ANATOMY OF A LINUX SYSTEM

> The cutting edge of open source software will belong to people who start from individual vision and brilliance, then amplify it through the effective construction of voluntary communities of interest.
>
> —Eric S. Raymond
> *The Cathedral & The Bazaar*

Linux® is named after Linus Torvalds, architect of the Linux kernel, the heart of the operating system. But a complete Linux distribution contains the work of hundreds of separate open source software projects. A surprise to many people is the amount of code that companies like Sun, SGI, and Digital (Compaq) have contributed. This poster peels back the skin of the Linux distribution to show many of the major projects and their relationship to each other and to the whole Linux anatomy.

The Big Ideas Behind Linux
Eric S. Raymond's *The Cathedral & The Bazaar* outlines the way a distributed network of programmers can build leading-edge, high-quality software without centralized control. Another O'Reilly book, *Open Sources*, contains essays on open source software development methodology by many of the key developers who have made it happen.

Magazines
Open Magazine: www.openmagazine.net
Linux Magazine: www.linux-mag.com
Linux Journal: www.linuxjournal.com
LinuxWorld: www.linuxworld.com
Maximum Linux: www.maximumlinux.com

Major Linux Distributors
Red Hat®: www.redhat.com
Debian GNU/Linux: www.debian.org
SuSE: www.suse.com
Linux-Mandrake™: www.linux-mandrake.com
Slackware® Linux: www.slackware.com
Storm Linux™ from Stormix Technologies, Inc.: www.stormix.com
OpenLinux™ from Caldera Systems: www.caldera.com
TurboLinux®: www.turbolinux.com
LinuxPPC: www.linuxppc.com
Yellow Dog Linux™: www.yellowdoglinux.com
Hard Hat™ Linux for Embedded Systems from MontaVista Software, Inc.: www.mvista.com
RTLinux™ from FSMLabs: www.fsmlabs.com
or www.rtlinux.org

Hardware/Systems
VA Linux Systems: www.valinux.com
Penguin Computing: www.penguincomputing.com
IBM: www.ibm.com/linux
Cobalt Networks, Inc.: www.cobalt.com

Conferences
O'Reilly Open Source Convention: conferences.oreilly.com
LinuxWorld New York: www.linuxworldexpo.com
LinuxWorld San Jose: www.linuxworldexpo.com
Atlanta Linux Showcase: ww.linuxshowcase.com
ApacheCon: www.apachecon.com
YAPC (Yet Another Perl Conference): www.yapc.org/America or www.yapc.org/Europe

Top Linux Web Sites
www.linux.com
www.linux.org
www.slashdot.org
www.linuxtoday.com
www.lwn.net
www.linuxgazette.com

O'Reilly Sites
www.oreilly.com
www.oreillynet.com
linux.oreilly.com

Project Hosting
www.collab.net
www.sourceforge.net

Downloads
www.ibiblio.org
www.freshmeat.net
www.themes.org
www.rpmfind.net/linux/RPM

Kernel (News and Notes)
www.kernel.org
www.kernelnotes.org

Philosophy
www.fsf.org
www.opensource.org
www.opencontent.org

C/C++ Programming Tools

The FSF's gcc C compiler is the single most important programming tool for Linux, since it is required for building the system and all the other tools. Other key tools include *make*, source code control tools like CVS and RCS, and editors like vi and Emacs. There are many variations of vi, including vim, nvi, elvis, and vile. Along with the C compiler, all systems need and rely on various libraries, including glibc and libstdc++.

Useful Books:
- *Programming with GNU Software* Mike Loukides, Andy Oram
- *Learning the vi Editor* Linda Lamb, Arnold Robbins
- *CVS Pocket Reference* Gregor N. Purdy
- *Learning GNU Emacs* Debra Cameron, Bill Rosenblatt, Eric S. Raymond
- *GNU Emacs Manual* Richard M. Stallman (FSF)
- *Managing Projects with make* Andrew Oram, Steve Talbott
- *Debugging with GDB: The GNU Source-Level Debugger* Richard M. Stallman, Cygnus Solutions (FSF)

Perl, Tcl, and Python

Scripting languages are widely used on Linux for everything from system administration to generating web content. Larry Wall's Perl is the most widely used but John Ousterhout's Tcl and Guido van Rossum's Python are also extremely popular languages, all included in a typical Linux distribution. Thousands of Perl modules are available from CPAN, the Comprehensive Perl Archive Network. The Tk toolkit allows the creation of graphical applications from any of these languages. ActiveState is developing a new IDE for Perl and Python called Komodo, which is based on Mozilla.

