2. No USB support—USB devices won't work in the virtual machine.
3. No USB-over-RDP support (I guess that makes sense given the first two limitations).
4. No shared folders—you will not be able to share data between the host and guest machines.
5. No iSCSI initiator—SCSI disks cannot be used as virtual disks.

Installation of VirtualBox

VirtualBox downloads are available for Windows, OS X and a wide variety of Linux distributions. I am running Ubuntu Gutsy Gibbon on my Dell Inspiron 9400 (1.83GHz Core Duo with 1GB of RAM), so I installed the 13.6MB `virtualbox_1.5.2-25433_Ubuntu_gutsy_i386.deb`. Note that the OSE is available only in a tarball, but because I am installing VirtualBox for personal and evaluation use, I am lazily installing the Debian binary (Figure 1).

The Gutsy Gibbon VirtualBox has some dependencies, which Synaptic took care of for me. If you run into dependency problems, ensure that you have `libxalan110` and `libxerces27` installed. Also note that at the end of the VirtualBox install, a dialog box containing instructions on how to set the permissions of the `/dev/vboxdrv` file

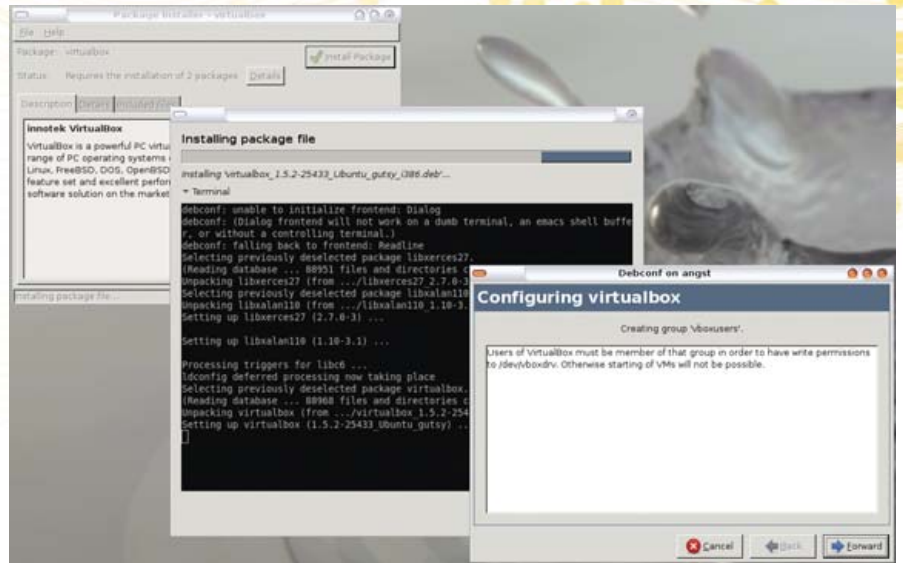


Figure 1. Installing the VirtualBox Gutsy Gibbon .deb Package

is displayed. Pay attention to those instructions as you will need them later.

After installation, I found VirtualBox under my Ubuntu Menu→System Tools slot.

Creating a New Virtual Machine

Because I opted for the personal evaluation binary, I had to accept the PELA and fill in a quick registration at first launch.

There are two parts to creating any virtual machine: the creation of the virtual machine container and then the installation of the OS into that container. To do the first part, start VirtualBox and click the New icon. This opens up the New Virtual Machine Wizard (Figure 2).

First, you must name your virtual machine. I know from

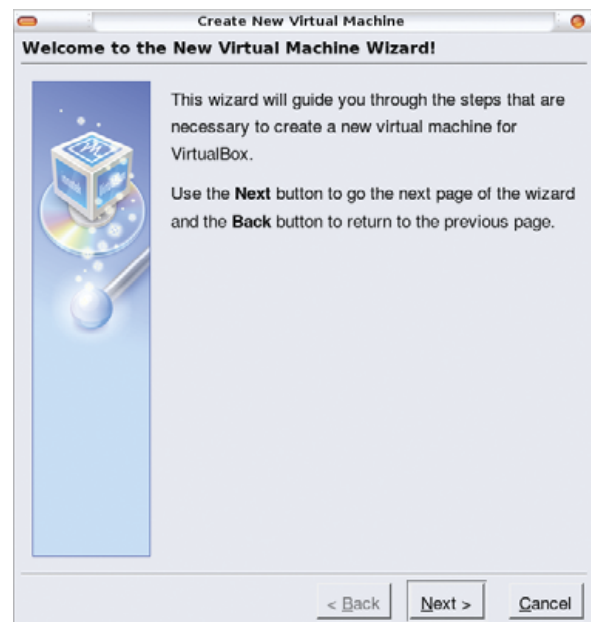


Figure 2. The VirtualBox New Virtual Machine Wizard

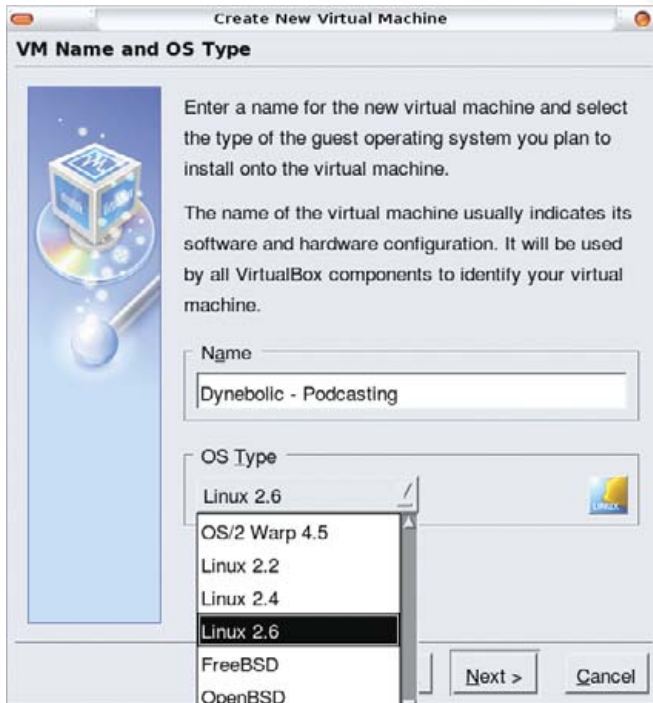


Figure 3. Select the virtual machine OS and set the name.

experience that I will create many, many virtual machines, so I have opted to name them each as the OS and then its function. In this case, I want to see whether I can run our podcasting rig out of a VirtualBox virtual machine, so I'm going to install dyne:bolic into my virtual machine. Therefore, I name this virtual machine Dynebolic-Podcasting and select the Linux 2.6 kernel OS (Figure 3).

Like a physical machine, a virtual machine needs some RAM. On the next screen, the wizard suggests 256MB of RAM. However, because I know that crunching a big Ogg file can take a system to its knees, I allocate 512MB to this virtual machine (Figure 4).

What does the virtual machine need next? Well, just like a regular machine, it needs a hard drive. I know from experience that I can use a CD/DVD ISO file (like my friendly dyne:bolic ISO) as the bootable hard drive, but VirtualBox's new machine wizard doesn't allow for that. I will be able to add my ISO later on, but not right now. So, as counter-intuitive as it seems, simply click the Next button without setting up any disks and acknowledge the error dialog that tries to stop this folly (Figure 5). Click Finish, and *voilà*, you have a shiny new dyne:bolic virtual container.

Sadly, this virtual machine will not boot, because I still don't have a hard drive or other bootable media. I want this thing to boot the dyne:bolic ISO image I have stored on my hard drive, so I have to mount a virtual CD/DVD drive into my virtual machine. This is one area where virtual and physical machines differ. If this were a physical machine, I would not only have to attach a CD/DVD drive to the machine, but I also would have to insert my dyne:bolic CD into the drive for the machine to boot in to it. With a virtual machine, I merely point to the ISO image, and VirtualBox is smart enough to understand that it is

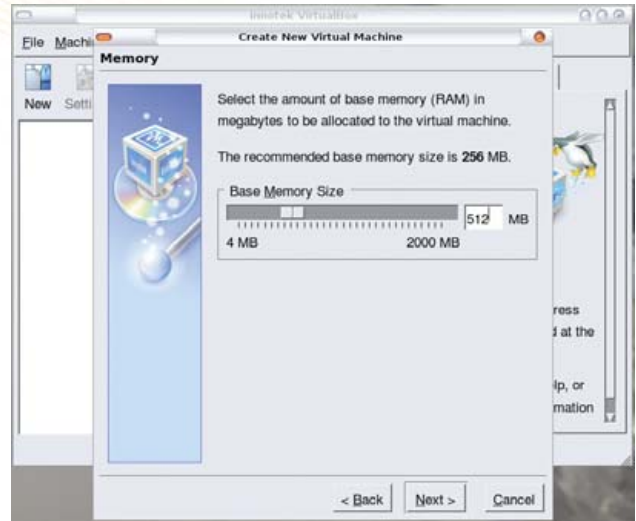


Figure 4. Setting the RAM of the Virtual Machine



Figure 5. VirtualBox's Caution Dialog about Not Creating a Hard Disk

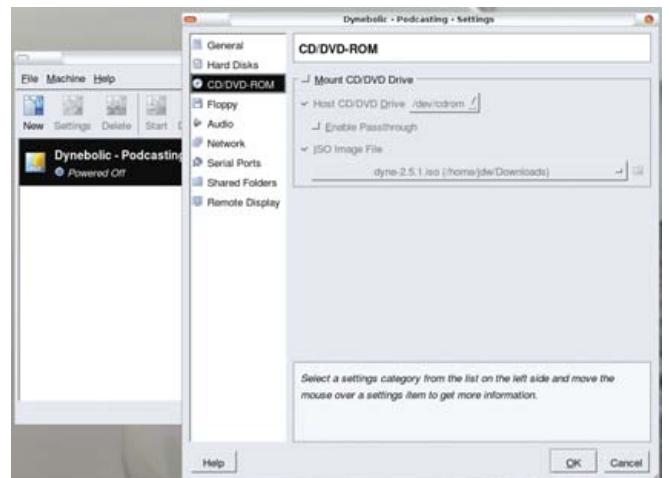


Figure 6. Create a Virtual CD/DVD ROM Drive

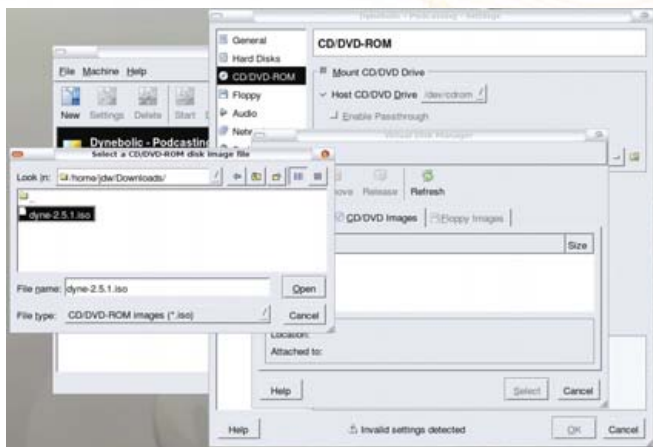


Figure 7. Use an ISO file from which to boot the virtual machine.

to boot from it. To accomplish this rather unintuitive task, follow these steps from the main VirtualBox screen:

1. Highlight the virtual machine you just created.
2. Click the blue CD/DVD-ROM label in the right column.
3. Click the Mount CD/DVD Drive check box (Figure 6).
4. Click the ISO Image File check box.
5. Click the file folder browse icon next to the ISO Image File drop-down list box.
6. Click the Add button and browse for your ISO image (Figure 7).

Creating a Hard Drive

Note that the screen from Step 2 above also provides the interface to add all sorts of ports and storage devices to the virtual machine. I'm happy running my dyne:bolic as a live CD, but if I wanted to install it into this virtual machine, I would have to create a hard drive into which to install it.

There are two types of hard drives in the virtual world: those that take up all the space they are allocated right off the bat, and those that can grow into their maximum allocated size as needed. I'm a big fan of the second type, because it allows me to allocate large amounts of disk space to my virtual machines without fearing that I will lose all that space from my host machine right away.

To create a growable hard disk, get yourself back to the main VirtualBox interface and ensure that the virtual machine is shut down. Then, follow these steps:

1. Highlight the virtual machine to which you want to add a drive.
2. Click the blue Hard Disks label in the right column.
3. To add a primary master drive, select the Primary Master check box.



Figure 8. Add a hard disk to the virtual machine.

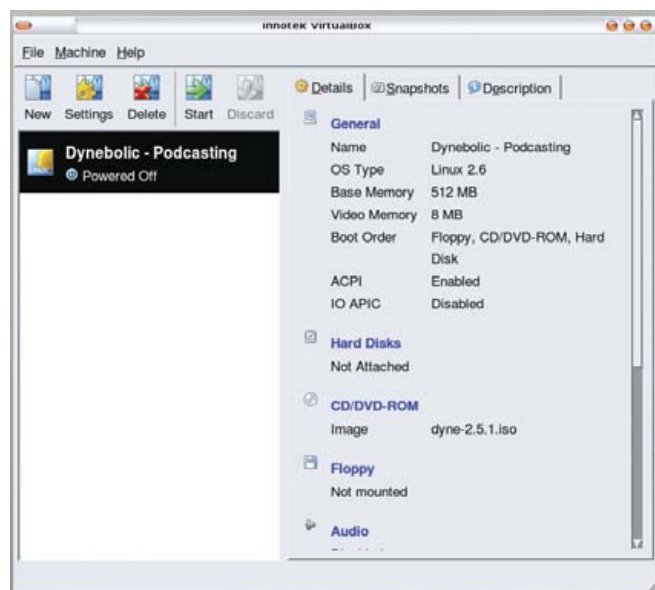


Figure 9. Newly Created Virtual Machine Overview

4. Select the file folder browse icon next to the Primary Master check box.
5. Click the New button.
6. Select the Dynamically expanding image check box, unless you have a reason to select the Fixed-size image, and click the Next button.
7. I recommend naming your disk file(s) with the same name as the virtual machine to which they belong. There are instances where you might want to share disks between

FEATURE VirtualBox: Bits and Bytes Masquerading as Machines

virtual machines, and therefore this naming convention might not make sense, but for the most part, the “one-disk-one-machine” philosophy works well.

8. Use the Image Size slider, or type in the desired disk size (Figure 8).
9. Click the Next button and the Finish button, and you now have a disk for your virtual machine.

If you’ve done everything right, the main VirtualBox screen will show your ISO under the CD/DVD-ROM label (remember, I didn’t actually create a hard disk for my machine, so it doesn’t show in my screenshot—Figure 9).

Turn Me On!

Let’s turn this baby on and see what happens. To turn on your new virtual machine, highlight it in the left pane of the VirtualBox interface, and then click the Start button in the toolbar. Note the Auto capture keyboard notification dialog. Essentially, it is notifying you that once you click anywhere inside your running virtual machine, all your keyboard keystrokes from that point on will be sent to the virtual machine rather than to the host machine. This may not seem like a big deal, but trying to shut down your machine or even resize a window when you have no keyboard or mouse control of the host machine is really quite difficult. In order to free your mouse and keyboard from the confines of your virtual machine, you have to press the right-**Ctrl** key. You can specify a different Auto capture escape key by going to **File**→**Preferences**→**General**→**Input**, if the default right-**Ctrl** key doesn’t suit you.

Your mileage may vary, but the next thing I encountered was another dialog box full of doom and gloom. This one informed me that my logged-in user (jdw) wasn’t part of the `vboxusers` group and therefore could not access the `/dev/vboxdrv` file. It would seem intuitive to me to add the user that is installing VirtualBox to the `vboxusers` group to avoid this error message, but that isn’t part of the installation process. To do it manually on my Ubuntu box, I simply edit my `/etc/group` file to add jdw to the `vboxusers` line and relog in (Figure 10).

My new group membership is all I need to launch my virtual



```
GNU nano 2.0.6 File: /etc/group
bash:~$ nano /etc/group
bash:~$
bash:~$ cat /etc/group
bash:~$
bash:~$
```

Figure 10. Adding My User to the `vboxusers` Group



Figure 11. Dyne:bolic Running in a VirtualBox Machine on a Gutsy Gibbon Desktop

machine successfully, and I am now running dyne:bolic in VirtualBox on my Ubuntu Gutsy machine (Figure 11). I have good reasons for wanting to run dyne:bolic in a virtual machine—primarily because dyne:bolic doesn’t like my actual video card, an ATI x1400, but it has no problems with the innotek virtual VGA driver. I also get the added benefit of not having to boot my system in dyne:bolic, so I retain access to my host machine while I’m using it.

Guest Additions

VirtualBox provides some applications called Guest Additions. If you’ve ever used VMware, you’ve likely stumbled across VMware Tools. Guest Additions are much the same thing in that they are installed into the guest machine and provide some functionality, such as arbitrary window size and better mouse control, to the guest session. You certainly can run a VirtualBox virtual machine without Guest Additions installed, but the experience is better with them.

Guest Additions are installed into the virtual machine itself rather than onto the host machine. Therefore, for every virtual machine you have, you’ll need to install the Guest Additions through the machine’s handy **Install Guest Additions** menu option in the **Devices** menu.

Snapshots

Earlier, I mentioned that one of the great benefits of using virtual machines over physical machines is that a complete bare-metal backup is as easy as file copy. There are other benefits in this area as well though, such as snapshotting.

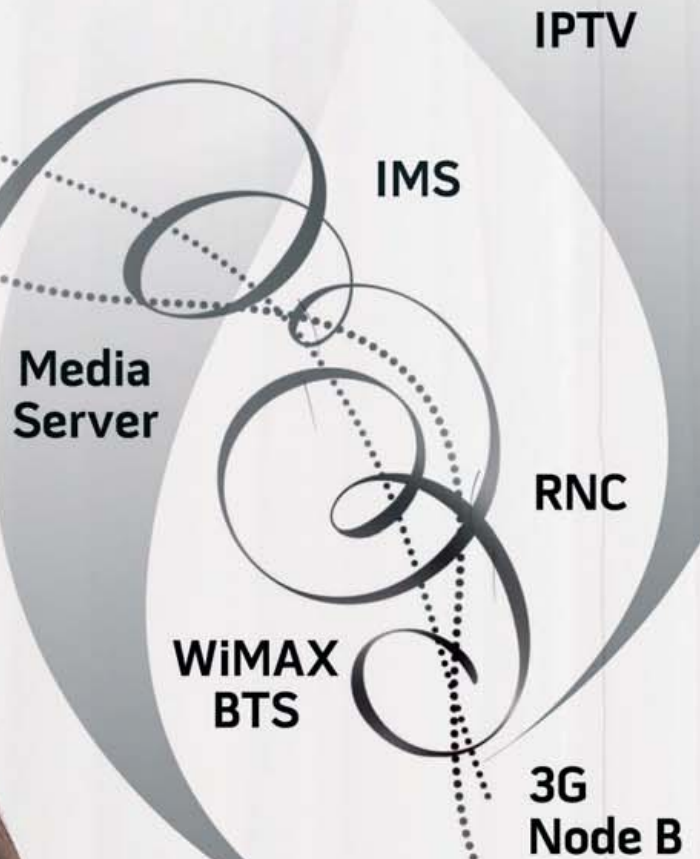
When a virtual machine is running, select **Machine**→**Snapshot** to take a quick restore point for the machine. This is invaluable when you’re installing new software, reconfiguring the virtual machine or doing something else dangerous. If you muck it up, you simply can shut down the virtual machine, select the **Snapshots** tab and restore the machine from any of the snapshots you’ve taken. To restore the machine to a previous snapshot, right-click the desired snapshot and select **Revert to Current Snapshot** (Figure 12).

Virtualization for a connected world

Now You Can Apply the Benefits of Virtualization To Connected Devices....

Virtualization makes software run more efficiently on hardware, reducing product and operational costs. VirtualLogix's VLX enables multiple operating systems to run simultaneously on the same single or multi-core processor, while maintaining hard real-time and high throughput requirements. The ability to share hardware resources and combine real-time operating systems with a Rich OS allows system designers to speed the integration of Linux and add more functionality quickly, safely, with reduced costs.

Find out how at VirtualLogix.com



DOWNLOAD A FREE WHITEPAPER!
www.virtuallogix.com/wp/linux_journal

Do you take

"the computer doesn't do that"

as a personal challenge?

So do we.

LINUX
JOURNAL™

Since 1994: The Original Monthly Magazine of the Linux Community

Subscribe today at www.linuxjournal.com

FEATURE VirtualBox

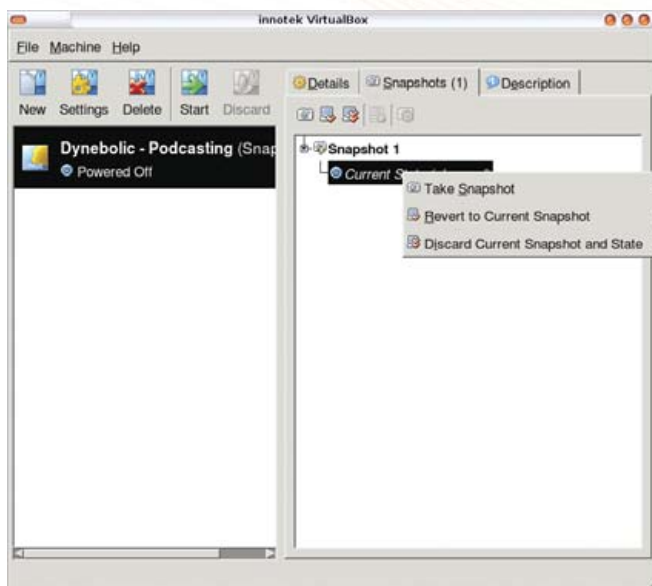


Figure 12. Select a snapshot to restore.

Getting Help

VirtualBox comes with a surprisingly complete help manual. Pressing F1 on any screen brings up the VirtualBox manual, although it isn't context-sensitive. The help manual is well laid out and easy to understand. There also are forums, IRC and a mailing list on the VirtualBox site as well as a public bug tracker.

Additionally, there's a complete set of technical documents available on the VirtualBox Web site that is aimed at developers and those who want to contribute to the OSE version of the product.

Your New Virtual World

Although the most common use of virtualization technologies is still certainly in the enterprise space, there are enough desktop virtualization applications out there now that a home user can join the fun. Whether you need an OS to study for a certification, want to run an OS that your host machine can't run directly (like me), want to run a server but don't have another physical machine to use, or just plain-old want to experiment, VirtualBox is a quick-and-easy way to jump into the fray. ■

Jon Watson is a CompTIA Linux+ certified Linux integration consultant. Jon lives in a 143-year-old house in the beautiful Canadian maritime province of Nova Scotia with his wife Kelly and their two dogs.

Resources

Open Source VirtualBox and Other Editions:
www.virtualbox.org/wiki/Editions

Downloading VirtualBox:
www.virtualbox.org/wiki/Downloads

VirtualBox Community:
www.virtualbox.org/wiki/Community



VERIO DEDICATED SERVERS: LINUX WITH A LINEAGE.

- **SAS70 Certification** for critical data security and scalability
- **Tier-1 Global IP network** and premier data centers for superior performance
- **Dedicated, certified support team 24/7/365** — no queue, no triage, just answers

Best of Breed.

When it's time to make the move to a dedicated server, experience is crucial. At Verio, we have a long-standing commitment to open source, dating back to our early work with FreeBSD. Verio — part of NTT Communications, the world's largest telecommunications company — is a strong, stable industry leader with more than a decade of trusted service and 500,000 business customers.

Call 1-800-269-3300 or visit dedicatedserver.com to receive a customized Technology Requirements Evaluation. And for a limited time, you can secure your data with a free backup solution.

Build on us.
VERIO
An NTT Communications Company

Verio and the Verio logo are trademarks and/or service marks of Verio Inc. in the United States and other countries. Linux® is the registered trademark of Linus Torvalds in the U.S. and other countries. The mark FreeBSD is a registered trademark of The FreeBSD Foundation and is used by Verio Inc. with the permission of The FreeBSD Foundation. All other names are trademarks or registered marks of their respective owners. ©2008 Verio Inc. All rights reserved.

Virtualize a Server with Minimal Downtime

Kyle Rankin



When it's time to convert a physical machine to a virtual one, use these steps to make the move safely and with a small maintenance window.

You don't have to look far in tech news to see that virtualization has become a big deal. After all, computers continue to become faster and more powerful, and as they do, the services that run on them often use fewer overall resources. On top of that, modern servers often need a fraction of their predecessors' power and cooling. With virtualization, you can get power savings and more efficient use of server resources, and you can create servers quickly without waiting on parts to arrive.