Useful Books:
- *Programming Perl* Larry Wall, Tom Christiansen, Jon Orwant
- *Learning Perl* Randal L. Schwartz, Tom Christiansen
- *Perl Cookbook* Tom Christiansen, Nathan Torkington
- *Perl in a Nutshell* Ellen Siever, Stephen Spainhour, Nathan Patwardhan
- *Programming the Perl DBI* Alligator Descartes, Tim Bunce
- *Learning Perl/Tk* Nancy Walsh
- *Object Oriented Perl* Damian Conway (Manning)
- *Tcl and the Tk Toolkit* John K. Ousterhout (Addison-Wesley)
- *Effective Tcl/Tk Programming* Mark Harrison, Michael McLennan (Addison-Wesley)
- *Tcl/Tk in a Nutshell* Paul Raines, Jeff Tranter
- *Exploring Expect* Don Libes
- *Learning Python* Mark Lutz, David Ascher
- *Programming Python* Mark Lutz
- *Python Pocket Reference* Mark Lutz
- *Python Essential Reference* David M. Beazley (New Riders)

Key Web Sites:
www.perl.com
www.cpan.org
www.perl.org
perl.oreilly.com
www.ajubasolutions.com
www.python.org
www.activestate.com
www.masonhq.com
www.pm.org

TCP/IP and the DNS

Many people don't realize that Bill Joy and his team at UC Berkeley originally developed the implementation of the TCP/IP stack that forms the foundation of the Internet, including all commercial versions. The Linux version is derived directly from the BSD code. BIND, the server that implements the domain name system, was designed by Paul Mockapetris and built by Paul Vixie, who still maintains it at the nonprofit Internet Software Consortium. ISC also maintains INN, the most widely used usenet news server software.

Useful Books:
- *Linux Network Administrator's Guide* Olaf Kirch, Terry Dawson
- *TCP/IP Network Administration* Craig Hunt
- *DNS and BIND* Paul Albitz, Cricket Liu
- *Managing Usenet* Henry Spencer, David Lawrence

Key Web Sites:
www.isc.org
www.ietf.org

Unix Command-Line Utilities

Linux includes a full complement of Unix command-line tools, courtesy of the Free Software Foundation's GNU project. What many people don't realize is that the free implementations of many of these tools were developed as part of Berkeley Unix and contributed to the GNU project from there, so the rivalry between Linux and BSD is overshadowed by deeper cooperation.

The *bash* shell is the most widely used command line shell for Linux. There are other versions of the classic Bourne shell, along with *tcsh*, a modern version of the C Shell (*csh*).

Useful Books:
- *Linux in a Nutshell* Ellen Siever, Stephen Spainhour, Jessica Hekman, Stephen Figgins
- *Running Linux* Matt Welsh, Matthias Kalle Dalheimer, Lar Kaufman
- *sed & awk* Dale Dougherty, Arnold Robbins
- *sed & awk Pocket Reference* Arnold Robbins
- *Effective awk Programming* Arnold Robbins (forthcoming from O'Reilly)
- *Learning the bash Shell* Cameron Newham, Bill Rosenblatt
- *Using csh & tsch* Paul DuBois

Key Web Site:
www.fsf.org

Web Server Software

Apache is the dominant web server not only on Linux but on the Web as a whole, with more than 60 percent of all visible web servers running Apache. Apache was created by the Apache Group and is now maintained by the Apache Software Foundation. In addition to the core Apache server project, there are projects for Perl on Apache, Java Server Pages (Jakarta) and XML. mod_perl, mod_serv, and PHP are widely used for generating dynamic content on Apache servers. Developers in the Apache GUI project are working on a cross-platform graphical tool called Comanche (which stands for **co**nfiguration **man**ager for Apa**che**) to help make Apache easier to configure. Zope is the most comprehensive software for web site design and management in the open source world. Using Python-based tools, it combines dynamic content management, shared development, and support for sophisticated management techniques like version control and staging.

Useful Books:
- *Apache: The Definitive Guide* Ben Laurie, Peter Laurie
- *Writing Apache Modules with Perl and C* Lincoln Stein, Doug MacEachern
- *HTML & XHTML: The Definitive Guide* Chuck Musciano, Bill Kennedy
- *JavaScript: The Definitive Guide* David Flanagan
- *Dynamic HTML: The Definitive Reference* Danny Goodman
- *CGI Programming with Perl* Scott Guelich, Shishir Gundavaram, Gunther Birznieks
- *Web Design in a Nutshell* Jennifer Niederst

Key Web Sites:
www.apache.org
www.comanche.org
xml.apache.org
www.apacheweek.com
www.php.net
www.xml.com
www.zope.org
www.w3.org

Kernel and Device Drivers

Linus's kernel is the heart of the system. The current version is 2.2, and is the same on all distributions, although the most recent kernel patch, 2.2.16, is not.