Although some people start from a clean slate and create brand-new virtual servers from scratch to replace old physical

machines, you simply may not have the extra time to invest to make the clean break to new versions of Linux with all-new packages. In many cases, it makes more sense to create a virtual clone of a physical server and keep all the files and services identical. If you want to go this route, there are a number of different methods you can use if you can handle unlimited downtime, but if you want to keep downtime and overall disruption to a minimum (and who doesn't?), there is a fairly simple process I've used to migrate a large number of servers to virtual machines with around 30 minutes of average downtime—well within most acceptable maintenance windows.

Caveats and Limitations

Before I get into the actual steps, there are some limitations to this approach that I should mention. First, this method has been designed and tested to work with VMware virtualization, specifically with its enterprise server products (although it would also work fine with their free server product). VMware works well for this process because it doesn't require that I modify my current Linux kernel to virtualize it—something that isn't always possible when you want to virtualize an old server. Having said that, these steps also could work with other virtual machine technologies that can use an unmodified Linux kernel. Second, this procedure has been tested with and is aimed at Red Hat-like distributions (Red Hat, CentOS and so on), but with a few tweaks I discuss later, it also could work with other distribution flavors. Finally, the actual amount of downtime you will need for this process probably will vary from my results, especially as you first test out each of the steps. Servers with large or slow disks and, specifically, servers that change large amounts of data frequently possibly will take longer to sync.

Along with these disclaimers, it's only fair to point out some of the benefits to this method:

- You can do a majority of the migration work against a live server.
- Standard Linux tools are used for the synchronization and other changes.
- The process protects your network from both servers showing up at the same time.
- You safely can leave the old server on the network and access its files while users use the new virtual machine.

Now that the disclaimers are out of the way, let me summarize the process in a few general steps. First, create a fresh virtual machine to replace the physical host. Then, boot in to the virtual machine with a rescue CD and partition the disk. Next, perform the main synchronization of files from the live physical server to the new virtual machine. Take down your physical machine and reboot in to a rescue CD, and perform a final synchronization from the off-line server. After that, change the boot settings on the virtual machine to suit its new environment, and then reboot in to your new virtual machine.

Create the Virtual Machine

To get started, first you must create a virtual machine container to replace your physical machine. Specific steps are different if you use VMware Server versus Virtual Infrastructure 3, but ultimately, what you want to do is to create a machine that mostly matches your physical machine's specifications. The specifications don't have to match exactly, and there actually are good reasons why you might want to tweak the settings a bit. For instance, if your server has 2GB of RAM, but you notice that it really needs only 1GB, now would be a good opportunity to change it. If your server is starting to run out of storage, this is a good time to increase it. If your physical server has a 32-bit or 64-bit processor, however, make sure the

virtual machine matches. Also, be sure that you match the operating system version you report to VMware with your actual OS if possible. For instance, if your server runs RHEL 3, don't tell VMware that it runs RHEL 4. You want to ensure that the OS will have drivers for the virtual devices that VMware presents, specifically for the disk subsystem. For instance, I've had numerous headaches due to RHEL 4's removal of the BusLogic SCSI module from the base OS (a virtual SCSI device that is a commonly used by VMware along with an LSI Logic virtual SCSI device).

After you set the specs for the virtual machine, edit the CD-ROM device so that it points either to an actual rescue CD in the VMware server or to an ISO. I prefer Knoppix for this procedure, but any live CD should work as long as it has the rsync and chroot tools, an SSH server and enough module support to access the disks on both the physical machine and the virtual machine. Now, boot the virtual machine into the rescue CD. Everything you need to do is done via the command line, so under Knoppix, type `knoppix 2` at the boot: prompt to bypass the GUI and go straight to a command line.

Partition the Virtual Machine's Disk

After Knoppix boots, you need to partition, format and mount the new partitions for this virtual machine. Use `fdisk` or `cdisk` from the command line to create your partitions to match your physical server. Again, you don't have to match the partition sizes exactly, as long as there is plenty of room to store all the files from the physical server. For this example, I will have a physical server with a single SCSI drive (`/dev/sda`) with three partitions: `/dev/sda1` for root, `/dev/sda2` for swap and `/dev/sda3` for `/home`. After you create the same partitions on the virtual machine, format them with the same filesystems you use on the physical machine, create mountpoints for them and then mount them:

```
$ sudo mkfs -t ext3 /dev/sda1
$ sudo mkfs -t ext3 /dev/sda3
$ sudo mkswap /dev/sda2
$ sudo mkdir -p /mnt/sda1 /mnt/sda3
$ sudo mount /dev/sda1 /mnt/sda1
$ sudo mount /dev/sda3 /mnt/sda3
```

First Sync

Now that you have created and mounted the partitions, you are ready for the first synchronization. For this to work, your virtual machine must have network access, and specifically, it needs to be able to access SSH on the physical machine. By default, Knoppix will attempt to get a DHCP lease if available, but otherwise, if your rescue disc is not able to get on the network, you need to make the necessary changes so that it can. This virtualization procedure reduces downtime by synchronizing the files twice—once while the physical server is running and once after it is off-line. The idea here is that a majority of files on most servers stay the same, at least over one or two days. If you perform the bulk of the file synchronization while the server is on-line, when you take it off-line, the final synchronization can occur much faster.

I use `rsync` for the synchronization, and for it to work, you need to allow (at least temporarily) for root SSH logins to occur on the physical machine. If it is disabled, edit

FEATURE Virtualize a Server with Minimal Downtime

`/etc/ssh/sshd_config` and change `PermitRootLogin` `no` to `PermitRootLogin yes`, and restart `sshd`. Otherwise, it will be difficult for `rsync` to copy all the files on the system. You will run an `rsync` command for each partition on the physical server, so in this example, that makes two `rsync` commands:

```
$ sudo rsync -avx --numeric-ids
↳--progress physicalhost:/mnt/sda1/
$ sudo rsync -avx --numeric-ids
↳--progress physicalhost:/home/ /mnt/sda3/
```

The `rsync` options I use here are chosen very deliberately, so it's worth understanding what each of them does. The `-a` option sets "archive mode", which essentially turns on a number of `rsync` options that preserve file ownership and permissions and other settings. The `-v` option makes `rsync` provide more output about what it is doing, and the `--progress` argument displays a progress meter so you can keep up with how long `rsync` will take. The other two arguments are rather important, and if you don't use `rsync` regularly, you might not come across them much. The `-x` argument tells `rsync` to stick to one filesystem. This is important particularly when you back up the `/` partition; otherwise, `rsync` happily will traverse into `/home` or any other partitions you have and copy them all into your local `/mnt/sda1` mountpoint, which probably will not have enough space to hold everything. The `--numeric-ids` argument sets file permissions on the destination files based on their numeric ID and not the matching user or group name. This is important as the Knoppix CD very likely has different user and group ID mappings than your server.

After these `rsync` commands complete, you are ready to take your physical server off-line. If you did need to schedule a maintenance window for the physical server, just leave the virtual machine running in its current state, and proceed to the next step when you are ready to take the physical machine off-line. If a number of days will pass until your maintenance window, you might want to run the above `rsync` commands again once you are close to the maintenance window, just so the final off-line `rsync` will happen more quickly.

Second and Final Sync

On the Physical Server: The last synchronization happens when the physical server is completely off-line, so you can make sure that no other files change on you. To do this, simply take a Knoppix CD (or your preferred rescue CD) to the physical machine and boot from it. All the commands you run will be from the command line, so you can boot in to Knoppix's terminal-only mode here as well. As Knoppix boots, it should detect your partitions automatically and create mountpoints under `/mnt` for them, but if it doesn't, just use the `mkdir` command to create them manually. Knoppix will not mount partitions automatically at boot, so you need to do that manually. In the case of this example, my physical server has two partitions to mount:

```
$ sudo mount /dev/sda1 /mnt/sda1
$ sudo mount /dev/sda3 /mnt/sda3
```

Now I need to set a password for the root Knoppix user

and then start the SSH server on this machine so I can run the `rsync`:

```
$ sudo passwd
$ sudo /etc/init.d/ssh start
```

Keep in mind that because I booted this machine into Knoppix, it most likely has gotten a different IP address via DHCP. Type `/sbin/ifconfig` to check which IP address the machine currently has, as you will need it for the final `rsync`.

On the Virtual Server: You now can start the final synchronization from the virtual server. The commands are very similar to what you used before, except this time, I add the `--delete` option so that `rsync` will remove any files on the virtual machine that were deleted from the physical machine since the last time I synced. Also notice that because the physical server is now booted in to Knoppix, I have to change the directory paths and the IP address for the remote host, as they changed since I booted in to Knoppix:

```
$ sudo rsync -avx --numeric-ids --progress
↳--delete 192.168.1.150:/mnt/sda1/ /mnt/sda1/
$ sudo rsync -avx --numeric-ids --progress
↳--delete 192.168.1.150:/mnt/sda3/ /mnt/sda3/
```

These commands could take a long time or a short time, depending on how many files have changed since the last time you ran `rsync`. Once it completes, you are ready to perform the final finishing touches on your virtual machine before bringing it into service.

Tweak Boot Settings

Even though the files on the virtual machine are identical to the physical machine, the virtual machine will not boot correctly at this point until you make some changes to the boot settings. This works best from within a chroot environment, so type:

```
$ sudo chroot /mnt/sda1
```

before you run the rest of the commands. Be sure to replace `/mnt/sda1` with the mountpoint for your root partition if it is different.

GRUB or LILO Changes

The first change you need to make within the chroot environment is to restore your bootloader. If you use GRUB, look at `/boot/grub/menu.lst` or `/boot/grub/grub.conf`. If you use LILO, look at `/etc/lilo.conf`. Check for any devices that may have changed. In particular, if you switched from an IDE to a SCSI device, from a RAID to a non-RAID or changed the root partition order, be sure to make changes here to reflect that. Next, if you use GRUB, type:

```
# grub-install /dev/sda
```

Change `/dev/sda` to match the primary disk device from which you will boot. If you use LILO, type:

```
# /sbin/lilo
```


After your bootloader has been installed, check `/etc/fstab` and confirm that any drive, partition or device changes you made in your bootloader config file also were changed here.

Re-create the `initrd`

Many servers these days use an `initrd` file to load modules that are essential for the boot process but that don't necessarily fit in the kernel image. Often, this `initrd` file contains only the modules that suit your hardware, so when you make the switch to new hardware, such as is the case with VMware's virtual SCSI controllers, you need to create a fresh `initrd` that has these new modules in it.

On a Red Hat system, edit either `/etc/modules.conf` or `/etc/modprobe.conf` for RHEL 4, and remove any references to `scsi_hostadapter` you find there. If you configured your virtual machine to use VMware's virtual BusLogic SCSI controller, replace those references with the following:

```
alias scsi_hostadapter BusLogic
```

If you chose VMware's LSI Logic SCSI controller, add the following lines instead:

```
alias scsi_hostadapter mptbase
alias scsi_hostadapter1 mptscsih
```

Obviously, these modules are specific to VMware virtualization, so if you want to attempt this with another virtualization technology, you will need to look up which SCSI modules it uses and make sure they are referenced here.

Now, you are ready to create a new `initrd`. Find the location of the `initrd` your server last used from your `/boot/grub/menu.lst`, `/boot/grub/grub.conf` or `/etc/lilo.conf` file, and then move it out of the way so you can create a new one safely. Then, run `mkinitrd` with the path to the `initrd` file to create and the name of the current kernel. For my example server, I am using the Red Hat 2.4.21-32.0.1.ELsmp kernel, so I would type:

```
# mv /boot/initrd-2.4.21-32.0.1.ELsmp.img
  ➔ /boot/initrd-2.4.21-32.0.1.ELsmp.img.bak
# mkinitrd /boot/initrd-2.4.21-32-0.1.ELsmp
  ➔ 2.4.21-32-0.1.ELsmp
```

As I said before, this is the method Red Hat uses to create `initrd` files. Unfortunately, different distributions use different methods. For instance, Debian's `mkinitrd` stores configuration files under `/etc/mkinitrd`, and the `mkinitrd` command uses slightly different options, so you might need to do some extra research to create a new `initrd` for your server's distribution.

At this point, you can reboot the virtual machine. Confirm that your physical machine no longer has its original IP address, or otherwise, simply power it off to be safe. If your server runs a hardware configuration service like `kudzu`, it most likely will prompt you at boot time because it has detected changes in the server's hardware. Be sure to select `Keep Configuration` for any old SCSI or network hardware it mentions, and select `Ignore` for any new SCSI or network hardware; however, you safely can remove old video, sound, USB and similar hardware if you are prompted.

Once the machine has booted completely, confirm that all system services have started and that you are connected to the network. I have noticed on some Red Hat systems that the network card's MAC address has been hard-coded into the configuration file, and as that has changed on the new virtual hardware, the network won't resume. In this case, simply edit the configuration file for your network card under `/etc/sysconfig/network-scripts/` (often `ifcfg-eth0`), and either remove the reference to the MAC address or change it to reflect the new MAC address. Then, restart the networking service.

Practice this procedure on a few test machines to be sure you have all the steps down for your particular network before attempting it on a live production machine. Nothing is worse than scrambling to fix strange `initrd` issues on a virtual machine while the physical server is down and your maintenance window is ticking away. You will find that the more often you perform these migrations, the faster you can do them—you even might be able to stagger them and complete a few at the same time. ■

Kyle Rankin is a Senior Systems Administrator in the San Francisco Bay Area and the author of a number of books, including *Knoppix Hacks* and *Ubuntu Hacks* for O'Reilly Media. He is currently the president of the North Bay Linux Users' Group.

COMPACT EMBEDDED SERVER



- x86 200MHz CPU
- 128MB SDRAM On Board
- 512MB CompactFlash™
- 10/100 Base-T Ethernet
- Reliable (No CPU Fan or Disk Drive)
- Two RS-232 & Two USB 1.1 Ports
- Optional Wireless LAN & Hard Drive
- Optional On Board Audio
- Dimensions: 4.5 x 4.5 x 1.375" (115 x 115 x 35mm)



2.6 KERNEL

Compact SIB (Server-In-a-Box) Starting at \$198.00 Quantity 1.

- EMAC Linux 2.6 Kernel
- Menu Drive Configuration Utility
- Eclipse Development Environment
- HTTP and FTP Servers
- PPP Dial In/Out Server & Client
- Telnet Server

Since 1985 OVER 22 YEARS OF SINGLE BOARD SOLUTIONS

EMAC, inc.

EQUIPMENT MONITOR AND CONTROL

Phone: (618) 529-4525 • Fax: (618) 457-0110 • www.emacinc.com

Building a Multisourced Infrastructure Using OpenVPN

How to use OpenVPN to take your hosting to the next level.
Dmitriy Samovskiy

Have you ever needed to expand your colocated servers at more than one provider and allow applications to communicate as if they were on the same LAN, possibly over multiple sets of firewalls and layers of NAT? Or, maybe you've wanted to move from one hosting service to another to take advantage of lower pricing or better uptime but would have preferred to do it gradually instead of in a single swoop (and a weekend-long maintenance window)? Or, maybe you've considered the Amazon EC2 cloud to host part, but not all, of your infrastructure? If your answer to any of these questions is yes, what you want is essentially a multisourced infrastructure.

Let's take a look at a simple distributed application, which consists of multiple services, a LAMP stack. Traditionally, you would start with Apache and MySQL on a single server. As your site grows, you would provision another server from your provider and add a second Apache instance. Later, you might want to provision yet another machine to be a dedicated database server to improve performance. This is a typical single-sourced

infrastructure—all services run within a single physical environment, controlled and supported by a single provider.

In contrast, with a multisourced infrastructure, you no longer are limited to one provider or one data center. You are free to mix and match hosting plans from different providers to suit your business and architecture better, and you can use as many providers as you like. Your applications still can communicate with one another, but instead of having a physical LAN, it's now a virtual LAN that sits on top of public Internet links. You can grow your services horizontally and achieve better geographic redundancy and fault tolerance at the same time, all without significant changes in your application. If it works in a single-sourced physical LAN, it most likely will work in multisourced virtual LAN as well.

Additionally, you can leverage the strengths of a particular provider for just a subset of your services. Going back to the LAMP stack as our example, with Amazon EC2, you can provision many Apache instances in response to the current load quickly; although you might prefer to run MySQL on bare

The Amazon EC2

The Amazon EC2 (Elastic Compute Cloud) is a Web service that allows users to provision new machines in an Amazon-hosted virtualized infrastructure in a matter of minutes, using a publicly available API. Users get full root access and can install almost any OS or application in their Amazon Machine Images. Web service APIs

allow users to reboot their instances remotely and scale capacity quickly if necessary, by adding tens or even hundreds of machines. Additionally, there is no up-front hardware setup costs—Amazon charges only for the capacity you actually use; there is no minimum fee. As more applications find their way to Amazon's virtual computing

environment, system administrators are looking for ways to provide secure connectivity over the public Internet between new machines in the Amazon EC2 and old machines in their regular data centers. This article describes one such technique—how to build a multisourced infrastructure based on OpenVPN.

metal elsewhere instead of in an EC2 virtual machine.

Finally, this method allows you to expand your corporate infrastructure outside your current data center or allow outside services to use applications in your corporate data center. Consider a remotely hosted data-crunching cluster that you rent by the hour, which uses your corporate data warehouse system for its input. As you can see, a multisourced infrastructure is more flexible and can accommodate various scenarios and needs.

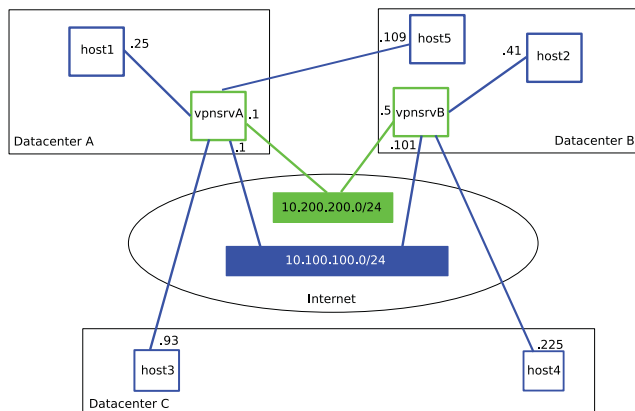


Figure 1. Multisourced Infrastructure: OpenVPN Virtual Links

In this article, I describe a particular implementation of the multisourced infrastructure concept that we at CohesiveFT (www.cohesiveft.com) developed using OpenVPN and that has been running in our production environment since mid-summer 2007. We chose OpenVPN primarily because it uses standard OpenSSL encryption, runs on multiple operating systems and does not require kernel patching or additional modules. The latter benefit is of key importance. Many Virtual Private Server (VPS) hosting solutions currently provide great service with pricing that is often better than other forms of hosting. These providers build guest OS kernels specifically tailored for their environment and method of virtualization. As a result, you probably want to avoid rebuilding the Linux kernel on your VPS as much as possible. Not that it can't be done, but you can save some time and probably get faster technical support if you don't do it.

Among the alternatives to OpenVPN, there is Openswan, a code fork of the original FreeS/WAN Project, but it requires a kernel patch to support NAT traversal, according to its wiki (wiki.openswan.org/index.php/Openswan/Install).

The OpenVPN protocol also is firewall-friendly, as it can pass all traffic over a single UDP tunnel (the default port is 1194). That feature, coupled with SSL encryption, makes this solution very difficult to attack when data packets pass through the public Internet.

OpenVPN turned out to be a great choice and offered us all the functionality we expected, except for one very important feature, fault tolerance. When you use a VPN to provide corporate network access to remote users, the solution is very simple—you deploy several OpenVPN servers and configure each server with its own network segment (for example, server 10.5.0.0 255.255.0.0 and server 10.6.0.0 255.255.0.0). In a typical scenario, the dynamic IP address assigned to a remote user will not matter much, as long as you configure firewalls, applications and services to allow both subnets.

When you build a multisourced infrastructure, however, this is not an acceptable solution, unless you want servers to change their IP addresses from time to time. To satisfy redundancy and fault-tolerance requirements, we needed an active-active pair of OpenVPN servers to share a common address space—all hosts must be able to access each other by static IP addresses at all times, no matter which OpenVPN server provides connectivity at either end of the communication. Then, if we lose one OpenVPN server, the other will provide all connectivity. And, if they are both up, both will be accepting connections from clients to share the load. This feature was not available as a part of the OpenVPN source distribution, so we developed a standalone dynamic routing daemon to facilitate active-active load balancing. You can find its source code, along with useful links, use-case scenarios and mailing lists, at www.cohesiveft.com/multisourced-infra.

Building Your Virtual LAN

You need two machines to run the OpenVPN daemon in server mode (we refer to them as vpnsvrA and vpnsvrB, and let's assume their physical IP addresses in your network are 192.168.7.1 and 192.168.17.1, respectively) and two new private subnets: data (for example, 10.100.100.0/24) and management (10.200.200.0/24). All of your applications and services will run in the data subnet, and vpnsvrA and vpnsvrB will exchange runtime status and routing information in the management subnet. Think of these two machines as virtual network switches for your virtual LAN. Also, note that these subnets do not have to be class C; you can choose a bigger data network, especially if you are planning to connect a large number of hosts.

Listing 1a. OpenVPN Server Configuration for vpnsrvA

```
server 10.100.100.0 255.255.255.0

ifconfig 10.100.100.1 10.100.100.2

push "route 10.100.100.0 255.255.255.0"
push "route 10.200.200.0 255.255.255.0"
dev tun
proto udp
user nobody
persist-key
persist-tun
dh keys/dh1024.pem
ca keys/ca.crt
cert keys/vpnsrvA-1.crt
key keys/vpnsrvA-1.key
comp-lzo
verb 3
keepalive 10 60
client-config-dir ccd
management tunnel 5656 /etc/openvpn/pass
```

Listing 1b. OpenVPN Server Configuration for vpnsrvB

```
mode server
tls-server
ifconfig 10.100.100.101 10.100.100.102
ifconfig-pool 10.100.100.410 10.100.100.251
route 10.100.100.0 255.255.255.0
push "route 10.100.100.0 255.255.255.0"
push "route 10.200.200.0 255.255.255.0"
dev tun
proto udp
user nobody
persist-key
persist-tun
dh keys/dh1024.pem
ca keys/ca.crt
cert keys/vpnsrvB-1.crt
key keys/vpnsrvB-1.key
comp-lzo
verb 3
keepalive 10 60
client-config-dir ccd
management tunnel 5656 /etc/openvpn/pass
```

Configure vpnsrvA and vpnsrvB as OpenVPN servers for the data subnet (Listings 1a and 1b). You may add more configuration options as needed. Note that the “server” line in the configuration file is a shortcut and cannot be used for both vpnsrvA and vpnsrvB. It actually expands to a set of commands that would have assigned 10.100.100.1 to both servers (see the OpenVPN man page for more details). We want an active-active configuration; therefore, we need vpnsrvA and vpnsrvB

Listing 2. OpenVPN Client Configuration

```
# Note: "remote" must point to servers' physical
# (not virtual) IP addresses
client
remote 192.168.7.1
remote 192.168.17.1
dev tun
proto udp
user nobody
persist-key
persist-tun
keepalive 10 60
comp-lzo
ca keys/ca.crt
cert keys/client-1.crt
key keys/client-1.key
ns-cert-type server
```

Listing 3. Parts of the Routing Table on vpnsrvA

10.100.100.2	0.0.0.0	255.255.255.255	UH	0 0 0	tun0
10.100.100.0	10.100.100.2	255.255.255.0	UG	0 0 0	tun0

to be in the same subnet but to have different IP addresses. To accomplish this, we explicitly expand the server definition for vpnsrvB and assign it the 10.100.100.101 IP address.

Another important note is that the client configuration directory (usually called ccd) and keys directory (called keys) need to be identical on both vpnsrvA and vpnsrvB. One of the easiest ways to accomplish this is to use rsync. rsync allows us to keep it simple and avoid extra variables in the mix. Plus, we always can switch the direction of rsync and promote either of the servers to be the master. For now, let’s assume that vpnsrvA is the master and that vpnsrvB mirrors the ccd and keys directories from vpnsrvA using rsync. You will create keys (preferably using the easy-rsa package that ships with OpenVPN) and update the ccd entries on the master server.

At this point, you can configure several hosts on your network as OpenVPN clients (Listing 2). Each host will have its own certificate/key pair, and the ifconfig-push directive in the ccd entry for this host will set its IP address (see Resources for a link to the OpenVPN HOWTO for a detailed explanation of how to set it up). We tie the virtual IP address to a host based on its certificate/key pair, in much the same way as in a DHCP configuration you would tie an IP address to a host based on its Ethernet MAC address. Therefore, each client must have its own unique certificate/key pair.

Note that we use OpenVPN’s built-in capability to round-robin between multiple servers and reconnect after connectivity failures, which is controlled by the keepalive option. Once this is done, you should be able to start the OpenVPN clients, and they should at least be able to communicate with their current OpenVPN server and refer to it by IP—10.100.100.1 or 10.100.100.101. If your client connects to vpnsrvA and you bring down the openvpn daemon on vpnsrvA, the client will detect it

and automatically reconnect to vpnsrvB.

A quick note about firewalls—in a virtual LAN, your main data interface will be called tun0. Therefore, all the rules you used to define for interface eth0 in a single-sourced configuration will need to be redefined for tun0. The Ethernet interface, however, will require additional rules to allow UDP on port 1194 (OpenVPN) from the client machines to both vpnsrvA and vpnsrvB.

The setup that we already have accomplished is somewhat fault-tolerant. If vpnsrvA becomes unavailable, all clients will reconnect to vpnsrvB, and connectivity will be restored. In other words, this is active-passive redundancy. But, what will happen if both vpnsrvA and vpnsrvB are up? Let's assume that host1 and host2 run the openvpn daemon in client mode. host1 connected to vpnsrvA and was assigned 10.100.100.25; host2 connected to vpnsrvB and was assigned 10.100.100.41. The routing table on vpnsrvA is shown in Listing 3. In this scenario, when host1 attempts to ping 10.100.100.101, its outgoing packets will be routed first to vpnsrvA but then will go back to the same tun0 interface, because vpnsrvA does not know about the existence of vpnsrvB. Similarly, when host1 attempts to ping host2, vpnsrvA also will send these packets back, as indicated by the 10.100.100.0/24 route. As a result, both operations will fail.

To address this issue, we developed a dynamic routing daemon called cube-routed (download it from www.cohesiveft.com/multisourced-infra). It shares routing information between vpnsrvA and vpnsrvB and adjusts routing tables depending on which client connects to which server in near real time. Its internal structure is not very complex. One thread connects to a local OpenVPN daemon process via its management interface (see the management option in the OpenVPN configuration file) and regularly runs the status command to update the list of clients connected locally. Another thread publishes this information for the remote instance of cube-routed. The third thread regularly reads a list of connected clients from the remote instance of cube-routed. Finally, the fourth thread adjusts the local routing table based on the following two rules: 1) adds a host route for every host connected to the remote OpenVPN server and 2) deletes the host route for every host connected to the local OpenVPN server.

cube-routed instances will exchange

Never tried SlickEdit? What's your excuse?

Top 10 Reasons *NOT* to use SlickEdit

10. The more I type, the better I get at it.
9. The office is so peaceful after 10PM.
8. Real programmers don't need no fancy tools.
7. Nothing beats formatting my code by hand.
6. I've always wanted Repetitive Stress Injury.
5. I love reading library reference manuals.
4. Milestones? What are milestones?
3. What would life be without syntax errors.
2. If I finish all my work on time, I'll just get more.
1. I've never tried it.

There's really only one good reason not to use SlickEdit — *you've never tried it.*

How can we be so sure? User feedback. For 19 years we've listened to power programmers and continued to try to exceed their expectations. Our multi-language development tools and advanced code editors are fast, powerful, and flexible — whether you're developing on Windows, Linux, UNIX, or Mac OS. Whatever your environment, SlickEdit has a tool for you. Our product suite includes SlickEdit®, SlickEdit® Tools for Microsoft® Visual Studio® 2005, and the SlickEdit® Plug-in for Eclipse™.

slickedit.

www.slickedit.com

Download a FREE trial at www.slickedit.com/trial
or call 1-800-934-3348.

Listing 4a. vpnsrvA cube-routed Configuration File

```
vpnsrvA
mgmt_interface = tun1
data_interface = tun0
remote_mgmt_ip = 10.200.200.5
remote_data_ip = 10.100.100.101
openvpn_mgmt_pass_file = /etc/openvpn/pass
openvpn_mgmt_port = 5656
cube_routed_port = 5657
```

Listing 4b. vpnsrvB cube-routed Configuration File

```
mgmt_interface = tun1
data_interface = tun0
remote_mgmt_ip = 10.200.200.1
remote_data_ip = 10.100.100.1
openvpn_mgmt_pass_file = /etc/openvpn/pass
openvpn_mgmt_port = 5656
cube_routed_port = 5657
```

information over the management subnet we selected earlier. Create a second tunnel tun1 between vpnsrvA and vpnsrvB. vpnsrvA can be a server with IP 10.200.200.1, and vpnsrvB is its client with IP 10.200.200.5. You can use the configuration files from Listings 1 and 2 as a basis, but remember to adjust the IP addresses and select a different port—for example, you could add port 11940 to both the server and client. Start both OpenVPN daemons, and use ping 10.200.200.1 and ping 10.200.200.5 to verify connectivity between them.

Now, create configuration files for cube-routed on both vpnsrvA and vpnsrvB, as shown in Listings 4a and 4b, and start both instances as root with the path to the configuration file as the only parameter (note that OpenVPN must already be running, and the tun0/tun1 interfaces on both vpnsrvA and vpnsrvB must be up).

Once you start everything and after several minutes of initial convergence time, host1 from the example above will be able to communicate with host2, even though they connected to different OpenVPN servers. Thus, you've achieved a fully fault-tolerant virtual LAN connectivity, with data traffic encryption as an added bonus.

LINUX JOURNAL IS LOOKING FOR A PART-TIME TECHNICAL EDITOR TO JOIN ITS TEAM OF EXPERIENCED WRITERS/STAFF

Linux enthusiasm is a must! Editing for technical accuracy, writing, project management and people skills required. Responsibilities include securing editorial content, interviewing, conducting product reviews, researching and writing articles, and promoting the magazine at events.

Please send your résumé, cover letter and two writing samples to ljeditor@linuxjournal.com.

Conclusion

This implementation is not without its limitations. First, applications that use broadcast or multicast will not work with OpenVPN's tun device. You can use the same network layout as described here, but instead of tun, experiment with OpenVPN's tap device to work around this. Second, latency of network links over the public Internet is significantly higher than that of Ethernet. If this is an inherent requirement for your application, you probably should leave this part of your infrastructure single-sourced. Third, because we use UDP-based tunnels, OpenVPN links will tend to go up and down more often than Ethernet, especially during times of network congestion. You can implement data caches, avoid long-lived TCP connections, focus on network exception-handling logic and experiment with TCP tunnels to reduce negative impact. Finally, there are exactly two OpenVPN servers in this setup. This generally should be sufficient, as it doesn't affect the number of actual hosts that you have connected to your multi-sourced infrastructure. If for some reason you need more than two, it becomes much more difficult to implement route sharing among cube-routed instances. In that case, you might want to consider a messaging system instead of raw sockets (for example, RabbitMQ). All in all, in our case, we found that the overall benefits of a multisourced infrastructure far outweighed the problems caused by these limitations, particularly if you design your architecture with these limitations in mind.

Multisourced infrastructure is a logical extension of its single-sourced predecessor, similar to the distributed service-oriented architecture, which came after monolithic applications and enabled greater flexibility, a faster development cycle and higher availability. It can help you design a smarter architecture and avoid a lock-in to a single hosting provider, on top of a standard time-tested open-source OpenVPN. ■

Dmitriy Samovskiy works at CohesiveFT (www.cohesiveft.com), an innovative maker of custom virtualized application stacks, where he focuses on open-source technologies, distributed applications, systems integration, Python and Ruby. You can reach him at dmitriy.samovskiy@cohesiveft.com.

Resources

OpenVPN: openvpn.net

OpenVPN 2.0 HOWTO: openvpn.net/howto.html

"Meet OpenVPN" by Hans-Cees Speel: www.linuxjournal.com/article/7949

"Introduction to OpenVPN" by David Bogen: www.osnews.com/story.php/5803/Introduction-to-OpenVPN

Openswan: www.openswan.org

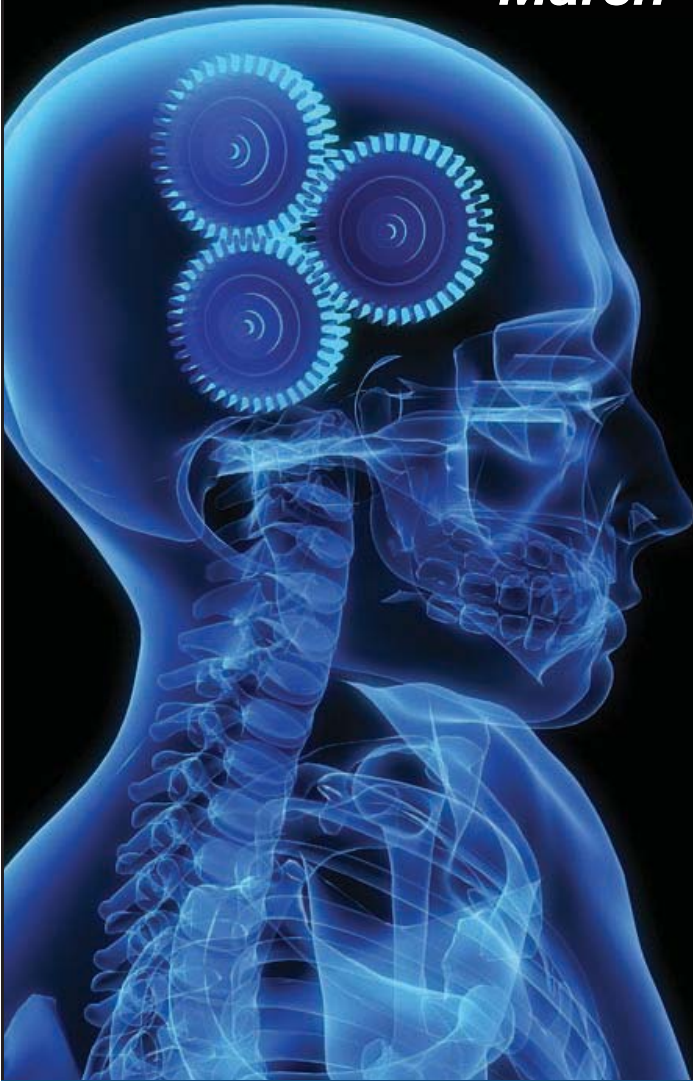
cube-routed: www.cohesiveft.com/multisourced-infra

Amazon EC2: aws.amazon.com/ec2

RabbitMQ: www.rabbitmq.com

Got ideas? Do it with PHP.

March 12-14, 2008 / Montreal, Canada



 Develop cool & efficient
Rich Internet Applications

 Secure your web
applications

 Build large & scalable
PHP apps

 and much
more...

Register online and get a 100\$ discount
<http://conf.phpquebec.com/lj>

Registration ends
March 7th



The conference
is sponsored by



ActiveState



Digging Up Dirt in the DNS Hierarchy, Part II

The examples used here were not invented. This article is really, really scary. RON AITCHISON

In the first part of this article [LJ, January 2008], we started the apparently simple journey of navigating our way to the IP address of `www.example.com` and its secure server `online.example.com` by traveling down the DNS hierarchy. We finally received a referral from the gTLD `.com` servers pointing us to the name servers `ns2.example.com`, an in-zone name server, and `ns1.example.net`, an out-of-zone (or out-of-bailiwick) name server.

So, let's restart our quest for the IP address of `www.example.com` and assume we have verified and obtained the IP address of `ns1.example.net`, which, because it is out-of-zone, we have tracked to its authoritative source via the `.net` gTLD servers. Now, it's time to check all our authoritative servers for the `example.com` domain to see what else we can find. First we check the front door:

```
dig @ns1.example.net version.bind txt ch
```

This command uses a legacy DNS resource record class called CH(AOS)—Internet addresses use the IN class—to try to obtain information about the software being used. We get this response:

```
; <<>> DiG 9.4.1-P1 <<>> @ns1.example.net version.bind txt ch
...
;; -->HEADER<<- opcode: QUERY, status: NOERROR, id: 8503
;; flags: qr aa rd ra; QUERY: 1, ANSWER: 1, AUTHORITY: 0, ADDITIONAL: 0
;; QUESTION SECTION:
;version.bind. CH TXT
;; ANSWER SECTION:
VERSION.BIND. 0 CH TXT "named 4.9.6-Rel-Tuesday-24-June-97..."
;; Query time: 25 msec
;; SERVER: 207.253.126.250#53(207.253.126.250)
...
```

And, we got lucky. This name server is telling us the supplier and version number of its software. If we were bad guys, we

would go and look up the alerts for this version of software and see if there were any juicy vulnerabilities. In this case, the news is extremely good (for the bad guys), because the server is running BIND 4, last updated in 1997! Between 3% and 7% of all the estimated 9 million name servers in operation still use this redundant software, which is full of bugs and exploit possibilities. Let's assume we repeat the command, substituting `ns2.example.com`—our second authoritative server—and we get back "my name is Bind, James Bind". The administrator of this DNS has a sense of humor and some knowledge of BIND configuration parameters. In the options clause of BIND's `named.conf` file (normally in `/etc/named.conf`), the following parameter will appear:

```
options {
...
version "my name is Bind, James Bind";
...
}
```

You can place any text here, and it will be supplied in place of the version number. If the statement is missing, BIND will return its version number, as shown in the previous example. As we shall see, this may not prevent us from discovering the information, but it does at least make it more than a trivial one-line command. Finally, although BIND servers respond to the above command, not all DNS software does, so we could have received a timeout.

Now, it's time to move on to the next check. We're going to use the second of our tools, `fpdns`, which is a DNS fingerprinting tool (see Part I of this article for more information on `fpdns`). `fpdns` uses a range of benign techniques to try to identify both the software vendor and, in many cases, the release version or version range. It is not infallible, but its accuracy is extremely good. Let's use it to check our reluctant Mr Bind:

```
fpdns ns2.example.com
```

And, we get the following:

```
fingerprint (ns2.example.com, 10.10.0.2): ISC BIND
9.2.0rc7 -- 9.2.2-P3
[recursion enabled]
```

Now, this potentially is serious as well. First, the current version of BIND at the time of this writing is 9.4.1-P1. So, we can simply check the security alerts for the version range quoted and see whether we have some handy poisoning possibilities. Second, this server is an open recursive server, which means that any request for name resolution will be accepted and acted on by this server, not only the names for which it is authoritative. We could test this using a dig command like the following:

```
dig @ns2.example.com some.obscure.domain
```

Why are open resolvers a serious problem? There are three reasons. First, we can load up the server for a simple Denial-of-Service (DoS) attack by sending it requests for external name resolution. It will be so busy following the referral chains that it will not have time to answer requests for the domain for which it is authoritative—effectively taking the domain off the air for at least a proportion of the traffic. Second, it can be used in Distributed Denial-of-Service attacks. In this type of attack, requests are sent for the same name to many open name servers (there are perhaps as many as one million open name servers on the Internet), each of which then sends a query to the DoS tar-

Between 3% and 7% of all the estimated 9 million name servers in operation still use this redundant software, which is full of bugs and exploit possibilities.

get. No one single request breaks any threshold monitoring, so it is difficult to identify all the sources. The net effect is that the target DNS is bombarded with traffic and cannot respond. Third, if I send a query to an open name server, I know what it is going to do; it's going to send a query to the target domain's name server. So, without even sniffing its traffic, I can start sending spoofed responses, and if I get lucky, I can poison the open server's cache (there are many documented weaknesses that I can exploit to increase my chances significantly).

The function of a caching server is to save the response until the Resource Record's TTL (Time to Live) expires and then re-read the record. If the TTL for the requested RR is long (30 minutes or more), I have a poisoning opportunity only every 30 minutes or more, but if the TTL is short, say, five seconds or even zero seconds, my odds of getting poisoned responses into the cache shoot up dramatically. And, of course, my poisoned response will not have a TTL of five seconds; it will be more like five weeks, so when it's there it stays there for a long time.

Now the real place to do this cache poisoning is not at the authoritative name server. Instead, I would go looking for

Hardware Systems for the open source community - Since 1989.

(Linux, FreeBSD, NetBSD, OpenBSD, Solaris, MS etc.)

The AMD Opteron[™] processors deliver high-performance, scalable server solutions for the most advanced applications. "Runs both 32-and 64-bit applications simultaneously".

Your Custom Appliance Solution!!

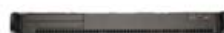
"Let us know your Needs..."

 "Your Logo Here"

"We will build you a Solution...."

AMD Opteron(TM) Value Server Starts at \$847

- 1U 14" Deep 260W.
- AMD Opteron 140 CPU.
- 512MB PC 3200 DDR ECC Unbuffered.
- Support upto 8GB DDR RAM.
- 40GB SATA Hard Disk.
- 2x 10/100 Mbps Lan.



Quad AMD Opteron(TM) Server Starts at \$2,812

- 1U AMD Opteron Model 840.
- 2GB Memory. Max 128 GB
- Supports upto 64GB FBDIMM.
- 80 GB SATA II Hotswap Hard Drive.
- 2x Integrated Dual 10/1000 LAN.



Dual AMD Opteron(TM) Storage Starts at \$4,120

- 5U Dual AMD Opteron Model 246.
- iSCSI or NAS Software Options.
- Support upto 18TB of Storage.
- Fail Hard Drive LED Indicator.



Dual AMD Opteron(TM) Storage Starts at \$8,445

- 8U AMD Opteron Model 246.
- 4TB of Storage (36TB Max).
- 1GB RAM
- 2 x 10/100/1000 Gigabit LAN.
- NAS or iSCSI Software Options.



Why Do Business With ASA?

"We Provide Approved EVAL Server..."

Since 1989, ASA has served customers like Cisco, Juniper, Caltech, Fermilab and most Universities. We provide a total custom solution with OS of your choice. Excellent pre and post-sales support.

"Reliable hardware at the most competitive prices".

Please call or contact us for your next hardware purchase.



2364 Calle Del Mundo, Santa Clara, CA - 95064

www.asacomputers.com

Email: sales@asacomputers.com

Tel: 1-800-REAL-PCS, Fax: 408-664-2910.



fpdns uses a range of benign techniques to try to identify both the software vendor and, in many cases, the release version or version range.

an open recursive name server at an ISP and try to poison the cache using the same technique, so that all the ISP's clients for `www.example.com` will come to my pharming site.

All name servers should be closed to external traffic to stop this behavior. If you are using BIND, there are three options:

1) If the name server is authoritative only (best practice recommends that you never mix caching and authoritative functions in the same DNS), add the following line to the `/etc/named.conf` file in the options clause:

```
options {
...
// BIND's default is recursion yes;

recursion no;
...
};
```

2) If your server does provide both authoritative and recursive services, limit who can use them by using the `allow-recursion` statement in an options clause:

```
options {
...
allow-recursion {192.168.2/24};
...
};
```

This statement limits the allowable IP addresses permitted to make recursive requests to `192.168.2.1–192.168.2.254`. It is worth pointing out that even if this statement is present, a recursive request from outside the defined IP range will return the correct result if it already exists in the cache (it previously was requested by a valid internal user). BIND 9's `view` clause also can be used to provide further control and separation in a mixed authoritative and caching configuration.

3) Finally, if the server only provides caching services, use the `allow-query` statement instead:

```
options {
...
allow-query {192.168.2/24};
};
```

Now, let's continue with our checks by requesting the IP address of our target from one of its authoritative servers:

```
dig @ns1.example.net www.example.com
```

And, we get this in response:

```
; <<>> DiG 9.4.1-P1 <<>> @ns1.example.net www.example.com
...

;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id:
49319

;; flags: qr rd ra aa; QUERY: 1, ANSWER: 2, AUTHORITY:
0, ADDITIONAL: 0

;; QUESTION SECTION:

;www.example.net. IN A

;; ANSWER SECTION:

www.example.com. 5 IN A 10.10.0.5

www.example.com. 5 IN A 10.10.0.6

;; Query time: 61 msec

;; SERVER: 192.5.6.30#53(192.5.6.30)
...
```

There are a couple of things to note in this response. First, the `aa` flag is set, which is what we would expect. If this flag is not set, this would be what is called a lame-server (a server defined in the parent as authoritative for a domain but that does not return the `aa` flag to a query for information in that domain). The master (primary) and slave (secondary) name servers for a domain must return the `aa` flag. There is no externally visible difference between master and slave server responses. This means you can use two or more slave servers to provide authoritative service and keep your master completely hidden. The second point to note is that the `ra` flag is set, thus offering a recursion service. Let's test it:

```
dig @dns1.example.net some.obscure.domain
```

And bingo, we get a response—this server is also open. The reason for using `some.obscure.domain` is to make sure the data is not already cached, in which case, depending on its configuration, the name server could return the desired results and still be closed as noted previously. Using an obscure name minimizes the possibility of a false positive. The corollary is also true. If we fire a request for a popular domain name, such as `google.com`, to an apparently closed DNS and get a valid result, we know this server is providing recursive services for some set of clients—unless of course it is the authoritative server for `google.com`! This knowledge gives us some, very modest, poisoning possibilities by looking at the TTL time of the response.

In passing, we also should note that the site sensibly has provided two IP addresses, albeit both on the same IP address block. This means that browsers automatically will fail over (in 2–3 minutes). If the first server fails, it uses a five-second TTL, which, apart from being of great assistance to potential cache poisoners, is entirely useless as Microsoft's browser will attempt to refresh its browser-cached IP addresses only every 30 minutes (one minute for Firefox).

So, `ns1.example.net` is using old, buggy software and is open. Can it get worse? Well, yes it can, and indeed, in this case, it does get worse.

So far, we have been emulating what a browser does and simply looking for ARRs; `dig` can be used to find any type of RR. In this case, the absence of an AUTHORITY SECTION is a tad suspicious, so let's experiment and try this command:

```
dig @ns1.example.net www.example.com ns
```

And, we get this response:

```
; <<>> DiG 9.4.1-P1 <<>> @ns1.example.net www.example.com ns
...
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 49319
;; flags: qr rd ra aa; QUERY: 1, ANSWER: 2, AUTHORITY: 0, ADDITIONAL: 2
;; QUESTION SECTION:
;www.example.com. IN NS
...
;; ANSWER SECTION:
www.example.com. 3000 IN NS ns3.example.com.
www.example.com. 3000 IN NS ns4.example.com.
...
;; ADDITIONAL SECTION:
```

```
ns3.example.com. 3000 IN A 10.10.0.8
ns3.example.com. 3000 IN A 10.10.0.9
...
;; Query time: 61 msec
;; SERVER: 192.5.6.30#53(192.5.6.30)
...
```

This result means that the user is trying to delegate `www.example.com` to an alternate set of DNS servers, `ns3` and `ns4.example.com`, but the delegation is invalid, so the defined DNS servers are not visible. The zone file probably has this construct:

```
$ORIGIN example.com.
...
; these A RRs should not be present in the example.com
; zone file but should be present in a www.example.com
zone file
www 5 IN A 10.10.0.5
www 5 IN A 10.10.0.5
...
; valid delegation for www.example.com
www 3000 IN NS ns3.example.com.
www 3000 IN NS ns4.example.com.
...
; required glue RRs for the delegation
ns3.example.com. 3000 IN A 10.10.0.8
ns3.example.com. 3000 IN A 10.10.0.9
```

BIND 9 (used by `ns2.example.com`) correctly will interpret this as a delegation and generate a referral to `ns3` and `ns4.example.com`. BIND 4 (`ns1.example.net`) will not, and thus, approximately 50% of the traffic will never even see the delegated servers, which if we perform our checks using the above techniques, we would see are solidly configured and using the latest versions of BIND (similarly with the name servers for `online.example.com`).

In summary, this user configured and maintained his or her internal name servers with great care and in a thoroughly professional manner but had entirely overlooked the route by which users arrived at the site. To put it another

TS-7800 High-End Performance with Embedded Ruggedness



\$269 qty 1 **\$229** qty 100

500 MHz ARM9

- New unbrickable design- 3x faster
- Backward compatible w/ TS-72xx
- Low power - 4W at 5V
- 128MB DDR RAM
- 512MB high-speed onboard Flash
- 12K LUT user-programmable FPGA
- Internal PCI Bus, PC/104 connector
- 2 USB 2.0 480 Mbps
- Gigabit ethernet
- 2 SD sockets
- 10 serial ports
- 110 GPIO
- 5 10-bit ADC
- 2 SATA ports
- Sleep mode uses 200 microamps
- Boots Linux in < 2 seconds
- Linux 2.6 and Debian by default

Design your solution with one of our engineers

- Over 20 years in business
- Never discontinued a product
- Engineers on Tech Support
- Open Source Vision
- Custom configurations and designs w/ excellent pricing and turn-around time
- Most products stocked and available for next day shipping

See our website for options, peripherals and x86 SBCs

INDEPTH

way, the castle was impregnable, the moat wide and deep, the walls thick, the defenses manned, but the front door wide open.

This problem may look pretty far-fetched, but it was real, took less than ten minutes to find and—in case you were wondering—is now fixed!