Useful Books:
- *Running Linux* Matt Welsh, Matthias Kalle Dalheimer, Lar Kaufman
- *Learning Red Hat Linux* Bill McCarty
- *Learning Debian GNU/Linux* Bill McCarty
- *Linux Device Drivers* Alessandro Rubini
- *Linux Multimedia Guide* Jeff Tranter
- *Understanding the Linux Kernel* Daniel P. Bovet, Marco Cesati
- *Building Linux Clusters* David HM Spector
- *Linux Application Development* Michael K. Johnson, Eric W. Troan (Addison-Wesley)

Key Web Site:
www.kernel.org

Mozilla

Mozilla, the free version of Netscape's web browser suite, is the one third-party graphical application certain to be on every Linux system. Though some people argue that Mozilla hasn't succeeded as an open source project, we believe its contributions are immense and will become more pervasive in the future. In addition to the browser, the Mozilla project is responsible for the JavaScript language, Bugzilla, and Tinderbox. Parts of Mozilla, including Gecko, are being incorporated into other applications and also into the GNOME 2.0 desktop.

Key Web Sites:
www.mozilla.org
www.mozillazine.org

The X Window System

The X Window System, developed at MIT by Jim Gettys, Bob Scheifler, and a host of contributors, is the foundation of all the Linux graphical user interface tools. The Linux version of X is maintained by the XFree86 project under the leadership of Dirk Hohndel of SuSE. High-level programming toolkits for X include GTK+, Qt,™ and Motif (which has recently been released as OpenMotif®).

Useful Books:
- *Volume 8: X Window System Administrator's Guide* Linda Mui, Eric Pearce
- *Linux X User's Guide* Ellen Siever
- *Programming with Qt* Matthias Kalle Dalheimer
- *The Concise Guide to XFree86 for Linux* Aron Hsiao (Que)

Key Web Sites:
www.xfree86.org
www.x.org
www.opengroup.org/openmotif

GNOME

Miguel de Icaza's GNOME (GNU Object Model Environment) is one of the most popular graphical desktop environments for Linux. Next-generation interfaces based on GNOME are now being developed by companies such as Helix Code and Eazel. GNOME includes a choice of window managers, including Enlightenment, Sawfish, and WM, plus facilities for creating applications with drag-and-drop support, pull-down menus, and other GUI features. GTK (The GIMP Toolkit) is the foundation for the GNOME programming language. GNU/GNOME is the default desktop environment for the Red Hat® and Debian Linux distributions.

Useful Books:
- *Learning Red Hat Linux* Bill McCarty
- *Learning Debian GNU/Linux* Bill McCarty
- *GTK+/Gnome Application Development* Havoc Pennington (New Riders)

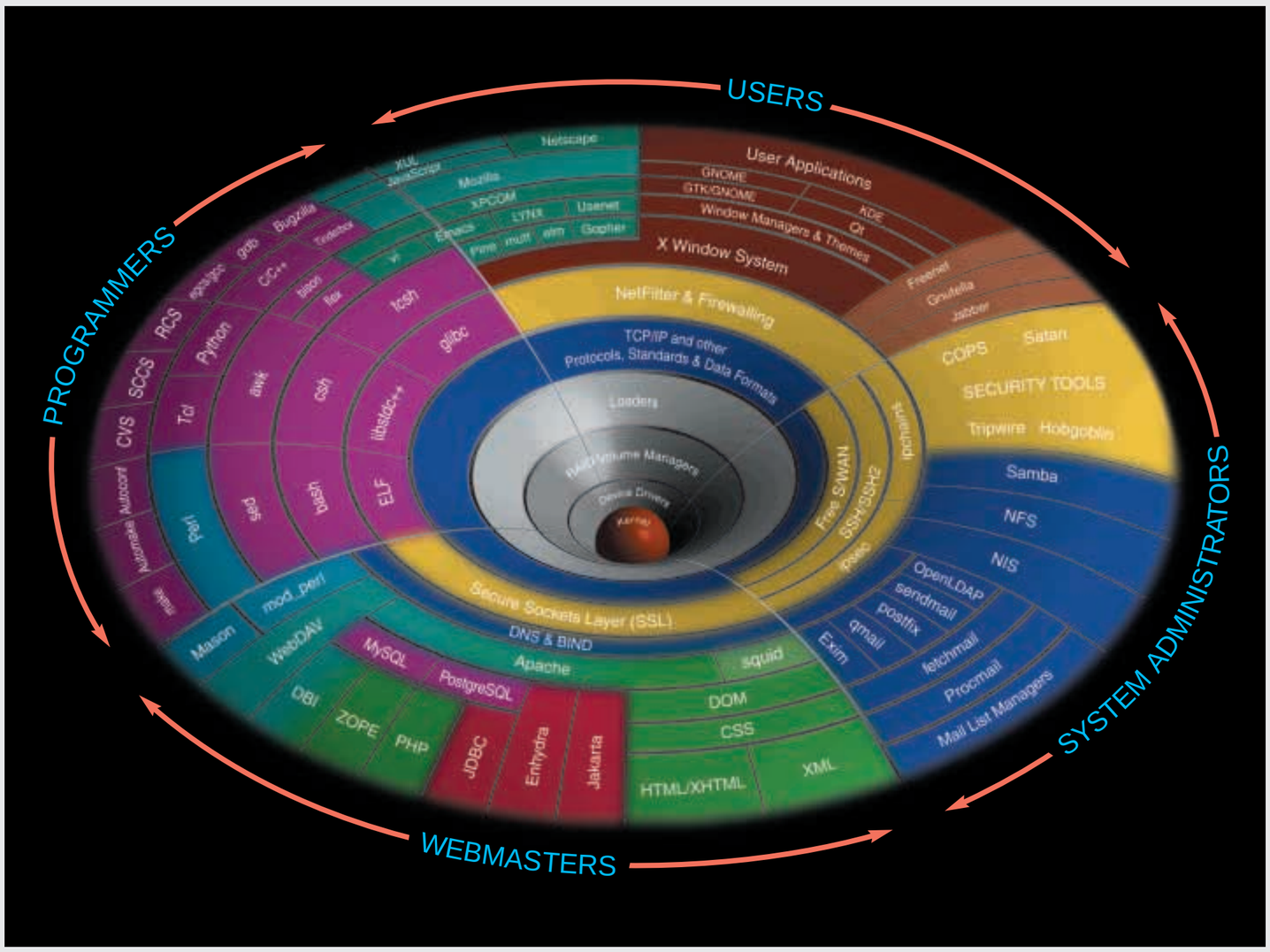
Key Web Sites:
www.gnome.org
developer.gnome.org
www.gtk.org
www.pango.org
www.helixcode.com
www.eazel.com