When performing this kind of analysis, you will develop your own methods and variations, but here are some things to look for that seem to make organizations especially vulnerable:

- Multiple domain names, for instance, example.com, secure-example.com and online-example.com, tend to make managing and monitoring more complex for the operator and, hence, are more likely to have DNS configuration errors.
- Backroom domains—many organizations elect to use unique domain names, for instance, support-example.com, to perform infrastructure functions, such as running their internal DNS systems, that are not visible to end users. For some strange reason, many of these organizations think end-user invisibility also applies to the DNS system.
- Many DNS servers—the more DNS servers, the more likely it is that at least one of them is running either badly configured or unpatched software.
- BIND 8 and open is a very common ISP configuration. BIND 8 is pretty buggy, represents approximately 20% of all DNS servers and is now officially deprecated. Whoopee for the bad guys.
- Always follow the transitive trust routes. The more there are, the more likely you are to find a problem.
- Outsourced DNS—there are highly professional DNS organizations to whom many large users subcontract a provision of DNS service and whose DNS configurations are invariably in very good shape. Many organizations use the outsourced DNS to delegate to internal DNS systems. These users can

exhibit the exact opposite characteristics of the example case—the internal name servers are a disaster. Further, in a surprising number of cases, even when outsourced, there is one internal name server or that of a local service provider on the primary authoritative list—almost invariably this additional name server has a problem.

The techniques used here are not aggressive; for example, they do not test for AXFR (zone transfer) vulnerability, because this not a friendly action and is likely to generate nasty responses, quite rightly, from DNS administrators. Treading lightly is the best technique.

We used a very small subset of dig's capability here. Read the man pages for more information. If you do find something suspicious or wrong, double-check, then either fix it immediately or, if it affects a third party, act responsibly and inform the relevant organization (though it is sometimes extremely difficult to get through to the right person). Theoretically, the SOA RR of the domain in question should contain the valid e-mail address of the right person in the organization.

I encourage you to experiment and modify the techniques for diagnosing and auditing your DNS systems—it will pay dividends time and time again—it's best that you get there before the bad guys. And, it can provide endless hours of fun as you sleuth around. ■

Ron Aitchison is the author of *Pro DNS and BIND* and loves nothing better than using dig to uncover bizarre DNS configurations. One day, real soon now, he is going to get a real life.

Resources

DNS Statistics:
dns.measurement-factory.com

BIND: www.isc.org

BIND Configuration:
www.zytrax.com/books/dns

fpdns: www.rfc.se/fpdns



We use our stuff.

visit our TS-7200 powered website at

www.embeddedARM.com

(480) 837-5200

SXSW INTERACTIVE

MARCH 7-11, 2008 • AUSTIN TX

SXSW INTERACTIVE FESTIVAL: CONNECT, DISCOVER, INSPIRE

Attracting digital creatives and new media entrepreneurs, the 15th annual South by Southwest (SXSW) Interactive Festival gives you both practical how-to information as well as unparalleled career inspiration. Attend this legendary gathering of the tribes to renew your link to the cutting edge.

Opening Remarks by Henry Jenkins on Saturday, March 8

The Co-Director of the Comparative Media Studies Program at MIT, Jenkins has also authored numerous books including "Convergence Culture: Where Old and New Media Collide" and "Fans, Bloggers and Gamers: Exploring Participatory Culture."

Confirmed speakers include **Lea Alcantara** (Lealea Design), **Natalie Zee Drieu** (Craft Magazine), **Tim Ferriss** (Author), **Jason Fried** (37signals), **Kelly Goto** (Goto Media), **Brewster Kahle** (Internet Archive), **Nathan Shedroff** (nathan.com), **Tatsuki Taomita** (Opera), **Susan Wu** (Charles River Ventures) and **Annalee Newitz** (Techsploitation).

REGISTER TO ATTEND SXSW 2008

Go to **sxsw.com** now to take advantage of early registration discounts and for up-to-date lists of panels, panelists, speakers and Web Awards finalists.

SOUTH BY SOUTHWEST INTERACTIVE FESTIVAL
March 7-11, 2008 | Austin, Texas | www.sxsw.com



Linux Powers *The Spiderwick Chronicles*

Linux software on Macintosh desktops with Linux renderfarms creates Paramount movie.

ROBIN ROWE

A Linux-based production pipeline is a perfect choice for a major motion picture like *The Spiderwick Chronicles*, with its many goblins and magical creatures. Hollywood has been the realm of Linux since 1997, when the movie *Titanic* proved that Linux can do big computer graphics jobs like rendering a sinking ocean liner. With an industry tradition of using UNIX-based operating systems for high-computation jobs, and due to the better, faster, cheaper nature of Linux, every major effects or animation movie today is produced using Linux. Visual effects facilities ILM and Tippett Studio each created visual effects for *Spiderwick*. Having multiple effects houses work on the same movie became common after 2003 when *The Matrix Reloaded* used a dozen effects houses.

Tippett Studio: Linux on Macintosh Desktops

The idea of switching Mac desktops to Linux is new in the film industry. The film industry routinely uses Macs running OS X for specialized tasks, such as art department concept artwork generated using Adobe Photoshop, picture editing with Final Cut Pro and sound editing with ProTools. When you scale past a few systems, the advantages of Linux for graphics become apparent, and Linux graphics PC desktops are the norm. The television series *South Park* is a notable exception, with Mac OS X desktops running Maya with a Linux renderfarm.

During the production of *Spiderwick*, Tippett Studio switched to Fedora Linux running on Macintosh desktops. "We currently have 119 Intel-based Apple Mac Pro workstations running Linux", says Tippett Computer Graphics Supervisor Russell Darling. "We decided to go with Apple hardware running Linux for our primary artist workstations on *The Spiderwick Chronicles*, although it might have been considered a risky endeavor for a show in production. We initially had some problems with sound on Maya and a few other minor issues, but they were resolved. We got a patch from Autodesk that took care of everything." Commercial Linux software vendors work closely with film studio clients.

Tippett chose Linux on Mac for many reasons. "There's the ability to run multiple operating systems, including Linux, OS X and Windows", says Darling, and he continues, "The systems are fast! That makes for more productive artists. The hardware is quiet and energy-efficient. It's cost-effective, with a good cost per rendermark [a renderfarm performance benchmark]. It's standardized hardware. And, there's a good support plan. Although the majority of our workstations run Linux, we have a handful of other systems running to support specific software. We use the ability to boot in to other operating systems,

but the ultimate goal is to move to a simultaneous multi-OS solution, such as Parallels."

To beat traditional alternatives, the Apple Mac Pro workstations had to meet a specific set of Tippett requirements. They had to run Fedora FC4 and XFS. They also had to run tools that Tippett uses, such as Maya with sound and in-house and third-party plugins (MEL scripts), Apple Shake with in-house and third-party plugins, SyFlex, cMuscle, RealFlow, JET, Flipper, rtTools and cineSpace. Internally developed software uses Python, Perl and C/C++. The platform must render frames identical to existing hardware. And, it has to support necessary peripherals, especially tablets.

ILM: Mulgarath, Thimbletack, the Griffin, the Sprites and Stray Sod

"The important thing with a fantasy genre is referencing nature", says ILM Art Director Christian Alzmann. "The Byron plumage is based on a red-tailed hawk. We're always drawing reference from nature. I did the early design of the Sprites with a bee next to them for scale, with two bees flying in formation. Mulgarath is part man, part bull, part goat, part trees. The warhog is a mean aggressive character, so we got pointy with him. And, he's a lot more distorted. We also use scale cues, such as a Chiquita banana sticker or Pepsi bottle cap."

"The Griffin has hair plus feathers and was rendered at 8k [images 8k pixels wide] to get detail", says ILM Animation Supervisor Tim Harrington. To achieve that level of detail meant 25- to 30-hour renders.

"*Spiderwick* took 215 artists and 15 months", says ILM Visual Effects Supervisor Tim Alexander. "It has 341 shots, 30 minutes, with 224 3-D shots."

Industrial Light & Magic occupies the 865,000 square-foot Letterman Digital Arts Center on the 23-acre San Francisco Presidio campus. Its data network has more than 300 10GB ports and 1,500 1GB ports, with fibre to every artist's desktop. There are 600 miles of cable throughout the four buildings on the campus. A 13,500 square-foot data center houses a Linux renderfarm with 3,000 AMD processors and more than 100TB of storage. Proprietary render management tools add Linux desktop workstations to the renderfarm pool after hours, expanding the processing capacity to more than 5,000 processors.

Tippett Studio: Hogsqueal, the Troll, Red Cap and His Army of Goblins and Bull Goblins

As Creature Supervisor for *The Spiderwick Chronicles*, visual effects pioneer Phil Tippett oversaw the design and develop-



The idea of switching
Mac desktops to Linux is
new in the film industry.

ment of the film's fantasy characters. "Phil Tippett was on set with me every day", says Director Mark Waters. "We were working on *Charlotte's Web* when Mark Canton gave us the script", says Tippett Studio Visual Effects Supervisor Joel Friesch. "When we saw the creatures, we had to do it. It's based on real creatures, not fantasy. We wanted Hogsqueal. We created a bull goblin marquette [a detailed statuette] that gave Mark something he could hold. The bull goblin is based on toads. We brought in real toads and photographed them. We created movies good for the animators, showing how the eyes move and the throat. We created a test scene with a goblin scratching the back of his leg. That took one month of modeling and one month of animation."

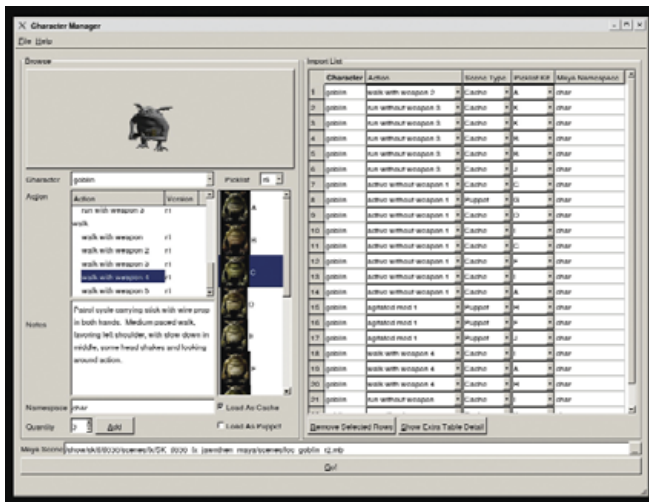


Figure 1. Tippett Studio's proprietary Creature Manager is used to maintain a library of creatures and animation cycles. The tool allows an artist to select and preview animation by pressing the larger creature button, then selecting a combination of an appropriate physical appearance for that creature from a predefined library and placing any number of selected creatures into a Maya scene.

Hand animation is a challenging laborious process. "One guy does blocking, like moving chess pieces", says Tippett Studio Animation Supervisor Todd Labonte. "You get it approved. We watch it over and over. You can go blind. We play it back in mirror image in our player or play it backward." Labonte demonstrates playing back a scene of goblins invading the house, shown in their Flipper playback software, which can display a mirror image or play in reverse to help catch animation inconsistencies. Flipper is used to view both QuickTime and image frame sequences of DPX, EXR or TIFF with synchronized AIF audio. Flipper predates commercial Linux flipbooks, such as FrameCycler. At older studios, like Tippett, it's common to find proprietary Linux tools created before commercial options were available. Tippett has a team of eight Linux programmers to maintain and develop tools.

"Creature Picklist is a GUI-based Maya plugin for creatures that allows animators to see visual representations of character, which they can select for their scene", says Darling.

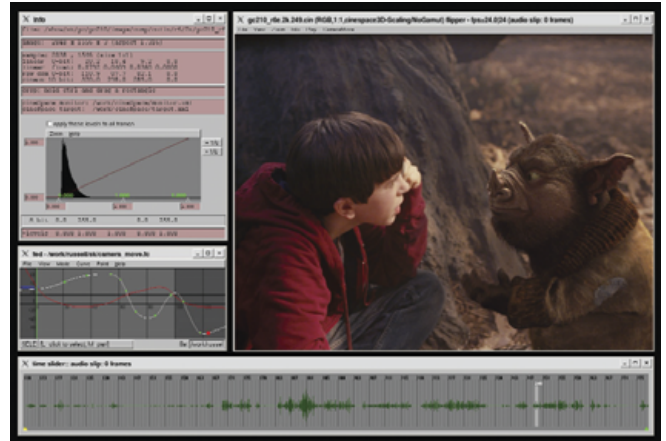


Figure 2. Flipper is Tippett Studio's proprietary flipbook image viewing tool. It allows an artist to view a series of individual image files as a continuous sequence. It also can be synchronized with audio, which is important for character animation. The artist can view the audio waveform to help with lip synchronization, as seen in the lower part of the screenshot. The tool also has a number of image and pixel comparison and analysis features, as seen in the dialog on the upper left. Post-camera moves can be previewed with Flipper before they are actually applied in the composite stage.

"In the case of *Spiderwick*, 'Goblin kits' were created as combinations of variants and blendshapes. We have shots that have more than 100 goblins. That's too many to animate using traditional methods. The numbers are also too small to make a commercial crowd system, such as Massive, a viable solution. We developed our own system called Swarm. For the *Spiderwick* shots, we instanced around 150 goblins and managed animation clip data to animate them as particles."

Furocious is Tippett Studio's proprietary hair, fur and feather system. It's a collection of plugins, scripts and executables used to place guide geometry onto scalp surfaces, visualize fur

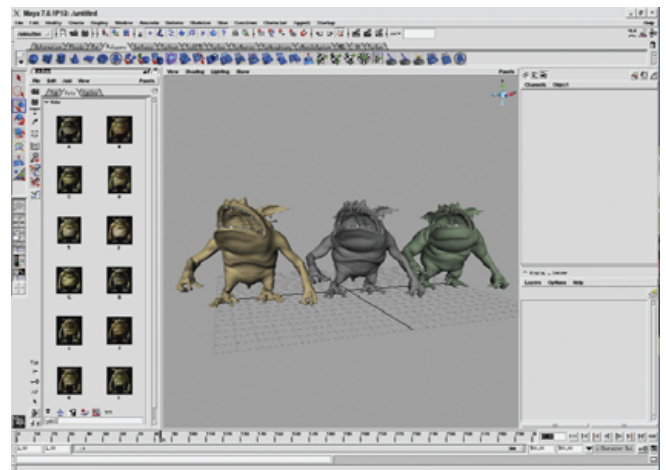


Figure 3. Tippett Studio's Picklist allows an animator to select creature variants from a library of different combinations of paint schemes and body parts.

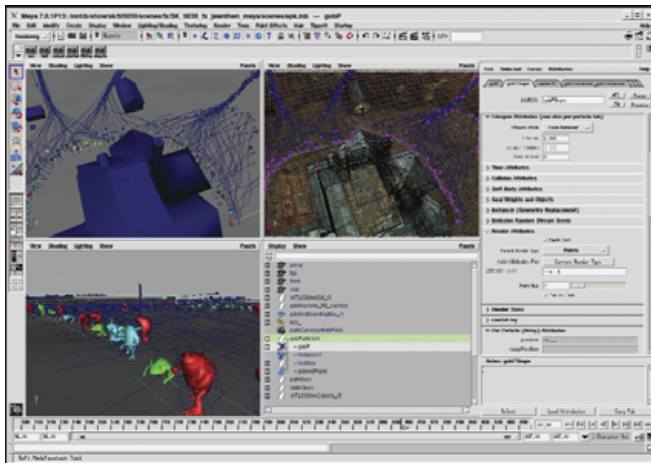


Figure 4. A view of Tippett's Swarm crowd system in Maya. The scene is choreographed by defining paths and actions for creatures to follow.

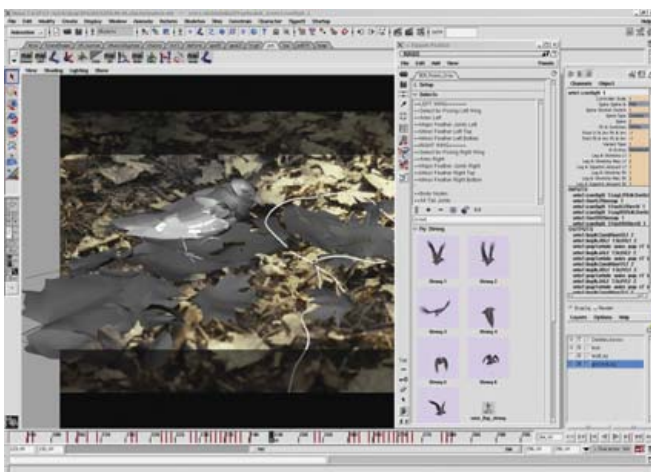


Figure 5. Tippett Studio's Picklist allows an animator to select specific predefined poses for creatures, such as this bird.

as GL curves in Maya, and grow procedural primitives at render/expansion time by interpolating neighboring guides at predetermined follicle root locations.

RaveHD: Embedded Linux for Cinema Playback

"We currently have six RaveHD systems at the studio", says Darling. "They're used to provide dailies in our screening rooms. We also use them for smaller reviews and on-demand playback for artists in special viewing rooms. And whenever we shoot with HD cameras on our stage, we use RaveHD to acquire the images and bring them on-line as individual digital frame files. We have been working with SpectSoft since 2002, before the RaveHD product really existed. RaveHD is an awesome system. It has a really nice design that allows it to integrate perfectly in our dailies pipeline—very reliable and easy to use". The RaveHD box is a Linux embedded system that plays cinema-grade motion pictures at data rates that would choke a PC.

"I originally met [SpecSoft partners] Jason and Ramona Howard at a Linux Movies meeting in Berkeley, shortly after I joined Tippett Studio", says Darling. "At that time, we were looking at developing a new dailies system. They had developed a Linux driver for the AJA Kona HD/SD card and had developed some DDR and editing tools. We were able to form a great relationship with SpectSoft where we provided specifications and requirements to them in order to help create a system that suited our purposes." LinuxMovies.org is an association of Linux motion-picture technologists founded in 2002 [by Robin Rowe, author of this article].

Tippett Studio: Linux Renderfarm

Tippett uses more than 1,200 processors in its renderfarm. "Tippett Studio has its own shading library built around the RenderMan Shading Language. Our pipeline tools are also centered around the RIB interface. The most important consideration for renderfarm configuration is that all jobs submitted the night before must be finished by the next morning", says Darling. "Each morning we do dailies to review the previous day's work. If the job is not finished, it can't be properly reviewed. There are occasional exceptions for special shots that may run long, but for the most part, we want everything to finish overnight." The Tippett renderfarm is managed by a proprietary batch-scheduling software, so that each computer in the farm is working on only one frame at a time. "Our distributed rendering system Batch-o-matic has been in use at the studio for ten years", says Darling.

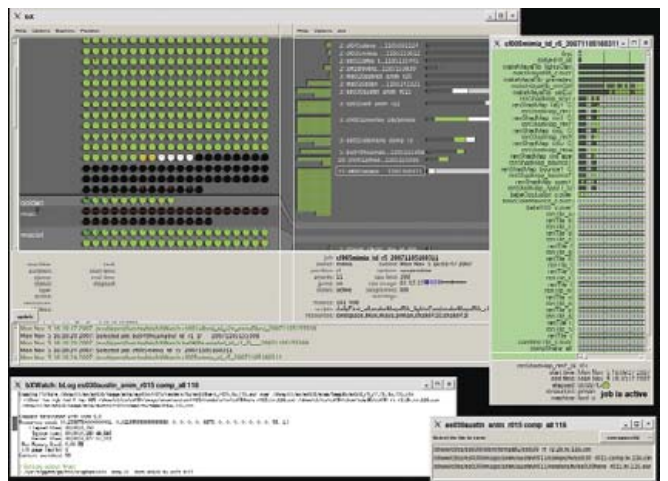


Figure 6. Tippett Studio's renderfarm is managed by its proprietary Batch-o-matic system. In its GUI interface, CPUs are represented as circles with color-coded indications of their activity level and status. Information about an artist's render jobs is represented by color-coded squares indicating number of CPUs in use, as well as an overall status bar. More detailed information is shown for a specific job, with status for each task and frame.

Tippett Studio: Linux Python Pipeline

JET is a proprietary Python-based system comprising software tools and scripts used to implement a visual effects and animation pipeline. "A visual effects and animation pipeline is an

There are commercial Linux software tools available to Linux animators beyond big packages like Maya.

assembly line of software used to organize, automate and facilitate the creation of computer-generated imagery”, says Darling. “The JET tool is highly customizable, featuring XML-based user-interface templates that can be modified to suit specific types of artists or production needs. JET uses modular template chunks to perform each of the tasks in the pipeline, such as rendering or compositing. The templates are implemented as Python objects and are centrally located. JET is not only implemented entirely in Python, but it’s also used to generate Python scripts automatically. These custom scripts form unique pipelines for each computer graphics job to run on the renderfarm.”

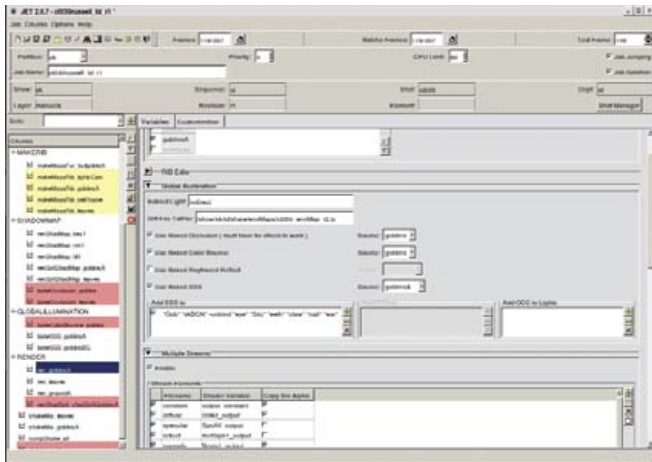


Figure 7. JET is Tippett Studio’s proprietary tool for creating jobs that are to be submitted to the renderfarm. Each job comprises a series of chunks (as shown in the panel on the left) that perform specific tasks, such as making a RIB, growing fur, rendering with RenderMan and compositing with Shake. The primary work area on the right provides an interface for the selected chunk.

Tippett Studio: Compositing with Apple Shake and Painting with Photoshop

Tippett Studio uses Apple’s Shake compositing software, which has been discontinued. “We took advantage of Apple’s offer of selling the Shake source code to us”, says Darling. “We don’t plan on modifying it, but it is good to have for an insurance policy. Shake is still very prominent in the visual effects industry. It’s useful to be able to share Shake projects with other studios we’re collaborating with. Shake has a really nice image-processing engine. In addition to standard compositing, we use it for all kinds of image-processing solutions in our pipeline. The product is very mature and feature-rich. It has support for plugins, which allows us to develop and enhance compositing nodes whenever we need something new. We don’t have any current plans to switch compositing packages, but we’re always keeping an eye on what’s out there.”

“We use Photoshop CS as the painting interface for our texture painting, as well as for matte paintings”, says Darling. “It’s used directly on Windows and Mac systems. Our goal is to use a system such as Parallels to permit artists to run tools outside their primary operating system. Deep Paint 3D is used to paint textures for 3-D models. The actual painting is done with Photoshop, but the interaction with the model is done in Deep Paint.” Deep Paint and Photoshop are both non-Linux commercial tools. Some studios, such as DreamWorks Animation, use Wine to run Photoshop on Linux. Another option is to run an open-source Linux paint package that supports industry-standard high-fidelity image formats DPX and OpenEXR (such as CinePaint). However, Deep Paint on Wine is untried, and there’s no open-source option.

“When painting in Photoshop, it is essential to be able to view the image with the same type of color management used to produce the final rendered image”, says Darling. “Photoshop supports ICC profiles for color management. Since we use cineSpace in the rest of the pipeline, we had a tool created that would allow us to produce an ICC profile that matched the cineSpace profile. We collaborated with Joseph Goldstone, who is a member of the International Color Consortium (ICC) to create this tool.”

Plentiful Commercial Linux Animation Tools

There are commercial Linux software tools available to Linux animators beyond big packages like Maya. “cMuscle is a system for simulating muscle movement”, says Darling. “We used it for muscle jiggle, skin jiggle, skin smoothing and sliding effects. We use SyFlex for the clothing worn by Red Cap and Hogsqueal. This system was augmented by some custom cloth/deformation software developed at Tippett Studio.”

“cineSpace allows us to have a specific, consistent look on all viewing devices that simulate our finaling medium—the look of the film that the director is using for final shots”, says

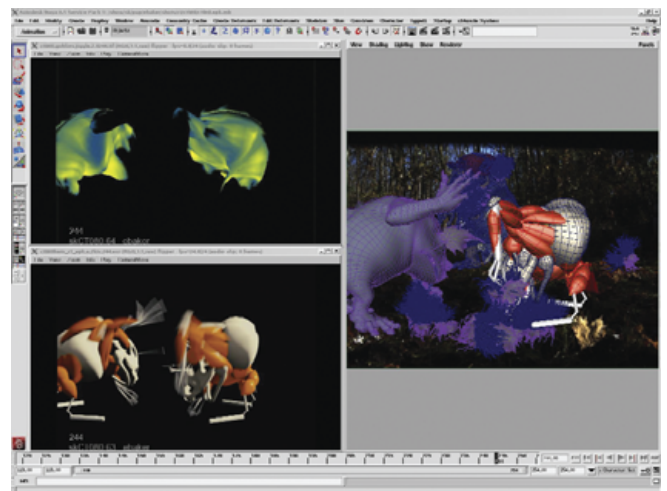


Figure 8. Creature skin, muscles and skeletal structures—the green and yellow images represent soft body jiggle maps. Along with the muscle system, we also have what is called tension controls. These are blendshapes of various body parts that when fired give those areas a more flexed look, giving the illusion of the skin tightening.



Enterprise and High-Performance Computing Under Your Control

Industry leading 4P x86 computing to boost performance and power efficiency with **Dual or Quad-Core** AMD Opteron™ Processors

4-Way XtremeWorkstation™

- AMD Opteron™ 8000 Series processors
- Up to 128GB of DDR2 533/667 memory
- Up to 6.0TB SATA or 2.4TB SAS
- Hot-swappable drives
- 2 PCI-E x16 slots for high-end graphics card
- Windows® or Linux OS



4-Way 3U XtremeServer™

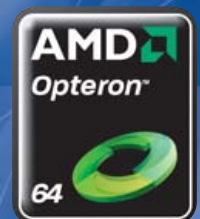
- AMD Opteron™ 8000 Series processors
- Up to 128GB of DDR2 533/667 memory
- Up to 4.5TB SATA or 1.8TB SAS
- 2 PCI-E x16 and 3 PCI-X slots
- Hot-swappable drives
- Redundant power supplies & fans
- ServerDome Management – IPMI 2.0
- Windows® or Linux OS



For more information, please visit www.appro.com
or call Appro Sales at 800-927-5464

AMD Opteron™
Processors:

- Quad-Core Ready - increase capacity without altering datacenter infrastructure
- Best performance per-watt with energy-efficient DDR2
- Optimized system performance with Direct Connect Architecture



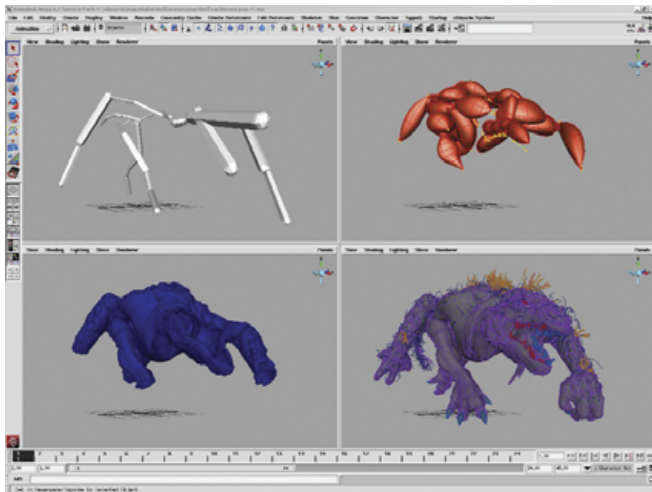


Figure 9. Creature muscles (red) and basic skeletal system (white). For this creature, Tippett Studio's hair and fur system Furocious was used to grow fur that had the appearance and behavior of grass, which is one of the unique camouflaging features of this character.

Darling. "Since film is still the master format, this means a combination of a specific film stock and a specific lab that is used to process and print that film. A cineSpace profile allows us to produce an image digitally that has the same look and characteristics of that same image if it were printed to film. It is

Tippett Desktop and Renderfarm Stats

Desktop hardware:

- Apple Mac Pro
- Quad Xeon5150 2.66GHz
- GeForce 7300 GT

Renderfarm hardware:

- Quad Xeon5160 3.0GHz
- Quad Opteron290 2.80GHz
- Quad Opteron265 1.80GHz
- Dual Athlon2800+ 2.15GHz
- Dual Athlon2400+ 2.00GHz
- Dual Athlon1900+ 1.60GHz

Operating system: Fedora Core 4

Kernel: 2.6.20

Compiler: GCC 4.0.2

Desktop: KDE 3.5.1

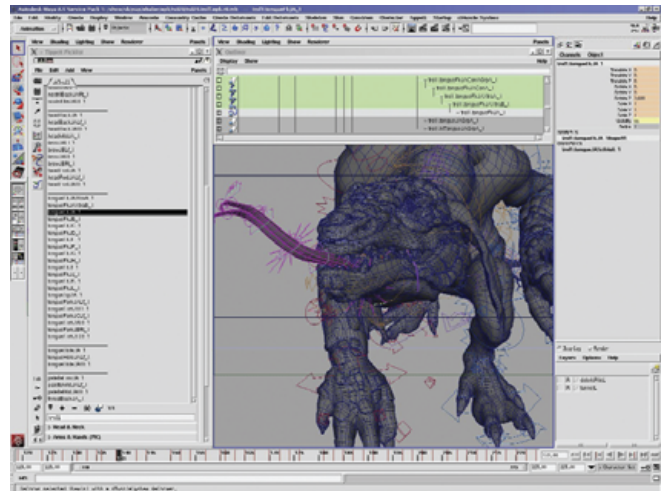


Figure 10. Animation controls for the Troll.

important that we are looking at the same thing that the director sees, as well as what will eventually be seen in the theater."

"Tippett Studio's primary operating system is Linux", says Darling. "Open-source software is very important to us. In addition to it being a part of our pipeline, we've contributed to a few projects over the years, such as Pixie. Open source is also something that can enable standards. An example of this would be the OpenEXR Project. We really needed a standardized high-dynamic-range image format that was supported in commercial and open-source CG software. The EXR image file format has now become a de facto standard image file format in our industry, simply because it was open-sourced."

The Creatures of *Spiderwick*

"In *Spiderwick* all unexplained phenomena in the world are due to fantastical creatures that are all around us", says director Mark Waters. The movie is based on the best-selling series of children's books written by Holly Black and illustrated by Tony DiTerlizzi. *The Spiderwick Chronicles* is a family adventure fantasy set at the dilapidated Spiderwick Estate, constructed in Canada at Quebec's Cap-Ste-Jacques Nature Park.

The movie opens to wide release on February 15, 2008. There's also a *Spiderwick* video game in development. ■

Robin Rowe is a partner in MovieEditor.com and a former DreamWorks Animation technologist. He's speaking at FOSDEM (fosdem.org) in Brussels, February 23, 2008. Robin is also the project manager for the open-source motion picture painting tool CinePaint.org.

Resources

cMuscle: www.cometdigital.com/cMuscleSystem_notice.php

SyFlex: syflex.biz

cineSpace: cinespace.risingsunresearch.com

METRO TORONTO CONVENTION CENTRE, CANADA

CONFERENCE: 7-9 APRIL 2008

EXPO: 8-9 APRIL

CULTIVATING
INNOVATION IN
THE IT INDUSTRY

it@360°

STAY AHEAD OF THE CURVE. GET ON BOARD!

www.it360.ca

PRESENTING:



ASTERISK
OPEN TELEPHONY
CONFERENCE



SPRING
CONFERENCE
2008

MEDIA
PARTNER:



IN
COOPERATION
WITH:

CIO Executive Council
The Professional Organization for CIOs

PREMIER MEDIA
SPONSORS:



MEDIA SPONSOR:



PLATINUM
ASSOCIATION
SPONSOR:



CONFERENCE
SPONSORS:



CERTIFICATION
SPONSOR:



ASTERISK &
OPEN TELEPHONY
SPONSOR:



Virtualization 2.0: Where the Sidewalk Ends

With VMware launching an IPO with no peak in sight and XenSource a red-hot acquisition by Citrix, it's clear that server virtualization is the flavor of the moment. But, additional challenges are faced by all hypervisor vendors and users: performance and connectivity to local data networks and to storage. The answer? Virtualization 2.0—infrastructure virtualization. **KEVIN EPSTEIN**

Having been indoctrinated in the ways of virtual machines (VMs) by VMware, where I spent the last four years, I feel like I'm committing heresy when I say that VMs aren't the sole answer to data-center ills. Beyond that, VMs actually can create some additional challenges, such as things to trip over when that moving sidewalk ends.

Don't get me wrong. I like VMware and Xen; I'm learning to believe in Solaris LDOMs on SPARC, and I'm a huge believer in the power of virtual machines. I think that for development, it's hard to beat the utility of having full multi-tier systems, virtually networked together, inside a single physical machine. And, it's certainly convenient to have packaged VMs to trade to other folks for easy replication of scenarios.

But, there are additional challenges faced by all hypervisor vendors (and users), performance and connectivity to local data networks and to storage.

An immediate aside is probably needed here. No, VMware VI3 doesn't even touch, much less solve, these issues—something that the world is beginning to recognize. Consider the following comments made by some well-known names in the industry:

- "Be forewarned—as soon as companies deploy wider virtualization, a completely new class of problems will arise. Here the market is still failing to offer qualified solutions, solutions that largely involve more sophisticated automation."—Alessandro Perilli, Virtualization.Info (world's leading virtualization blog)
- "The ability to do server repurposing is critical for customers who want to implement a real-time infrastructure."—Donna Scott, VP and Distinguished Analyst, Gartner Group
- "Servers virtualized? Great! Now you need to virtualize your entire data center. Virtualization isn't happening only within servers. Infrastructure virtualization applies a software abstraction layer across the entire data center."—Rachel Chalmers, The451Group

In short, it's not enough to have virtualized servers. You need virtual connectivity and real servers too—aka infrastructure virtualization.

But, let's consider the two challenges again, performance

and connectivity to local data networks and storage. The performance challenge is straightforward, and therefore, it's not worth much discussion.

Hypervisors take up CPU cycles, are a potential point of failure (if the hypervisor fails, the whole set of VMs are lost), and due to the round-robin nature of their scheduling, they actually can mask race conditions in tests. Honest hypervisor vendors always will admit this point quite openly—in fact, I'm proud that during my tenure, VMware quite openly stated "don't virtualize applications above a certain threshold CPU, network or disk I/O level".

The point? To test and run many multi-tier production applications, such as Citrix, SQL Server, SAP and so on, you want to run on bare metal. This, unfortunately, raises the challenge of movement between physical and virtual (P2V & V2P), as well as raising the second challenge, networking.

Networking is a more subtle challenge, because it's out-of-box—outside the physical machine in which a hypervisor (okay, let's be honest, it's an operating system) lives.

Again, it's time for an aside: hypervisors are an operating system. A hypervisor, like VMware ESX Server or Citrix XenSource Enterprise Server, is installed on a bare-metal server-class computer. The computer is cabled in to a LAN (data) network switch and a SAN (storage) switch. The computer is turned on, boots up with the hypervisor and then can run multiple full servers, in virtual machines, each with its own operating system and applications (for example, Windows/Exchange e-mail server and a Linux/Apache Web server) concurrently on top of that hypervisor operating system.

Hypervisors, therefore, control events happening in-box, within a computer. The moment data leaves the physical machine, bound out to another physical box via NIC to LAN or to storage via HBA to SAN (or NIC to NAS), it has left the hypervisor's control.

What does this mean or matter? Well, envision your nice multi-tier application—Web servers, app servers, databases—built in VMs and "cabled" inside single physical machine A. Works great.

Now, take one of those VMs, and put it on a different physical machine B (running a hypervisor) that's somewhere else.

Unless that physical machine is on the same LAN subnet, with the same SAN access, you've just broken your data

Hear Yourself Think Again!



WhisperStation™ **Cool... Fast... Silent!**

For 64-bit HPC, Gaming and Graphic Design Applications

Originally designed for a group of power hungry, demanding engineers in the automotive industry, WhisperStation™ incorporates two dual core AMD Opteron™ or Intel® EM64T™ processors, ultra-quiet fans and power supplies, plus internal sound-proofing that produce a powerful, but silent, computational platform. The WhisperStation™ comes standard with 2 GB high speed memory, an NVIDIA e-GeForce or Quadro PCI Express graphics adapter, and 20" LCD display. It can be configured to your exact hardware specification with any Linux distribution. RAID is also available. WhisperStation™ will also make a system administrator very happy, when used as a master node for a Microway cluster! Visit www.microway.com for more technical information.

Experience the "Sound of Silence".

Call our technical sales team at 508-746-7341 and design your personalized WhisperStation™ today.



Microway
Technology you can count on™

center. The server (VM) you just moved is reaching out to LAN and SAN paths that don't physically exist anymore, because machine B doesn't have them.

Not to mention that little part about machine B already having to be running a hypervisor. (And, if you can't get to machine B to turn it on and install that hypervisor, such as in a DR case, then what? Don't say, "oh gosh, that will all be fixed by virtual HBAs and NPIV", either. NPIV requires a full data-center rollout, supported across all hardware—got legacy switches, anyone?)

Virtual HBAs make a bad problem worse. Now you have even more bound HBAs. (Heaven forbid that you're allowed to change virtual HBAs like virtual MACs, that'd create an even worse problem, as without locking, people would duplicate HBAs and kill off entire SANs by accident.)

But, what's the solution? Ideally, you'd be able to do a few things:

- Move easily between physical and virtual. Enable four-minute P2V/V2P conversions to increase hypervisor proliferation and adoption. Seamlessly move servers from running on physical machines to running on virtual machines—and back. This functionality provides a safety net for applications not suited for hypervisors, allows unlike hardware to be used for disaster recovery and high availability, and removes potential application vendor-support issues for hypervisors by providing a vehicle to test applications quickly on bare-metal hardware.
- Move what was running on your physical machines around as easily as what was running on your virtual machines. So, you could remote-start machine B, above. (For bonus points, remote-start hypervisors like ESX Server or Xen too.) This then provides high availability on a per-physical-machine level. Note that the VMware VirtualCenter HA option (and the Xen equivalent) depends on the availability of another physical machine running ESX server with the right physical network and storage connectivity. Scalent creates such a machine in real time.
- Take the network topology, LAN and SAN, with you. So, no more SAN configuration adjustments or opening all SAN LUNs to all machines (good grief). Ideally, you'd be able to install hypervisors that could communicate within your existing multi-tier networks and have real-time access to storage to enable features like VMotion. (Again, VMware ESX requires that physical machines share common pools of storage and the same network subnet if they are to share virtual machines in a VMware ESX farm or cluster.) This can present logistical or security concerns when applied to current data-center architectures. A good solution would allow VMware ESX physical machines to be in physically disparate locations—behind layers of LAN switches, in different racks or data centers, connected to storage LUNs as needed—thus allowing a simple ESX farm or cluster creation in existing data-center architectures.

This solution would manifest as infrastructure virtualization (IV) software—effectively, a data-center topology manager. IV software turns on physical computers, sets up the networks between them, sets up their connections to storage and points them at the right software (operating system/application package) to run.

For example, extending the above hypervisor example, at any time, IV software could turn on any three physical computers, with no software running on them and set computer one to network A/storage LUN 1, running the Windows/Exchange e-mail server package; set computer two to network B/storage LUN 2, running the Linux/Apache Web server; and set computer three to network C/storage LUNs 3 and 4, running a hypervisor, with the hypervisor load of virtual machines.

How would this look in real life? Let's consider a disaster recovery scenario.

Visualize a data center. There are many physical computers running various operating systems (Windows, Linux, Solaris, AIX, VMware ESX, Xen and so on) on various hardware (x86, SPARC, PowerPC), connected to various data networks and various storage LUNs.

Suddenly, a rack of computers in the corner goes up in flames.

Luckily, the customers are running IV software, as well as their existing automation, virtualization and bare-metal operating systems. That infrastructure virtualization or server repurposing software immediately realizes that the physical computers are down, hunts around the data center and finds some computers that are off or some that are running low-priority jobs, like print servers, that the IT staff has designated "repurposable".

The IV software turns on or reboots those physical machines, assigns them the networks and storage connections of the burned machines and tells the computers to load the software to which the IV points them.

Unfortunately, there are only ten new working computers—20 went up in smoke. So, the IV software loads two of the new computers with a hypervisor and runs six of the burned systems as virtual machines on top of the physical computers running the hypervisor. This is possible because the IV software set up all the necessary network and storage connectivity for all those servers and associated them with the physical machine running the hypervisor in real time.

Pipe dream, you say? Not at all. The scenario just described is alive and well, running on Linux and controlling multiple other operating systems, in some of the world's largest data centers today. And, disaster recovery is only one of the benefits; others include performance, real-world test harness automation and eased hypervisor deployment.

So, enjoy server virtualization. Go deploy hypervisors; they're a good thing. But, don't be caught unaware when the sidewalk ends. There's more to virtualization than servers. Your infrastructure is waiting. ■

Kevin Epstein is the VP Marketing & Products for Scalent Systems, makers of Infrastructure Virtualization software. Prior to Scalent, he served as a Director for VMware, Inc., from 2002 until 2006, and previously for Inktomi Corporation's Network Products division, RealNetworks, Netscape and others. Kevin holds a BS degree in High Energy Physics from Brown University and an MBA from Stanford University, and is the author of *Marketing Made Easy*, a popular trade book from Entrepreneur Magazine Press/McGraw Hill.

SD WEST 2008

Software Development
Conference & Expo West

March 3-7, 2008

Santa Clara Convention Center
Santa Clara, CA

**SUPER EARLY
BIRD DISCOUNT**

Register by JAN 11

**SAVE UP
TO \$500**

**EARLY BIRD
DISCOUNT**

Register by FEB 8

**SAVE UP
TO \$300**

JOIN THE REVOLUTION

FEATURING
OVER
200
EDUCATIONAL
SESSIONS

PLUS:

Keynotes, Expo, Birds-
of-a-Feathers, Panels,
Case Studies, Parties
and other Special
Events!

EDUCATIONAL SESSIONS INCLUDE:

LANGUAGES & PLATFORMS:

- C++
- JAVA PROGRAMMING
- .NET DEVELOPMENT
- RUBY

NEXT-GEN WEB:

- WEB 2.0/WEB DEV
- WEB SERVICES/SOA

SOFTWARE DEVELOPMENT LIFECYCLE:

- MODELING & DESIGN
- PEOPLE, PROCESS &
METHODS
- REQUIREMENTS &
ANALYSIS
- SECURITY
- TESTING & QUALITY

SOFTWARE BUSINESS & CAREER:

- BUSINESS OF SOFTWARE
- **NEW** IT AND THE
DEVELOPER
- **NEW** EMERGING
TECHNOLOGY

REGISTER TODAY AT WWW.SDEXPO.COM



CMP

United Business Media

System Minimization

Strategies for reducing Linux's footprint, leaving more resources for the application or letting engineers further reduce the hardware cost of the device. **GENE SALLY**

"How small can you make this?" is a question frequently heard by embedded engineers at the start of their projects. Most of the time, the person asking this question is concerned with reducing the RAM and Flash resources with the goal of reducing a device's unit costs or energy requirements.

Because Linux, and the surrounding environment, originally was intended for desktop or server systems, its default configuration isn't optimized for size. However, as Linux is finding itself in more embedded devices, making Linux "small" isn't as daunting a task as it once was. There are several different approaches for reducing the memory footprint of a system.

Many engineers start by reducing the size of the kernel; however, there is lower-hanging fruit at hand. This article goes into detail about how to reduce the size of the kernel, mostly by removing code that won't even be used in a typical embedded system.

A root filesystem (RFS) can be the largest consumer of memory resources in a system. A root filesystem contains the infrastructure code used by an application as well as the C library. Selecting the filesystem used for the RFS itself can have a large effect on the final size. The standard, ext3, is frightfully inefficient on several axes from an embedded engineer's perspective, but that's a topic for another article.

Realistically, How Small?

Even the smallest Linux distribution has at least two parts: a kernel and root filesystem. Sometimes, these components are colocated in the same file, but they're still separate and distinct components. By removing nearly all features from the kernel (networking, error logging and support for most devices) and making the root filesystem just the application, the size of a system easily can be less than 1MB. However, many users choose Linux for the networking and device support, so this isn't a realistic scenario.

Kernel

The Linux kernel is interesting in that although it depends on GCC during compilation time, it has no dependencies at runtime. Those engineers new to Linux confuse the initial RAM disk (so-called initrd) with a kernel runtime dependency. The initrd is mounted first by the kernel, and a program runs that interrogates the system in order to figure out what modules need to be loaded in order to support the devices, so that the "real" root filesystem can be mounted. In fact, the two-step mounting, the initrd followed by the real root filesystem, rarely finds its way into embedded systems as the gain in flexibility in a system that does change isn't worth the additional space or time. But, this topic falls under the rubric of the root filesystem and is discussed later in this article.

Most of the effort in reducing kernel size lies in removing what's not needed. Because the kernel is configured for desktop and server systems, it has many features enabled that wouldn't be used in an embedded system.

Loadable Module Support

Kernel loadable modules are re-locatable code that the kernel links into itself at runtime. The typical use cases for loadable modules are allowing drivers to be loaded into the kernel from user space (typically after some probing process) and allowing the upgrade of device drivers without taking down the system. For most embedded systems, once they're out in the field, changing the root filesystem is either impractical or impossible, so the system's designer links the modules directly into the kernel, removing the need for loadable modules. The space-saving in this area isn't limited to the kernel, however, as the programs managing loadable modules (such as insmod, rmmod and lsmod) and the shell script to load them aren't necessary.

Linux-tiny Patches

The Linux-tiny set of patches has been an on-again-off-again project that originally was spearheaded by Matt Mackall. The Consumer Electronics Linux Forum (CELF) has put effort into reviving the project, and the CELF Developer's Wiki has patches for the 2.6.22.5 kernel (at the time of this writing). In the meantime, many of the changes in the Linux-tiny Project have been included in the mainline kernel. Even if many of the original Linux-tiny patches have made it into the kernel, some substantial space-saving patches haven't, such as:

1. Fine-grain printk support: users can have control over what files can use printk. This allows engineers to reap the size benefits of excluding printk for the kernel at large while still having access to their favorite debugger in the places where it's needed most.
2. Change CRC from calculation to use table lookup: Ethernet packets require a CRC to validate the integrity of the packet. This implementation of the CRC algorithm uses table lookups instead of calculations, saving about 2K.
3. Network tweaking: several patches reduce the supported network protocols, buffer sizes and open sockets. Many embedded devices support only a few protocols and don't need to service thousands of connections.
4. No panic reporting: if the device has three status lights and a serial connection, the user won't be able to see, much less act on, panic information that appears on a

(nonexistent console). If the device has a kernel panic failure, the user simply will power-cycle the device.

5. Reduction of inlining: an inline is where the compiler, instead of generating a call to a function, treats it as a macro, putting a copy of the code in each place it is called. Although the inline directive is technically a hint, GCC will inline any function by default. By suppressing inline functions, the code runs slightly slower, as the compiler needs to generate code for a call and return; in exchange, however, the object file is smaller.

The Linux-tiny patches are distributed in a tar archive that can be applied with the quilt utility or applied individually.

Additional Kernel Tweaking Recommendations

Although the Linux-tiny Project covers a lot of ground, several additional configuration changes will result in substantial footprint reductions:

1. Remove ext2/3 support and use a different filesystem: the ext2/3 filesystem is large, a little more than 32K. Most engineers enable a Flash filesystem, but don't disable ext2/3, wasting memory in the process.

2. Remove support for sysctl: sysctl allows the user to tweak kernel parameters at runtime. In most embedded devices, the kernel configuration is known and won't change, making this feature a wasted 1K.
3. Reduce IPC options: most systems can do without SysV IPC features (grep your code for msgget, msgctl, msgsnd and msgrcv) and POSIX message queues (grep for mq_*[a-z]), and removing them scores another 18K.

Seeing the Effect of Your Changes

The size command reports the amount of code and data in an object file. This is different from the output of the ls command, which reports the number of bytes in the filesystem.

For example, a kernel compiled with an armv5l cross-compiler reports the following:

```
# armv5l-linux-size vmlinux  
  
text    data    bss     dec     hex    filename  
2080300 99904   99312   2279516 22c85c vmlinux
```

The text section is the code (discovering the historical

NEW THREATS. NEW SOLUTIONS. BLACK HAT STYLE.

The stakes are high for today's network defenders.

New security threats to governments and businesses emerge hourly, from sophisticated wireless attacks to remotely organized Bot armies.

Black Hat DC brings together the best minds in security to define tomorrow's information security landscape.

Featuring many new tracks, new training sessions and a special emphasis on wireless security and attack analysis, DC 2008 is a whole new Black Hat.



FEBRUARY 18-21, 2008 ★ WESTIN DC CITY CENTER

Diamond Sponsor

Microsoft

Gold Sponsors

IOActive
COMPREHENSIVE COMPUTER SECURITY SERVICES

NORMAN SAINT

Klocwork

l0rmin
SECURITY



reason the code is in the text section is an exercise left to the reader) emitted by the compiler. The data section contains the values for globals and other values used to initialize static symbols. The bss section contains static data that is zeroed out as part of initialization.

Although this data is revealing, it doesn't show what parts of the system are consuming memory. There isn't a way to query vmlinux for that information, but looking at the files linked together to create vmlinux is the next best thing. To get this information, use `find` to locate the built-in.o files in the kernel project and pass those results to `size`:

```
# find . -name "built-in.o" | xargs armv5l-linux-size
↳--totals | sort -n -k4
```

The output of this command is similar to the following:

text	data	bss	dec	hex	filename
189680	16224	33944	239848	3a8e8	./kernel/built-in.o
257872	10056	5636	273564	42c9c	./net/ipv4/built-in.o
369396	9184	34824	413404	64edc	./fs/built-in.o
452116	15820	11632	479568	75150	./net/built-in.o
484276	36744	14216	535236	82ac4	./drivers/built-in.o
3110478	180000	159241	3449719	34a377	(TOTALS)

This technique makes spotting code that occupies a large amount of space obvious, so engineers working on a project can remove those features first. When taking this approach, users shouldn't forget to do a clean make between builds, as dropping a feature from the kernel doesn't mean that the object file from the prior build will be deleted.

For those new to the Linux kernel, a common question is how to associate some built-in.o file with an option in the kernel configuration program. This can be done by looking at the Makefile and the Kconfig file in the directory. The Makefile will contain a line that looks like this:

```
obj-$(CONFIG_ATALK) += p8022.o psnap.o
```

which will result in the files on the right-hand side being built when the user sets the configuration variable `CONFIG_ATALK`. However, the kernel configuration tool typically doesn't readily expose the underlying configuration variable names. To find the link between the variable name and what's visible, look for the variable name, sans the `CONFIG_`, in the files (Kconfig) used to drive the kernel configuration editor:

```
find . -name Kconfig -exec fgrep -H -C3 "config ATALK" {} \;
```

which produces the following output:

```
./drivers/net/appletalk/Kconfig-#
./drivers/net/appletalk/Kconfig-# Appletalk driver configuration
./drivers/net/appletalk/Kconfig-#
./drivers/net/appletalk/Kconfig:config ATALK
./drivers/net/appletalk/Kconfig- tristate "Appletalk protocol support"
./drivers/net/appletalk/Kconfig- select LLC
```

```
./drivers/net/appletalk/Kconfig- ---help---
```

There's still some hunting to do, as the user needs to find where "Appletalk protocol support" appears in the configuration hierarchy, but at least there's a clear idea of what's being sought.

Root Filesystem

For many embedded engineers new to Linux, the notion of a root filesystem on an embedded device is a foreign concept. Embedded solutions before Linux worked by linking the application code directly into the kernel. Because Linux has a well-defined separation between the kernel and root filesystem, the work on minimizing the system doesn't end with making the kernel small. Before optimization, the size of the root filesystem dwarfs that of the kernel; however, in the Linux tradition, this part of the system has many knobs to turn to reduce the size of this component.

The first question to answer is "Do I need a root filesystem at all?" In short, yes. At the end of the kernel's startup process, it looks for a root filesystem, which it mounts and runs the first process (usually `init`; doing `ps aux | head -2` will tell you what it is on your system). In the absence of either the root filesystem or the initial program, the kernel panics and stops running.

The smallest root filesystem can be one file: the application for the device. In this case, the `init` kernel parameter points to a file and that is the first (and only) process in userland. So long as that process is running, the system will work just fine. However, if the program exits for any reason, the kernel will panic, stop running, and the device will require a reboot. For that reason alone, even the most space-constrained systems opt for an `init` program. For a very small overhead, `init` includes the code to respawn a process that dies, preventing a kernel panic in the event of an application crash.

Most Linux systems are more complex, including several executable files and frequently shared libraries containing code shared by applications running on the device. For these filesystems, several options exist to reduce the size of the RFS greatly.

Change the C Library

Combined with GCC, most users don't think of the C library as a separate entity. The C language contains only 32 keywords (give or take a few), so most of the bytes in a C program are those from the standard library. The canonical C library, `glibc`, has been designed for compatibility, internationalization and platform support rather than size. However, several alternatives exist that have been engineered from inception to be small:

- `uClibc`: this project started as an implementation of the C library for processors without a memory management unit (MMU-less). `uClibc` was created from the beginning to be small while supplying the same functionality of `glibc`, by dropping features like internationalization, wide character support and binary compatibility. Furthermore, `uClibc`'s configuration utility gives users great freedom in selecting what code goes into the library, allowing users to reduce the size further.

- **uClibc++:** for those using C++, this library is implemented under the same design principles. With support for most of the C++ standard library, engineers easily can deploy C++-based applications onboard with only a few megabytes.
- **Newlib:** Newlib grew out of Red Hat's foray into the embedded market. Newlib has a very complete implementation of the math library and therefore finds favor with users doing control or measurement applications.
- **dietlibc:** still the smallest of the bunch, dietlibc is the best kept secret among replacements for glibc. Extremely small, 70K small in fact, dietlibc manages to be small by dropping features, such as dynamically linked libraries. It has excellent support for ARM and MIPS.

Using an Alternate C Library

Both Newlib and dietlibc work by providing a wrapper script that invokes the compiler with the proper set of parameters to ignore the regular C libraries included with the compiler and instead use the ones specified. uClibc is a little different as it requires that the toolchain be built from source, supplying tools to do the job in the buildroot project.

Once you know how to invoke GCC so it uses the right compiler, the next step is updating the makefiles or build scripts for the project. In most cases, the build for the project resides in a makefile with a line that looks like this:

```
CC=CROSS_COMPILE-gcc
```

In this case, all the user needs to do is run `make` and override the `CC` variable from the command line:

```
make CC=dietc
```

This results in the makefile invoking `diet` for the C compiler. Although it's tempting, don't add parameters into this macro; instead, use the `CFLAGS` variable. For example:

```
make CC="gcc -Os"
```

should be:

```
make CC=gcc CFLAGS="-Os"
```

This is important, because some rules will invoke `CC` for things other than compilation, and the parameters will not make sense and result in an error.

Back to the Root Filesystem

After selecting the C library, all of the code in the root filesystem needs to be compiled with the new compiler, so that code can take advantage of the newer, smaller C library. At this point, it's worth evaluating whether static versus shared libraries are the right choice for the target. Shared libraries work best if the device will have arbitrary code running and if that code isn't known at the time of deployment; for example,

the device may expose an API and allow end users or field engineers to write modules. In this case, having the libraries on the device would afford the greatest flexibility for those implementing new features.

Shared libraries also would be a good choice if the system contained many separate programs instead of one or two programs. In this case, having one copy of the shared code would be smaller than the same code duplicated in several files.

Systems with a few programs merit closer consideration. When only a few programs are in use, the best thing to do is create a system each way and compare the resulting size. In most cases, the smaller system is the one with no shared libraries. As an added benefit, systems without shared libraries load and start running programs faster (as there's no linking step), so users benefit from an efficiency perspective as well.

Summary

Although there's no magic tool for making a system smaller, there is no shortage of tools to help make a system as small as possible. Furthermore, making Linux "small" is more than reducing the size of the kernel; the root filesystem needs to be examined critically and paired down, as this component usually consumes more space than the kernel. This article concentrated on the executable image size; reducing the memory requirements of the program once it is running constitutes a separate project. ■

Gene Sally has been working with all facets of embedded Linux for the last seven years and is co-host of LinuxLink Radio, the most popular embedded Linux podcast. Gene can be reached at gene.sally@timesys.com.

Resources

Linux-tiny Patches: www.selenic.com/linux-tiny. A series of small patches to the kernel to reduce the image size and runtime resources. Many of these patches already have made their way into the kernel.

GNU C Library: www.gnu.org/software/libc. The GNU C Standard Library is the canonical implementation of the C library. The need for this to run on nearly every platform with backward compatibility resulted in a Lib C that's bigger than most.

uClibc: www.uclibc.org. A well supported smaller implementation of Lib C.

Newlib: sourceware.org/newlib. Red Hat's small C library.

dietlibc: www.fefe.de/dietlibc. The smallest C library of the bunch. It works well with an existing cross-compiler, as the install creates a "wrapper" program for GCC, invoking it with the right parameters to make building with dietlibc very easy.

Virtualization with KVM

Introducing KVM, its internals and how to configure and install it. IRFAN HABIB

Virtualization has made a lot of progress during the last decade, primarily due to the development of myriad open-source virtual machine hypervisors. This progress has almost eliminated the barriers between operating systems and dramatically increased utilization of powerful servers, bringing immediate benefit to companies. Up until recently, the focus always has been on software-emulated virtualization. Two of the most common approaches to software-emulated virtualization are full virtualization and paravirtualization. In full virtualization, a layer, commonly called the hypervisor or the virtual machine monitor, exists between the virtualized operating systems and the hardware. This layer multiplexes the system resources between competing operating system instances. Paravirtualization is different in that the hypervisor operates in a more cooperative fashion, because each guest operating system is aware that it is running in a virtualized environment, so each cooperates with the hypervisor to virtualize the underlying hardware.

Both approaches have advantages and disadvantages. The primary advantage of the paravirtualization approach is that it allows the fastest possible software-based virtualization, at the cost of not supporting proprietary operating systems. Full virtualization approaches, of course, do not have this limitation; however, full virtualization hypervisors are very complex pieces of software. VMware, the commercial virtualization solution, is an example of full virtualization. Paravirtualization is provided by Xen, User-Mode Linux (UML) and others.

With the introduction of hardware-based virtualization, these lines have blurred. With the advent of Intel's VT and AMD's SVM, writing a hypervisor has become significantly easier, and it now is possible to enjoy the benefits of full virtualization while keeping the hypervisor's complexity at a minimum.

Xen, the classic paravirtualization engine, now supports fully virtualized MS Windows, with the help of hardware-based virtualization. KVM is a relatively new and simple, yet powerful, virtualization engine, which has found its way into the Linux kernel, giving the Linux kernel native virtualization capabilities. Because KVM uses hardware-based virtualization, it does not require modified guest operating systems, and thus, it can support any platform from within Linux, given that it is deployed on a supported processor.

KVM

KVM is a unique hypervisor. The KVM developers, instead of creating major portions of an operating system kernel themselves, as other hypervisors have done, devised a method that turned the Linux kernel itself into a hypervisor. This was achieved through a minimally intrusive method by developing KVM as kernel module. Integrating the hypervisor capabilities into a host Linux kernel as a loadable module can simplify management and improve performance in virtualized environ-

ments. This probably was the main reason for developers to add KVM to the Linux kernel.

This approach has numerous advantages. By adding virtualization capabilities to a standard Linux kernel, the virtualized environment can benefit from all the ongoing work on the Linux kernel itself. Under this model, every virtual machine is a regular Linux process, scheduled by the standard Linux scheduler. Traditionally, a normal Linux process has two modes of execution: kernel and user. The user mode is the default mode for applications, and an application goes into kernel mode when it requires some service from the kernel, such as writing to the hard disk. KVM adds a third mode, the guest mode. Guest mode processes are processes that are run from within the virtual machine. The guest mode, just like the normal mode (non-virtualized instance), has its own kernel and user-space variations. Normal kill and ps commands work on guest modes. From the non-virtualized instance, a KVM virtual machine is shown as a normal process, and it can be killed just like any other process. KVM makes use of hardware virtualization to virtualize processor states, and memory management for the virtual machine is handled from within the kernel. I/O in the current version is handled in user space, primarily through QEMU.

A typical KVM installation consists of the following components:

- A device driver for managing the virtualization hardware; this driver exposes its capabilities via a character device `/dev/kvm`.
- A user-space component for emulating PC hardware; currently, this is handled in the user space and is a lightly modified QEMU process.
- The I/O model is directly derived from QEMU's, with support for copy-on-write disk images and other QEMU features.

How do you find out whether your system will run KVM? First, you need a processor that supports virtualization. For a more detailed list, have a look at wiki.xensource.com/xenwiki/HVM_Compatible_Processors. Additionally, you can check `/proc/cpuinfo`, and if you see `vmx` or `smx` in the `cpu flags` field, your system supports KVM.

How KVM Compares to Existing Hypervisors

KVM is a fairly recent project compared with its competitors. In an interview with Avi Kivity, the main developer, he compared KVM with alternative solutions:

In many ways, VMware is a ground-breaking technology. VMware manages to fully virtualize the notoriously

complex x86 architecture using software techniques only, and to achieve very good performance and stability. As a result, VMware is a very large and complex piece of software. KVM, on the other hand, relies on the new hardware virtualization technologies that have appeared recently. As such, it is very small (about 10,000 lines) and relatively simple. Another big difference is that VMware is proprietary, while KVM is open source.

Xen is a fairly large project, providing both paravirtualization and full virtualization. It is designed as a standalone kernel, which only requires Linux to perform I/O. This makes it rather large, as it has its own scheduler, memory manager, timer handling and machine initialization.

KVM, in contrast, uses the standard Linux scheduler, memory management and other services. This allows the KVM developers to concentrate on virtualization, building on the core kernel instead of replacing it.

QEMU is a user-space emulator. It is a fairly amazing project, emulating a variety of guest processors on several host processors, with fairly decent performance. However, the user-space architecture does not allow it to approach native speeds without a kernel accelerator. KVM recognizes the utility of QEMU by using it for I/O

How Virtualization Works

Platform virtualization is an old technology; however, in recent years, the hardware and operating systems have matured to the point of making the promise of virtualization a reality. The most fundamental part of virtualization is the hypervisor. The hypervisor acts as a layer between the virtualized guest operating system and the real hardware. In some cases, the hypervisor is an operating system, such as with Xen; in other cases, it's user-level software, such as VMware. The virtualized guest operating system, or the virtualized instance, is an isolated operating system that views the underlying hardware platform as belonging to it. But, in reality, the hypervisor provides it with this illusion.

Processor Support for Virtualization

Due to the resurgence of interest in virtualization technology, microprocessor manufacturers have updated their processors to have native support for virtualization. Doing so allows the processor to support a hypervisor directly and simplifies the task of writing hypervisors, as is the case with KVM. The processor manages the processor states for the host and guest operating systems, and it also manages the I/O and interrupts on behalf of the virtualized operating system.

Advertiser Index

For advertising information, please contact our sales department at 1-713-344-1956 ext. 2 or ads@linuxjournal.com.
www.linuxjournal.com/advertising

Advertiser	Page #	Advertiser	Page #
ABERDEEN, LLC www.aberdeeninc.com	7	O'REILLY MYSQL USERS CONFERENCE www.mysqlconf.com	83
2008 ANNUAL WEB SERVICE/SOA www.webservicesonwallstreet.com	33	PHP QUEBEC conf.phpquebec.com	57
APPRO HPC SOLUTIONS appro.com	69	POLYWELL COMPUTERS, INC. www.polywell.com	3
ASA COMPUTERS www.asacomputers.com	59, 85	THE PORTLAND GROUP www.pggroup.com	27
AVOCENT CORPORATION www.avocent.com/remotectl	1	RACKSPACE MANAGED HOSTING www.rackspace.com	C3
BLACK HAT BRIEFINGS (CONFEX PARTNERS LTD) www.blackhat.com	77	R CUBED TECHNOLOGIES www.rcubedtech.com	95
CARLNET www.cari.net	39	R1SOFT, INC. www.r1soft.com	91
CORAID, INC. www.coraid.com	21	SD BEST PRACTICES www.sdexpo.com	75
EMAC, INC. www.emacinc.com	51	SDG SYSTEMS www.sdgsystems.com	31
EMPERORLINUX www.emperorlinux.com	5	SERVERS DIRECT www.serversdirect.com	9
GENSTOR SYSTEMS, INC. www.genstor.com	28	SILICON MECHANICS www.siliconmechanics.com	19, 87
HURRICANE ELECTRIC www.he.net	17	SLICKEDIT INC www.slickedit.com	55
INTEL www.intel.com	C2	SXSW FESTIVALS AND CONFERENCES www.sxsw.com	63
IT360 www.it360.ca	71	TECHNOLOGIC SYSTEMS www.embeddedx86.com	62
LOGIC SUPPLY, INC. www.logicsupply.com	6	USENIX ASSOCIATION www.usenix.com/events/fast08	25
LPI www.lpi.org	89	VERIO www.verio.com	47
MICROWAY, INC. www.microway.com	C4, 73	VIRTUAL LOGIX www.virtuallogix.com	45

However, the project is continuing at a rapid pace, and according to Avi Kivity, KVM already is further ahead than Xen in some areas and surely will catch up in other areas in the future.

hardware emulation. Although KVM is not tied to any particular user space, the QEMU code was too good not to use—so we used it.

KVM, however, is not perfect due to its newness; it has some limitations including the following:

- At the time of this writing, KVM supports only Intel and AMD virtualization, whereas Xen supports IBM PowerPC and Itanium as well.
- SMP support for hosts is lacking in the current release.
- Performance tuning.

However, the project is continuing at a rapid pace, and according to Avi Kivity, KVM already is further ahead than Xen in some areas and surely will catch up in other areas in the future.

Installing KVM

KVM has been added to many distribution-specific repositories, including OpenSUSE/SUSE, Fedora 7 (which comes with KVM built-in), Debian and Ubuntu (Feisty).

For other distributions, you need to download a kernel of version 2.6.20 and above. When compiling a custom kernel, select Device Drivers→Virtualization when configuring the kernel, and enable support for hardware-based virtualization. You also can get the KVM module along with the required user-space utilities from sourceforge.net/project/showfiles.php?group_id=180599.

I have installed the OpenSUSE packages; hence, filenames used in the examples in this article may be different from those in your release.

Creating the Guest OS

Using the compiled kernel with virtualization support enabled, the next step is to create a disk image for the guest operating system. You do so with `qemu-img`, as shown below. Note that the size of the image is 6GB, but using QEMU's copy-on-write format (`qcow`), the file will grow as needed, instead of occupying the full 6GB:

```
# qemu-img create -f qcow image.img 6G
```

Instantiation of a new guest operating system is provided by a utility called `qemu-kvm`. This utility works with the `kvm` module, using `/dev/kvm` to load a guest, associate it with the virtual disk (a regular QEMU `qcow` file in the host operating system), and then boot it. In some distributions this utility may be called `kvm`.

With your virtual disk created, load the guest operating system into it. The following example assumes that the guest operating system is on a CD-ROM. In addition to populating the virtual disk with the CD-ROM ISO image, you must boot the image when it's done:

```
# qemu-kvm -m 384 -cdrom guestos.iso -hda image.img -boot d
```

The I/O in the current release of KVM is handled by QEMU, so let's look at some important QEMU switches:

- `-m`: memory in terms of megabytes.
- `-cdrom`: the file, ideally an ISO image, acts as a CD-ROM drive to the VM. If no `cdrom` switch is specified, the `ide1` master acts as the CD-ROM.
- `-hda`: points to a QEMU copy-on-write image file. For more hard disks we could specify:

```
#qemu-kvm -m 384 -hda vmdisk1.img -hdb vmdisk2.img -hdc vmdisk3.img
```

- `-boot`: allows us to customize the boot options; the `-d` switch boots from the CD-ROM.

The default command starts the guest OS in a subwindow, but you can start in full-screen mode, by passing the following switch:

```
-full-screen
```

Additionally, KVM allows low-level control over the hardware of the virtualized environment. You can redirect serial, parallel and USB ports to specific devices by specifying the appropriate switches. Sound in the VM is supported as well, and you can pass your sound card to the VM via the `-soundhw` switch to enable sound.


The following are some keyboard shortcuts:

- Ctrl-Alt-F: toggle full screen.
- Ctrl-Alt-N: switch to virtual console N.
- Ctrl-Alt: toggle mouse and keyboard.

Conclusion

With the introduction of KVM into the Linux kernel, future Linux distributions will have built-in support for virtualization, giving them an edge over other operating systems. There will be no need for any dual-boot installation in the future, because all the applications you require could be run directly from the Linux desktop. KVM is just one more of the many existing open-source hypervisors, reaffirming that open source has been instrumental to the progress of virtualization technology. ■

Irfan Habib is student of software engineering at the National University of Sciences and Technology, Pakistan. He loves to code in Python, which he finds to be one of the most productive languages ever developed.



April 14 –17, 2008
Santa Clara, California

Harness the Power of MySQL

The **2008 MySQL Conference & Expo** brings over 1,600 open source and database enthusiasts, developers, gurus, experts and users together to harness the power of MySQL and celebrate the huge MySQL ecosystem:

- Big ideas, know-how, and connections packed into four concentrated days and nights
- More than 100 sessions encompassing a wide range of skill levels
- In-depth tutorials to help develop technical skills and learn new tricks
- Inspirational keynotes explore open source in the enterprise and beyond
- An Expo Hall featuring dozens of the latest tools and products from cutting edge vendors
- Opportunity to become MySQL Certified
- A DotOrg pavilion showcasing community projects at the forefront of MySQL innovation

Join us at the **2008 MySQL Conference & Expo**—you'll benefit not just from gathering with your peers and experts from a wide variety of companies and countries, but also by learning from some of the leading visionaries in the open source world.



Register and save 10% when you use discount code: **mys08ljr**
www.mysqlconf.com

The Best of Both Worlds

Running Linux inside Windows using QEMU. DASHAMIR HOXHA

I recently bought an IBM ThinkPad laptop with 1GB of RAM and Windows XP preinstalled. Because I have been using only Linux for many years, I immediately thought about making it a dual-boot system (actually a multiboot system, because I usually install several copies of Linux on my computer).

As I said, I mainly use only Linux, but I also keep a copy of Windows around, because other people may need to use my computer who are not able to use Linux. Also, being a computer specialist, I like knowing all the ways of using a computer, not only the best one, and as many people still use Windows, I want to understand their points of view.

So, I now can reboot and switch from Linux to Windows and from Windows to Linux. However, I thought it would be

useful to run both systems in parallel, instead of switching from one to the other. One of the reasons for this is the Windows XP Home Edition that was installed on my laptop is customized by IBM specifically for this laptop, and there are some tools developed by IBM that make things more convenient. Another reason is that I wanted to test a client-server network with Windows as the client and Linux as the server. I'm sure you can think of other reasons for doing this as well.

After some research and testing, I decided to use QEMU. Now I can run any of the Linux distributions that are installed on the other partitions on Windows. I also can access Windows from the Linux system. I can access the Internet from the Linux system, and I can access any of the Linux services

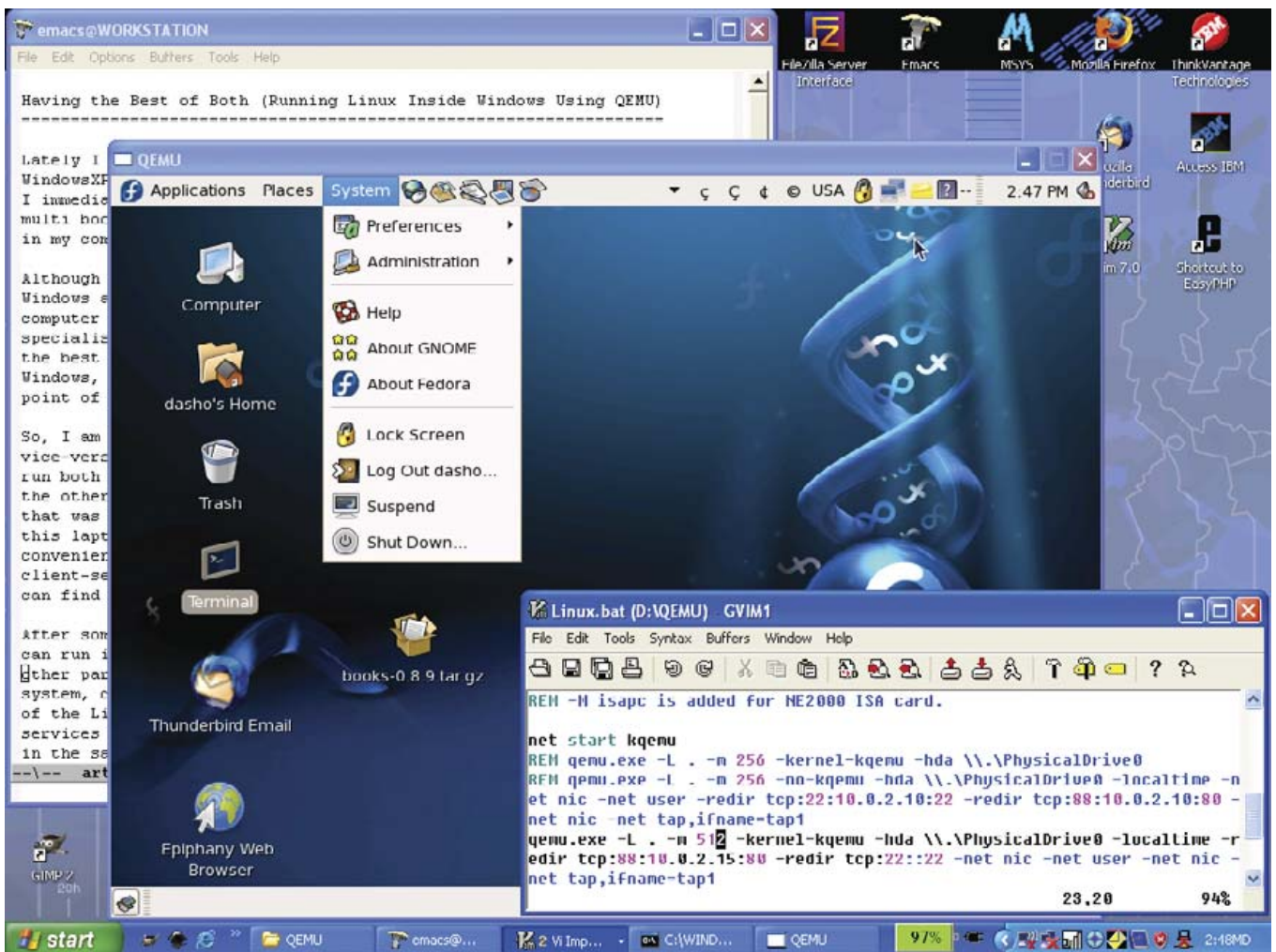


Figure 1. Fedora Core 6 Running inside Windows XP through QEMU

from Windows. Additionally, I can access certain Linux services from the network. It's like having two systems running on the same machine at the same time.

Running Linux inside Windows using QEMU is not difficult; however, doing it well requires some tricks that I didn't discover immediately.

Installing and Running QEMU

Installing QEMU in Windows is easy. I downloaded qemu-0.9.0-windows.zip and extracted it in D:\QEMU. I didn't forget to read README-en.txt (always read READMEs). Then, I made a copy of the batch file (script) qemu-win.bat and renamed it start-linux.bat. To access it more easily, I created a shortcut (link) for it on the desktop by doing a right-click and selecting Send to→Desktop. Then, I modified the last line of start-linux.bat to look like this:

```
qemu.exe -L . -m 128 -hda \\.\PhysicalDrive0 -soundhw all -localtime
```

The modification consists of replacing the parameter -hda linux.img with the parameter -hda \\.\PhysicalDrive0. Now, when I start QEMU by running this script, instead of using the file linux.img as a virtual hard disk, it uses my real hard disk and boots from it. Then, I see the beautiful GRUB menu that is installed in the MBR of my hard disk, and I select and boot one of my Linux systems. Isn't it great?

Be careful not to boot Windows again inside Windows. According to the documentation, using the same disk image in more than one machine can corrupt it.

Running Linux

The system that I usually boot inside Windows is Fedora Core 6. The parameter -m 128 tells QEMU to use up to 128MB of RAM for the emulated system. With 128MB of RAM, Fedora isn't able to run in graphic mode and falls back to text mode. However, with 256MB of RAM, it works. If you have 1GB of RAM in your machine, like me, you could be generous and give 512MB to Linux.

The graphical interface is important to me, but I am quite happy with command-line Linux. In order to run Fedora in text mode, even though it has 256MB of RAM, I pass the 3 parameter to the kernel, which tells it to boot in run-level 3. Initially, I did this manually, with these steps:

- Select Fedora in the GRUB menu.
- Press E to edit it.
- Select the kernel line.
- Press E to edit it.
- Append 3 at the end of the kernel line, and press Enter to return.
- Press B to boot the modified Fedora entry.



Want your business to be more productive?
The ASA Servers powered by the Intel Xeon Processor provide the quality and dependability to keep up with your growing business.

Hardware Systems for the Open Source Community Since 1989
(Linux, FreeBSD, NetBSD, OpenBSD, Solaris, MS, etc.)

1U Woodcrest/Clovertown Storage Server Starts at - \$1,741



- 1TB Storage installed. Max - 3TB.
- 1U Dual core 5030 CPU (Qty-1). Max - 2 CPUs.
- 1GB 667MGZ FBDIMMs Installed.
- Supports 16GB FBDIMM.
- 4X250GB htswap SATA-II Drives installed.
- 4 port SATA-II RAID controller.
- 2X10/100/1000 LAN onboard.

2U Woodcrest/Clovertown Storage Server Starts at - \$3,791

- 4TB Storage installed. Max - 12TB.
- 3U Dual core 5050 CPU.
- 1GB 667MGZ FBDIMMs Installed.
- Supports 16GB FBDIMM.
- 16 port SATA-II RAID controller.
- 16X250GB htswap SATA-II Drives installed.
- 2X10/100/1000 LAN onboard.
- 800w Red PS.



3U Woodcrest/Clovertown Storage Server Starts at - \$3,991



- 4TB Storage installed. Max - 12TB.
- 3U Dual core 5050 CPU.
- 1GB 667MGZ FBDIMMs Installed.
- Supports 16GB FBDIMM.
- 16X250GB htswap SATA-II Drives installed.
- 16 port SATA-II RAID controller.
- 2X10/100/1000 LAN onboard.
- 800w Red PS.

5U Woodcrest/Clovertown Storage Server Starts at - \$6,691

- 6TB Storage installed. Max - 18TB.
- 5U Dual core 5050 CPU.
- 4GB 667MGZ FBDIMMs Installed.
- Supports 16GB FBDIMM.
- 24X250GB htswap SATA-II Drives installed.
- 24 port SATA-II RAID, CARD/BBU.
- 2X10/100/1000 LAN onboard.
- 930w Red PS.



8U Woodcrest/Clovertown Storage server Starts at - \$11,191



- 10TB Storage installed. Max - 30TB.
- 8U Dual core 5050 CPU.
- 2X5050 installed.
- 1GB 667MGZ FBDIMMs.
- Supports 32GB FBDIMM.
- 40X250GB htswap SATA-II Drives installed.
- 2X12 Port SATA-II Multilane RAID controller.
- 1X16 Port SATA-II Multilane RAID controller.
- 2X10/100/1000 LAN onboard.
- 1850 W Red Ps.

All systems installed and tested with user's choice of Linux distribution (free). ASA Collocation—\$75 per month



2354 Calle Del Mundo,
Santa Clara, CA 95054
www.asacomputers.com
Email: sales@asacomputers.com
P: 1-800-REAL-PCS | FAX: 408-654-2910

Intel®, Intel® Xeon™, Intel Inside®, Intel® Itanium® and the Intel Inside® logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Prices and availability subject to change without notice. Not responsible for typographical errors.



Later, I added another entry to the menu, with the 3 parameter appended to the kernel line in order to boot it more quickly, which looks like this:

```
title Fedora Core TextMode (2.6.18-1.2798.fc6)
    root (hd0,7)
    kernel /boot/vmlinuz-2.6.18-1.2798.fc6 ro root=/dev/hda8 rhgb quiet 3
    initrd /boot/initrd-2.6.18-1.2798.fc6.img
```

Sometimes I see several error messages and failures while Linux is booting (for example, when I tried Scientific Linux), but I ignore them. The reason for this is the hardware of the emulated machine (which is being emulated by QEMU) is somewhat different from the hardware of the real machine. The same thing happens when the hard disk is taken from one machine and placed in another. Linux autodetects the machine's devices and reports that some devices are missing and new devices are added (for example, network cards). I simply keep the configurations of the "removed" devices and let Linux autoconfigure the new devices it finds.

Making QEMU Run Faster

To make the emulated system run faster, I installed `kqemu`. I downloaded `kqemu-1.3.0pre11.tar.gz` from the QEMU download page and extracted it inside `D:\QEMU`. Then, I clicked `kqemu.inf` with the right-mouse button and selected `Install`. Next, I added, in `start-linux.bat`, the command `net start kqemu` and added the parameter `-kernel-kqemu` to `qemu.exe`. Now, the last two lines of `start-linux.bat` look like this:

```
net start kqemu
qemu.exe -L . -m 256 -kernel-kqemu -hda \\.\PhysicalDrive0
    ↪-soundhw all -localtime
```

Note: Scientific Linux 4.4 does not work at all with the parameter `-kernel-kqemu`, and the problem seems to be an incompatibility of the kernel with the BIOS file (which is named `bios.bin`, and I think that represents the BIOS configuration of the emulated system). When I replaced it with the BIOS of Puppy Linux, it worked. It is strange that the original BIOS is 128KB and Puppy's BIOS is 64KB, which is older as well.

Accessing Windows and the Internet from Linux

The default QEMU parameters for the network are `-net nic -net user`. This means that it will emulate a virtual interface on the Windows side and create a network interface `eth0` for the emulated Linux system. Both of these interfaces have a virtual connection

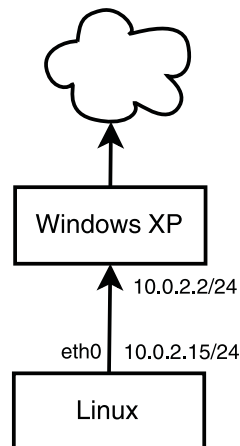


Figure 2.
Network Diagram

between them, allowing them to communicate with each other. The IP of the Windows virtual interface is `10.0.2.2/24`, and QEMU also creates a virtual DHCP server connected to it. To get an IP for Linux from the QEMU DHCPD, I log in as `root` and give the command `dhclient`. Then, the Linux interface gets IP `10.0.2.15/24`, gateway `10.0.2.2` and DNS `10.0.2.3`. Afterward, Windows and the Internet can be accessed from Linux without a problem.

To check the network configuration, try the commands `ip address ls`, `ip route ls` and `cat /etc/resolv.conf` in Linux. Here's example output from those commands:

```
[root@fedora6 ~]# ip address ls
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 16436 qdisc noqueue
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast
    qlen 1000
    link/ether 52:54:00:12:34:56 brd ff:ff:ff:ff:ff:ff
    inet 10.0.2.15/24 brd 10.0.2.255 scope global eth0
3: sit0: <NOARP> mtu 1480 qdisc noop
    link/sit 0.0.0.0 brd 0.0.0.0
```

```
[root@fedora6 ~]# ip route ls
10.0.2.0/24 dev eth0 proto kernel scope link src 10.0.2.15
default via 10.0.2.2 dev eth0
```

```
[root@fedora6 ~]# cat /etc/resolv.conf
; generated by /sbin/dhclient-script
nameserver 10.0.2.3
```

To test the network connection, try the commands `ping 10.0.2.2` and `wget http://www.google.com/`.

If you try to ping `www.google.com` (or any other IP), it won't work. However, the network connection is okay and working, and you can verify it with other tools, such as `wget` or `lftp`. It simply means that ping is not working for some reason. This has been very misleading to me, because the usual way to check for network connectivity is to ping something out there.

Accessing Windows Files from Linux

Because I can ping Windows from Linux as `10.0.2.2`, I also can access any service (daemon) that runs on Windows. In particular, I can access any file-sharing services. Usually, I run Apache as a Web server on Windows. It can be installed easily with EasyPHP. Then, I use `wget` to retrieve any files that are accessible through the Web server.

Another service I use is FTP. I install and configure it using the FileZilla FTP server. From Linux, I can access the folders (directories) that are shared by the FTP service, using `lftp` (you can use any other ftp client as well). This is better than using a Web server, because I can transfer files both ways—from Windows to Linux and from Linux to Windows.

I've even used `svnserve` to run a Subversion service in Windows. From Linux, I could access the Subversion repositories

in Windows. Think of it as a way to transfer files between Windows and Linux, as you can access the svn repositories from both Windows and Linux, although it is not a very straightforward way to transfer files.

I tried transferring files between Linux and Windows using a fat32 partition, which can be accessed from both systems. Theoretically, there is no reason why it should not work, and actually it does work; however, it does not work so well. The problem is that the modifications that are done to fat32 from Linux are not "seen" from Windows until it is restarted, and the modifications that are done from Windows are not "seen" from Linux until QEMU is restarted, which makes this solution impractical and unusable.

Accessing Linux's httpd and sshd

In order to access the Web server and Linux's secure shell, I added these parameters to the qemu.exe command:

```
-redir tcp:88:10.0.2.15:80 -redir tcp:22::22
```

The first -redir parameter makes QEMU answer any requests to port 88. Actually, it is not going to answer it itself but redirects it to server 10.0.2.15, port 80, which is the Linux

Web server. I chose port 88 (different from 80) in case I need to run any other Web service (such as EasyPHP) in Windows, so they don't have conflicts with each other. To test that it's working, open <http://127.0.0.1:88/> in a browser. Make sure that the Linux network interface has been configured (with dhclient) after the Linux server has been started.

The second -redir parameter makes QEMU redirect any connections to port 22 (secure shell) to Linux's port 22. If the server IP is missing, the default value is 10.0.2.15, which corresponds to the IP given by the emulated DHCPD of the emulated system (Linux). To access the shell of a Linux server from Windows, I use PuTTY by connecting to 127.0.0.1 port 22. Accessing the Linux shell through PuTTY is much more convenient than accessing it through the QEMU console, because I can open several terminals at the same time, and I can copy/paste between Linux and Windows. I also can enlarge PuTTY terminals and adjust fonts and colors. It also is possible to use pscp to copy files between Windows and Linux through SSH.

If you want to make these Linux services (httpd and sshd) accessible to the network as well (so they can be accessed from other computers on the local network), open the Windows firewall for them: Control Panel→Windows

Expert Included.

As product manager for HPC clusters, Kirtan helps to make the complicated process of sourcing, pricing, and acquiring a cluster as straightforward as possible for customers. Kirtan is pleased to be collaborating with Intel and other members of the Intel Cluster Ready program. He works to develop cluster solutions that will support multiple Intel Cluster Ready ISV applications.

He is proud that Silicon Mechanics has developed an industry-first cluster configurator, part of the RackScale Clusters product line, which allows users to dynamically configure complete cluster solutions – online, interactively.

The combination of Silicon Mechanics hardware, the RackScale Clusters online configurator, and the Intel Cluster Ready program offers customers tremendous value by reducing installation, deployment, and maintenance times for applications in their clusters.

When you partner with Silicon Mechanics, you get more than innovative Intel solutions — you get an expert like Kirtan.



SILICON MECHANICS

visit us at www.siliconmechanics.com

or call us toll free at 866-352-1173

Silicon Mechanics and the Silicon Mechanics logo are registered trademarks of Silicon Mechanics, Inc. Intel, the Intel logo, Xeon, Xeon Inside, Intel Cluster Ready and the Intel Cluster Ready logo are trademarks or registered trademarks of Intel Corporation in the US and other countries.



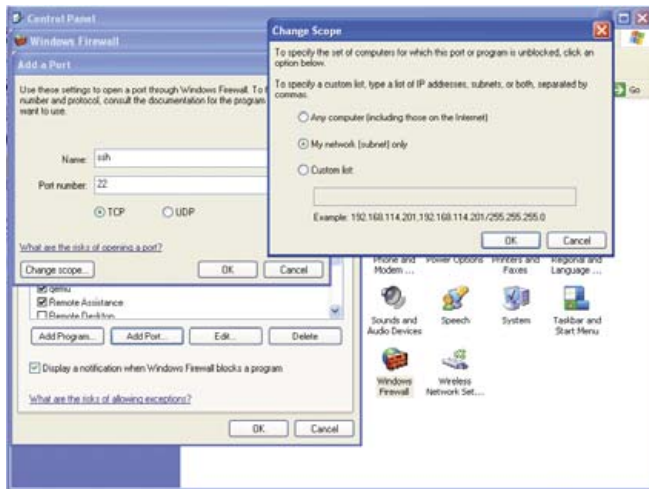


Figure 3. Opening Port 22 in the Windows Firewall

Firewall→Exceptions→Add Program. Then browse, select D:\QEMU\qemu.exe, and press OK. Next, open Control Panel→Windows Firewall→Exceptions→Add Port, and add ports 88 and 22. Also check the box Change scope... when you add or edit a program or port.

More Complex Networking between Windows and Linux

I also want to access other Linux services, such as Samba and FTP. Adding another `-redir` parameter for each port that I want to access is not convenient, and it's not an elegant solution anyway. I want to be able to access Linux from Windows without any restrictions. It does not seem to be so easy, because all that Windows can see is the `qemu.exe` process, and it has no idea what goes on inside it. So, how can Windows communicate directly with the Linux that runs inside QEMU? It is possible by creating a tap virtual Ethernet adapter using OpenVPN.

I downloaded `openvpn-2.0.9-install.exe` and installed it. During the installation, I checked only the components TAP-Win32 Virtual Ethernet Adapter, Add OpenVPN to Path and Add Shortcuts to Start Menu, because I didn't need the others. I changed the destination folder to D:\QEMU\OpenVPN, because I prefer to group the related tools together. I received some warnings that this software has not passed Windows testing, but I continued anyway, trusting that open-source testing is stronger than Windows testing.

After installation, I selected the menu Start→OpenVPN→Add a new TAP-Win32 virtual Ethernet adapter to create a new tap interface. Again, I received the same warnings, but continued anyway. Now, in Network Connections, I find a new network connection named Local Area Connection 1. I right-click on it and rename it `tap1`.

Then, I modified `start-linux.bat` by adding these parameters to QEMU:

```
-net nic,vlan=0
-net tap,vlan=0,ifname=tap1
```

```
-net nic,vlan=1
-net user,vlan=1
```

The parameter `-net nic` tells QEMU to create a new network interface for the emulated system. Because this parameter has been used twice, Linux is going to run in a machine that has two network interfaces, `eth0` and `eth1`. The parameter `-net user` creates a virtual interface on the other side (the Windows side). It is the network interface that was created previously by default (the one that has a built-in DHCP server associated with it), even though we didn't specify any `-net` parameters. The parameter `-net tap` tells QEMU to use the virtual Ethernet adapter `tap1`, which we created previously, and to connect it to the virtual network. The `vlan` options that are used with the `-net` parameters tell QEMU how to connect these virtual interfaces to each other. All the interfaces that have the same `vlan` number are connected to the same virtual hub/switch. So, we have two switches in our virtual network that is emulated by QEMU.

The last two lines of `linux-start.bat` now look like this:

```
net start qemu
qemu.exe -L . -m 256 -kernel-kqemu -hda \\.\PhysicalDrive0
➤-localtime -redir tcp:88:10.0.2.15:80 -redir tcp:22::22
➤-net nic,vlan=0 -net tap,vlan=0,ifname=tap1
➤-net nic,vlan=1 -net user,vlan=1
```

Note that the QEMU parameter `-soundhw all` is now missing. I removed it because one of the sound devices was creating conflicts with the network devices, so they were not recognized properly as `eth0` and `eth1`. If you can't do without a sound device, at least append it at the very end of the line; the parameters' order does matter.

The order of the `-net` parameter declarations matters as well. I have noticed that if `-net user` is declared before `-net tap`, `eth0` and `eth1` are switched with each other, and there is also a failure to initialize `eth0` during the Fedora initialization scripts. Keep this in mind, in case you have any similar problems.

After starting QEMU, we have a (virtual) physical network (Figure 4).

To check the "physical" connections of the network, press `Ctrl-Alt-2` to switch to the QEMU monitor. Then, in the `(qemu)` prompt, give the command `info network`. Finally, press `Ctrl-Alt-1` to get back to the Linux console. Here's the output from the command:

```
VLAN 0 devices:
tap: ifname=tap1
```

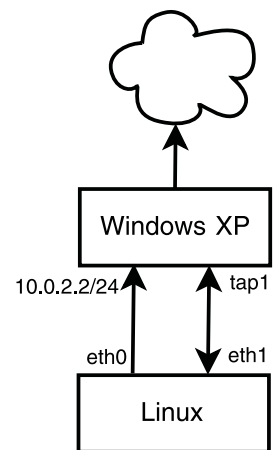


Figure 4. Physical Network

Growing a World of Linux Professionals



We at the Linux Professional Institute believe the best way to spread the adoption of Linux and Open Source software is to grow a world wide supply of talented, qualified and accredited IT professionals.

We realize the importance of providing a global standard of measurement. To assist in this effort, we are launching a Regional Enablement Initiative to ensure we understand, nurture and support the needs of the enterprise, governments, educational institutions and individual contributors around the globe.

We can only achieve this through a network of local "on the ground" partner organizations. Partners who know the sector and understand the needs of the IT work force. Through this active policy of Regional Enablement we are seeking local partners and assisting them in their efforts to promote Linux and Open Source professionalism.

We encourage you to contact our new regional partners listed above.

Together we are growing a world of Linux Professionals.



Stable. Innovative. Growing.


```

ne2000 pci macaddr=52:54:00:12:34:56
VLAN 1 devices:
user redirector
ne2000 pci macaddr=52:54:00:12:34:57

```

Now, we need just to configure the network settings properly, such as IPs, gateway and DNS.

The user redirector interface on the Windows side is configured automatically by QEMU, with IP 10.0.2.2/24, and we don't have access to it, so we cannot modify it. If you check in Network Connections, you will find that the virtual interface tap1 now appears to be connected. To configure it, right-click on it and select Properties, then select Internet Protocol (TCP/IP) and Properties again. In the configuration window, set a fixed-IP address of 192.168.10.10 and netmask 255.255.255.0. It's just like a usual network interface configuration.

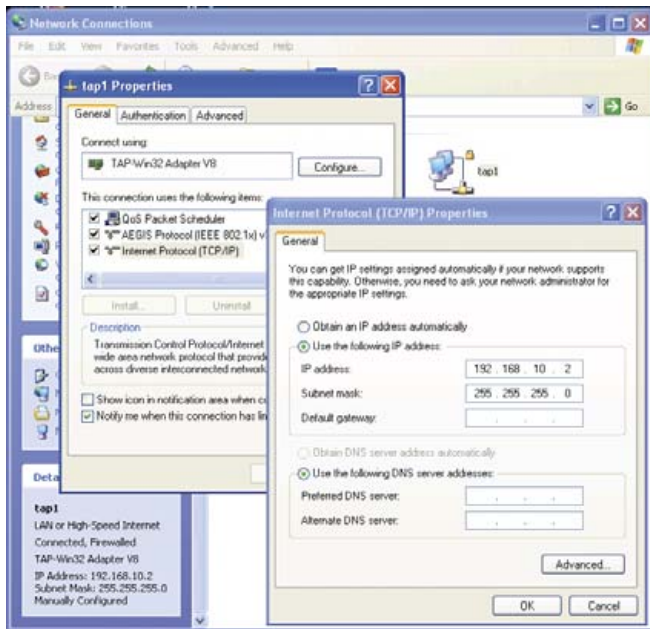


Figure 5. Configuration of the Network Interface tap1

To check that the network configuration is okay, run `ipconfig` at the command prompt, and you'll see this output for tap1:

```

Ethernet adapter tap1:

    Connection-specific DNS Suffix . . . : 
    IP Address . . . . . : 192.168.10.2
    Subnet Mask . . . . . : 255.255.255.0
    Default Gateway . . . . . :

```

This output is displayed when QEMU is running; otherwise, the information for tap1 will be something like: Media disconnected.

Now, we're done with network configuration on the

Windows side. This has to be done only once. All that is left is configuring the network interfaces on the Linux side.

First, log in as root. To check that you already have two network interfaces, run `ip addr`, and it should list eth0 and eth1. You can configure eth0 automatically, like this: `dhclient eth0`, as we did previously, and it will get an IP from QEMU's built-in DHCP server. Then, you can continue with eth1's manual configuration.

However, I prefer to use scripts whenever possible, and I want to make sure that eth0 always gets the IP 10.0.2.15/24, no matter what, because this is important for the `-redir` parameters shown previously. So, I do the network configuration on the Linux side by running this script (which has to be rerun whenever the system is rebooted):

```

bash# cat /usr/local/config/net-config-qemu.sh
#!/bin/bash
### configure the network when Linux is being
### emulated from Windows by QEMU

### network settings
IP0=10.0.2.15/24      ## eth0
IP1=192.168.10.10/24 ## eth1
GW=10.0.2.2         ## gateway
DNS=10.0.2.3

### configure eth0
ip link set dev eth0 up
ip address flush dev eth0
ip address add $IP0 dev eth0

### configure eth1
ip link set dev eth1 up
ip address flush dev eth1
ip address add $IP1 dev eth1

### set the gateway
ip route add to default via $GW

### set the DNS server
echo "nameserver $DNS" > /etc/resolv.conf

```

To check that the network configuration is okay, run `ip addr`, `ip route` and `cat /etc/resolv.conf`. Here's output from these commands:

```

[root@fedora6 ~]# ip addr
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 16384 qdisc noqueue
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
2: eth1: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc
    pfifo_fast qlen 1000
    link/ether 52:54:00:12:34:56 brd ff:ff:ff:ff:ff:ff
    inet 192.168.10.10/24 scope global eth1
3: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc

```

Continuous Data Protection

The Future of Data Centers

Can your backup software do this?	R1Soft CDP Server	Acronis® True Image	EMC Retrospect®
Daily Backups	✓	✓	✓
Hourly Backups	✓	Not Supported	Not Supported
Open File Backups	✓	✓	Not Supported
Bare-Metal Restore	✓	✓	Not Supported
Continuous Data Protection	✓	Not Supported	Not Supported
Restore Linux LVM	✓	Not Supported	Not Supported
Restore Linux Software RAID	✓	Not Supported	Not Supported
Easy To Use Web Interface	✓	Not Supported	Not Supported
Manage Thousands of Servers	✓	Not Supported	Not Supported
Control Panel Integration	✓	Not Supported	Not Supported
	\$80 - \$100 /server	\$699 /server	You Can't Afford It

Data Centers serious about uptime and performance use R1Soft.

For more information visit: www.r1soft.com or call us at **800-956-6198**



R1Soft

Continuous Data Protection

For **LINUX & WINDOWS**

```

#pfifo_fast qlen 1000
link/ether 52:54:00:12:34:57 brd ff:ff:ff:ff:ff:ff
inet 10.0.2.15/24 scope global eth0
4: sit0: <NOARP> mtu 1480 qdisc noop
   link/sit 0.0.0.0 brd 0.0.0.0
[root@fedora6 ~]#
[root@fedora6 ~]# ip route
10.0.2.0/24 dev eth0 proto kernel scope link src 10.0.2.15
192.168.10.0/24 dev eth1 proto kernel scope link src 192.168.10.10
default via 10.0.2.2 dev eth0
[root@fedora6 ~]#
[root@fedora6 ~]# cat /etc/resolv.conf
nameserver 10.0.2.3
[root@fedora6 ~]#

```

Now, all that remains is making sure the network is working as expected. The first check is to ping from Linux 10.0.2.2. If it is not working, it's possible that you need to switch eth0 and eth1. Sometimes, the network interface with MAC 52:54:00:12:34:56 is recognized by Linux as eth0, and the other as eth1, and sometimes it is recognized as eth1 and the other as eth0. This depends on the Linux distribution (Fedora, Slackware or whatever else). So, it is possible that eth0 and eth1 have gotten the wrong IP addresses from the configuration script, and in that case, the ping won't work. To solve this problem, modify the IP addresses that are assigned to eth0 and eth1 in the script `/usr/local/config/net-config-qemu.sh`, and run it again.

Now that the ping with 10.0.2.2 is working, try to ping 192.168.10.2 (tap1) from Linux. In general, it doesn't work. This is strange, because the ping to 192.168.10.10 from the command prompt in Windows does work. The problem is with the Windows firewall. To fix this, open the Control Panel, double-click Windows Firewall, select Advanced tab, select tap1 and click on Settings, then choose the ICMP tab, and here, check Allow incoming echo request. After this, the ping to 192.168.10.2 should work.

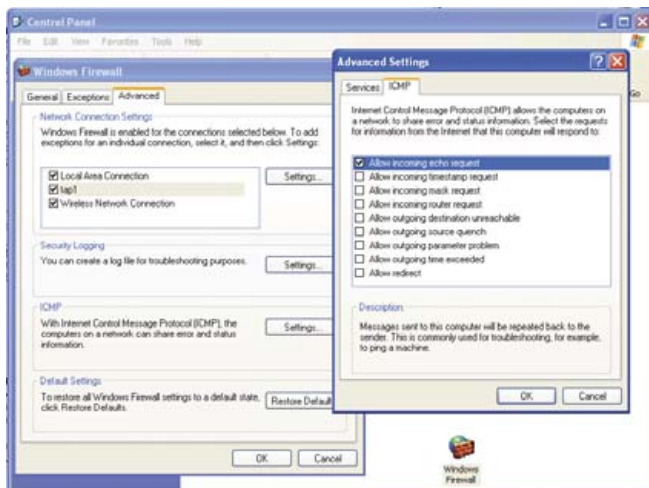


Figure 6. Allowing the tap1 Interface to Be Pinged

Don't try to ping to 10.0.2.15 from the command prompt in Windows, because it can't possibly work. Do you wonder why? Me too.

The next thing to try is accessing some Linux services from Windows using the IP 192.168.10.10. Try to open `http://192.168.10.10` in a browser, and you will see the pages that are served by the Linux Web server. Try also to log in through PuTTY to 192.168.10.10, port 22, and you will access the Linux shell.

Finally, we have a first-class bidirectional network connection between Windows and Linux, which can be used to access any Linux services from Windows. ■

Dashamir Hoxha has been a Linux specialist for many years, but occasionally, he has to use Windows as well. He has experience with server administration and network configuration.

Resources

QEMU Open-Source Processor Emulator:
fabrice.bellard.free.fr/qemu

QEMU on Windows: www.h7.dion.ne.jp/~qemu-win

OpenVPN Download: openvpn.net/download.html

Using Tap:
www.h7.dion.ne.jp/~qemu-win/TapWin32-en.html

How to Use Network: www.h7.dion.ne.jp/~qemu-win/HowToNetwork-en.html

qemu-0.9.0-windows.zip: www.h6.dion.ne.jp/~kazuw/qemu-win/qemu-0.9.0-windows.zip

QEMU Download:
fabrice.bellard.free.fr/qemu/download.html

EasyPHP Download:
www.easyphp.org/telechargements.php3

FileZilla Download:
sourceforge.net/project/showfiles.php?group_id=21558

Subversion Download: subversion.tigris.org/servlets/ProjectDocumentList?folderID=260&expandFolder=74

PuTTY Download: www.chiark.greenend.org.uk/~sgtatham/putty/download.html

Did you know Linux Journal maintains a mailing list where list members discuss all things Linux? Join LJ's linux-list today: <http://lists2.linuxjournal.com/mailman/listinfo/linux-list>.

Video Codecs and the Free World

How codecs are hurting multimedia, how Linux is dealing with it, and why free codecs can save it. **SETH KENLON**

Few video producers ever would have guessed that the term codec would become a household term, but with so many codecs on the market, average computer users have little choice but to be painfully aware that if their computer does not have the correct codec installed, they will not be able to view their favorite Web or DVD video content. Therefore, any computer enthusiast, professional sysadmin or video producer should be familiar with codecs, why they exist and how to deal with them.

The term codec is a combination of two words: code and decode. The concept is simple; if I had only one page of paper upon which I wanted to write two pages of content, I might write a note in some kind of code, leaving out certain letters or words. This would fit the content onto the one page allotted, but it would make no sense to intended readers, unless I gave those readers a key on how to decode the writing so they could piece it all back together and understand what had been written. This is precisely what a codec does with video.

Ideally, a codec effectively delivers high-quality video to end users in a reasonable amount of download time. However, large companies often opt to use a codec for its exclusivity so they can charge for the key to decode that video, such as with DVDs and streaming video. So the reason that many codecs exist at all is not to further the quality and effectiveness of video compression and delivery, but to hinder delivery to the nonpaying audience—sometimes even at the expense of video quality.

Acquiring the Right Codec

End users, system administrators and video content producers are all affected by codec compatibility and availability. Being familiar with codecs is important, as computers are now the hub of many people's entertainment centers.

A missing codec typically affects the entire system. If, for example, Xvid is not installed on the system, Xvid will not be available for Web browsers, media players or media editors. Install the Xvid component, and all of the applications on a GNU/Linux system will recognize it and utilize it when needed.

With proprietary software, even though a system has a codec installed, proprietary applications may not utilize the codec, simply because it has been programmed not to use the codec for political reasons. There is no practical reason, for instance, that Digidesign's Avid or Apple's Final Cut Pro cannot recognize and edit Ogg files, especially given the price tags of these applications.

Keep in mind that there is a difference between a codec and a file format. File formats, such as .mp4, .mov or .avi, are actually just containers for video and audio streams. So an .mp4 file, for example, may use a codec like Xvid, h.264,

x264, Decklink and so on, or it may, in fact, use the actual MPEG-4 codec (or any number of other codecs). Although the system usually can detect the actual codec being used within a file, it may confuse the user if a file format (container) is taken to be the same thing as a codec (encoder).

Most systems come with a certain amount of codecs ready to use with the OS's default Web browser and media player. Apple and Microsoft bundle their own proprietary and supported codecs with their systems, and Linux distros distribute their supported open-source codecs. Still, in all three cases, end users or system administrators are going to have to download non-bundled and unsupported codecs at some point.

On today's systems, acquiring codecs often is reduced to a few clicks of the mouse. The smoothest experience for end users may easily be on the Linux desktop, as distributions, such as Ubuntu, Linspire, Fedora and OpenSUSE, are so sensitive to the Linux-doesn't-play-media preconception that they have made it almost automated.

To get a codec to watch DVDs on your Linux machine, simply place a DVD into the computer, and the default player opens (such as Totem or Kaffeine or Xine). If an MPEG-2 decoder is not installed already, the system offers to download and install it. A few clicks of the mouse, and the codec is installed, and the movie player, usually without so much as a relaunch, plays the DVD.

There is no practical reason, for instance, that Digidesign's Avid or Apple's Final Cut Pro cannot recognize and edit Ogg files, especially given the price tags of these applications.

Similarly, most Web browsers will detect a missing codec and either direct users to the Web site containing the downloadable installer for the codec, or, even better, the system will intervene and offer to download and install the package automatically. Again, with a few clicks of the mouse, the codec is installed and ready to enable the video in the browser.

If the Web browser attempts to play a video without the proper codec installed, there are cases where the only feedback the user will get is a blank space where the video should be. You can, however, easily force the system to prompt for a codec download.

First, right-click on the empty space in the browser where the video should be playing, and open that movie (it may be missing sound and picture though) in your distro's default movie player. When the system's media player is

The smoothest experience for end users may easily be on the Linux desktop, as distributions, such as Ubuntu, Linspire, Fedora and OpenSUSE, are so sensitive to the Linux-doesn't-play-media preconception that they have made it almost automated.

unable to play the video, it will offer to find the appropriate codec and install it.

If the Linux system does not recognize the codec being used, or if your distro of choice does not offer an automated codec solution, you can do a little detective work and discover what codec you need to find and install it manually.

This kind of codec forensics is most easily done with VideoLAN's VLC player; it's free, open source and prepackaged for most popular Linux distros. Once you've installed VLC player on your system, open the movie you want to play. Most likely, VLC will play the video, but if you still need to discover its codec so that the video can play on your system outside the VLC player, go to the View menu and select Stream and Video Info. This opens a comprehensive list of the streams (video, audio, timecode, subtitle and so on) contained in the file and what codec was used in creating each stream.

Armed with this information, go back to your distribution's package manager, and search for the codec that VLC has revealed is being used. Or, if you are running Slackware or Gentoo or something similar, seek out the codec on-line. A typical open-source solution will be a GStreamer package, containing a number of open-source decoders for popular proprietary codecs. Again, these can be installed either via your package manager or manually.

If you are trying to achieve in-browser playability, be aware that sometimes a decoding package, such as Flash, will be specific to the Web browser in which you are trying to view video. So, take care to install the correct package for compatibility with your browser (Firefox, Konqueror, Opera and so on).

After you've installed the codec, relaunch your browser if necessary, and try to play the video again.

Encoding and Transcoding Video

Whether you are encoding video you have produced and want to distribute or are taking video from one source and transcoding it into a format more friendly to your system, there are handy open-source encoders that generate video files playable by many popular consumer content players. The most effective is the FFmpeg program—sometimes wrapped in a GUI and sometimes used strictly on the command line. The FFmpeg man page is extensive but fairly easy to use.

Video encoding is best learned by doing, and although there are many important variables when encoding video, there are no magic settings; they change in relation to the target size of the video as well as the actual content (for example, how frequently pixels change chroma or luma values from

one video frame to another).

To encode a video, first make sure you have FFmpeg installed on your system. If not, add it via the command line (`sudo apt-get install ffmpeg` on Debian systems), or use your distribution's package manager.

The command to encode looks like this:

```
ffmpeg -i [filename] -vcodec [codec to use] -s [target frame size]
➤ -b [target bit rate in kbps] -r [target frames per second]
➤ -acodec [audio codec to use] -pass 2 -ar [target audio sample rate]
➤ -ab [audio bit rate in kbps] -f [target file format] [output filename]
```

For example:

```
ffmpeg -i bigmovie.m4v -vcodec xvid -s 720x480 -b 3000k -r 24
➤ -acodec libfaac -pass 2 -ar 128k -f .mp4 freemovie.mp4
```

The example above transcodes a video from a proprietary MPEG-4 format into the open-source codec of Xvid with libfaac sound, in a standard definition frame size and a fairly high video quality.

The important variables are frame size (-s), as it determines scalability of the video, and bit rate (-b for video, and -ar for audio), which determines how much information is being sent per video frame—the higher the bit rate, the sharper and nicer the image will be but also the file size will be larger. Finally, the -pass variable determines how much preprocessing the encoder will do to the video. In one-pass mode, the encoding is done fairly quickly and not always optimally. With two-pass encoding, FFmpeg reviews the video file once, gathers necessary data, and then does the actual encoding on the second pass. The end result is a higher quality compression, but you can expect double or triple the encoding time.

The other popular way of transcoding video is with GUI programs that rip video from encoded DVDs, making them viewable in an unrestricted codec. This also is ideal for creating a home media server, with your entire movie collection digitized and ready to play at any moment. Obviously, the legality of this varies from day to day and from country to country. However, the GUI programs are plentiful and utilize the same variables as FFmpeg. You will need to set the video codec to which you want to save your video, the audio codec, the bit rates of each, frame size and so on.

The great codecs of the Free Software movement, Ogg Vorbis and Ogg Theora, are obviously very well supported on GNU/Linux. Generating them is done easily with `ffmpeg2theora`. The command-line variables for `ffmpeg2theora` are similar to those for FFmpeg:

```
ffmpeg2theora [filename] -x [target horizontal pixel count]
➤ -y [target vertical pixel count] -V [target bit rate in kbps]
➤ -A [audio bit rate] -c [audio channels] -H [audio sample rate]
➤ -o [output filename]
```

Using and Promoting Free Codecs

When it became clear to the Open Source movement that video codecs were doomed to remain proprietary and counter-

productive, the Ogg format was born. Open to all and freely available to any system, Ogg Vorbis (for sound) and Ogg Theora (for video) are advanced and fully featured codecs.

A common argument against using Ogg is that it requires users of Microsoft and Apple products (or, the majority of computer users) to seek out a suitable player. Yet, it's clear by now that requiring users to download a media player or media plugin is not at all uncommon and will quite probably become even more common as content delivery becomes more Internet-reliant and computer-centric. People today expect to have to download a video player to watch certain video content. The real problem with Ogg is that there is no ubiquitous media player for the format; RealPlayer has Real Media, QuickTime has QuickTime Player, Windows Media has Windows Media Player, Flash has Flash Player and so on. People easily can find those, but where do they go for Ogg playback?

Promote both free software and free codecs by promoting Ogg formats, but don't fail to promote a player that easily and effectively plays media on all major platforms. One of the better players for this job is VLC player, which installs on Windows, Mac OS X, GNU/Linux, BSD, BeOS and Solaris. Another is miro, an iTunes-like aggregator of video podcasts, IP TV and YouTube, as well as media on your local machine. Both play Ogg, so send a link to the player along with the Ogg clips you distribute.

Video Codecs and the Free World

The fact is, and will remain, that open-source tools are the saving grace of video professionals and system administrators working in a multiplatform multimedia world. FFmpeg and VLC Player have both been trusted playback, video-analyzing and video-conversion programs in my video toolkit for years, and they solve postproduction problems that proprietary, overpriced editing packages introduce with exclusive codec licenses and incompatibility.

Here are two contrasting examples:

- A recent update to Final Cut Suite,

Apple's premier video production suite, dropped support for a number of codecs while adding support for Apple's proprietary ones. Being a closed system, there is no solution to this problem, only the work-around of transcoding the source material.

- The open-source application Blender supports any codec that its host system supports, and updates can be requested of programmers on Blender's IRC channel, often resulting in a patch within days. Final Cut Studio is well over a thousand dollars per client license. Blender, of course, is free.

As long as the primary market for codecs are the companies that continue to desire to protect their digital content, new codecs will continue to be developed that will require a separate license to use. This will result in myriad codecs on the Web and in the video production world. And, as long as licenses are required to use proprietary codecs, the more divided and convoluted the delivery methods will become.

Utilize free codecs fearlessly and unify video production as well as delivery. The Open Source movement is stronger than ever, and the Creative Commons ideals in the art world are getting serious press for encouraging freely distributed works by big-name acts like Radiohead and Prince, adventurous independents like the movie *Rune*, bountiful pod-safe music, the Internet Archive and so on. The climate is such that free codecs have the unique opportunity to become the popular choice for maximized compatibility and end-user freedom. ■

Seth Kenlon is a film and video editor, systems consultant and software trainer. Concurrently with all of that, he is a Linux user, supporter and promoter.

Resources

VideoLAN—VLC Player:
www.videolan.org

miro: www.getmiro.com

Blender: www.blender.org

Linux Laptops

Starting at \$799



Linux Desktops

Starting at \$375



Linux Servers

Starting at \$899



**DON'T BE SQUARE!
GET CUBED!**



R³ Technologies
Making Stuff to Bring Technology to Life

309.34.CUBED
shopcubed.com

Life in the Vast Lane

What lives past the Web 2.0 bubble. DOC SEARLS

Back when the O'Reilly folks were pumping up "Web 2.0" as a meme—successfully, as it turned out—I was asked for some quotage on the subject. My reply, oft-quoted since, was "It's what we'll call the next crash." I'm still sure about that prediction, in part because the Web 2.0 bubble is fed by the same kind of gas that inflated the dot-com bubble nearly a decade ago: a combination of exit-oriented venture investment and tech media fantasizing.

Back in the late '90s, everybody was talking up "portals", and advertising was going to pay for everything. Now, in the getting-late '00s, everybody is talking up "social networks", and advertising is going to pay for everything. As I write this, the Facebook social network has attracted a reported 50 million faces and has just announced plans to make money by somehow "monetizing" those faces' social connections. It's wacky stuff, but not wacky enough to scare away value-inflating buzz, especially by the Big Boy media.

Venture capitalists, now as then, angle their investments less toward revenues than toward get-rich IPOs or acquisitions by existing giants, such as Google and News Corp. Thanks to a complicated \$250 million advertising deal with Microsoft, which also acquired a 1.6% share of Facebook—hottest of the social networks—the *New York Times* and *CNN Money* both figured that Facebook was suddenly worth \$15 billion—a figure that went viral almost instantly. Never mind that Facebook's revenues were only \$150 million. Never mind that the details of the Microsoft deal were not fully revealed, and that Microsoft had an obvious interest in pricing a potential Facebook acquisition beyond the reach of everybody, especially Google—a company whose stock price in November passed \$724 per share, approaching a market cap of .25 trillion dollars. Both Facebook and Google are Web 2.0 companies with valuations inflated by bubble gas. Meet the new tank, same as the old tank.

One difference between the first bubble and this one is advertising. During the dot-com bubble, advertising was mostly promise. In the Web 2.0 bubble, advertising is delivering, big-time. But, can advertising pay for everything forever? More to the point of social networks, will on-line society put up with it? As Alan Patrick puts it, "Planet Advertising desperately wants to believe we will all trust all our 'friends' who start spamming us with ads, but they misunderstand the entire dynamic of trusted networks. We trust friends precisely because they don't do this sort of thing."

The 50-million-member Facebook jury is out on that one, because Facebook's monetize-your-friendships system (code name: Beacon) isn't running yet. But, even if it works, it's still just advertising. In the long run, there's going to be a lot more money in helping demand find supply than in helping supply find (or create) demand—simply because the efficiencies involved in helping money-in-hand find places to go exceed the guesswork that defines advertising at its core. That even goes for Google, which introduced the radical notion of accountability, but still involves mountains of wasted placements (by countless Linux servers pushing gazillions of tiny text ads into the margins of blogs and search results). I'm not saying that advertising ends, by the way, just that its fate is to become part of an informational ecosystem that supports the buying intentions of customers at least as well as it supports the selling intentions of vendors.

Problem is, a symmetrical market ecosystem will have trouble emerging from a networked world optimized for cable television. Both cablecos and telcos have been pushing crippled asymmetrical "last mile" Internet service for the duration. They've shown little inclination (at least here in the US) toward supporting the Net's original end-to-end architecture, which supports abundant supply and abundant demand all over the place, and not just flowing down from a few industrial giants to a zillion "consumers".

By focusing on battles between sell-side giants (even in the "social space"), big media hasn't helped. But, here and there we see exceptions. One example is a November 2007 *Forbes* column by Peter Huber titled "Web 50.0". Here's an excerpt:

...digital life's two woeful deficiencies are both centered on the last yard of the network, where the bit meets the brain. Our eyes can process images far faster than the wire plugged in to the back of the box can deliver them. And our brains can crank out signals—through vocal cords or fingers—with far greater speed, dexterity and sensitivity than today's man-machine interfaces can match. This one area of computing is still stunningly primitive. Touch typing is a century old. A mouse is only a modest advance over a telegraph key.

But laser light channeled through glass



fiber can move pictures faster than the eye can see. And with micromechanical and semiconductor sensors, it's now possible to build gloves as digitally sensitive as human hands, or systems that move the image on the screen in response to how you move your head and eyes. The digital revolution is now waiting for these technologies to converge and proliferate. Then the revolution starts all over again....

For a hint at what's to come, Huber says:

...sneak into your teenager's bedroom. Ignore the dusty Dell, Mac and even the iPhone—marvel instead at Microsoft's Xbox, Sony's PlayStation or Nintendo's Wii. For a preview of what you'll be doing on such a machine, don't waste time trying to type a letter—where's the keyboard, anyway?—or run a spreadsheet, or Google a search, or crawl through eBay or Amazon. Go kill someone in *Halo 3*.

What matters isn't that these are games. It's that they're live. They involve data-thick interaction in real time.

I've maintained for some time that the most important step forward in the Net's recent history is not the generational progression from 1.0 to 2.0, but the branching of the Live Web off the Static Web. The big challenge is building out the Live Web, and it's not one we should leave up to the Big Boys, even as we run it over their glass.

That's because the critical enabling feature of the Live Web won't be technical. It will be the moral and political feature we call freedom. That's not something the Big Boys are going to give us. It's something that comes from ourselves. ■

Doc Searls is Senior Editor of *Linux Journal*. He is also a Visiting Scholar at the University of California at Santa Barbara and a Fellow with the Berkman Center for Internet and Society at Harvard University.



Russ Barnard, President, FlapDaddy Productions

“Fanatical Support™ saved me from my own mistake.”

“Not long ago, I reformatted one of our servers. Not until I was driving home did I learn that I brought our entire site down in the process. I called my guy at Rackspace and he said, ‘We’re already on it.’ By the time I pulled in the driveway, my site was back up. Now that’s Fanatical Support.”

Keeping little mistakes from causing big problems is one definition of Fanatical Support. What will yours be?

Watch Russ’s story at www.rackspace.com/fanatical
1-888-571-8976

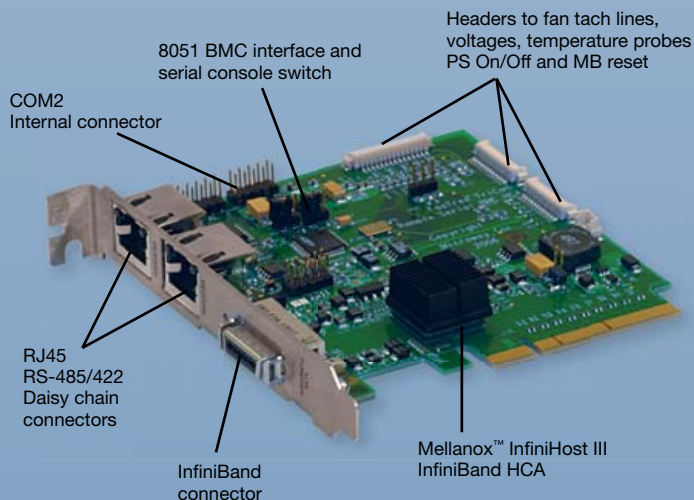


Affordable InfiniBand Solutions

4 Great Reasons to Call Microway NOW!

1 **TriCom™**

- DDR/SDR InfiniBand HCA
- "Switchless" serial console
- NodeWatch web enabled remote monitor and control



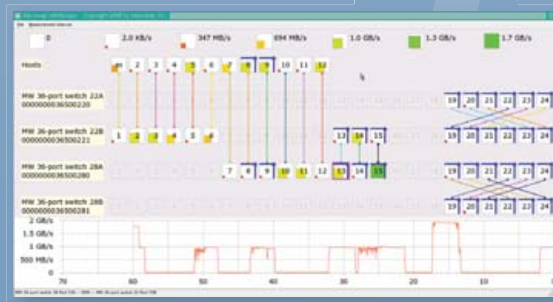
2 **FasTree™**

- DDR InfiniBand switches
- Low latency, modular design
- 24, 36 and 48 port building blocks



3 **InfiniScope™**

- Monitors ports on HCA's and switches
- Provides real time BW diagnostics
- Finds switch and cable faults
- Lane 15 interface
- Logs all IB errors



4 **ServaStor™**

- Extensible IB based storage building blocks
- Redundant and scalable
- Parallel file systems
- Open source software
- On-line capacity expansion
- RAID 0,1,1E, 3, 5, 6, 10, 50



Upgrade your current cluster, or let us design your next one using Microway InfiniBand Solutions.

To speak to an HPC expert
call **508 746-7341** and ask
for technical sales or email
sales@microway.com
www.microway.com

**Microway**
Technology you can count onsm