KDE

The K Desktop Environment (KDE) was the first comprehensive graphical environment for Linux, and is still one of the most popular. KDE is built on top of the Qt™ Toolkit, which is now available under the Q Public License (or QPL). Qt is a product of Norway's Trolltech AS and is the foundation of the KDE desktop. KDE is the preferred environment on SuSE, Mandrake, and Corel Linux. Like GNOME, the KDE Group has developed a suite of office applications called KOffice, which includes a word processor and programs for creating presentations, spreadsheets, illustrations, and much more.

Useful Book:
- *KDE Application Development* Uwe Thiem (MTP)

Key Web Sites:
www.kde.org
developer.kde.org
www.konqueror.org
koffice.kde.org
www.trolltech.com
www.trolltech.com/qpl



Java™

To some, Java and Linux might sound like an odd combination, but in reality, there's a lot going on in this space. Sun has a long-standing link to Linux, supplying developers and source code for the kernel, in addition to their recent contribution to the community by GPLing StarOffice.™ Other Java-Linux connections include the Blackdown Project, which is a world-wide community of volunteer developers who are working to bring the Java platform to Linux, and Kaffe, a cross-platform implementation of the Java Virtual Machine (JVM). Founded by Tim Wilkinson, Kaffe supports Sun's own Solaris, Linux, Microsoft Windows® and Windows CE, and even DOS, and is finding a new home in embedded systems. Other Java-related open source projects include Java Server Pages (JSP/Jakarta), and Enhydra (a Java/XML-based application server and development environment).

Useful Books:
- *Database Programming with JDBC and Java* George Reese
- *Java in a Nutshell* David Flanagan
- *Jini in a Nutshell* Scott Oaks, Henry Wong
- *Java™ Programming on Linux* Nathan Meyers (Waite Group Press)
- *Java and XML* Brett McLaughlin

Key Web Sites:
www.sun.com/linux
www.blackdown.org
www.kaffe.org
jakarta.apache.org
www.openoffice.org

XML and HTML

Tim Berners-Lee's introduction of the HyperText Markup Language (HTML) took the Internet by storm in 1993 when people outside of technical circles realized its potential for spreading information over the then-fledgling World Wide Web (WWW). In the years since, the HTML standard has been through a few iterations, and the Web has evolved into a media-rich environment dominated by things like the Document Object Model (DOM), JavaScript, and other advancements like Cascading Style Sheets (CSS). But HTML isn't flexible enough for today's web, and the eXtensible Markup Language (XML) was at first deemed to be far ahead of itself. That's changed. The new version of HTML, XHTML, builds upon its existing structure by adding some of XML's powerful features. XML is quickly becoming the defacto markup for transporting all kinds of data over the Internet and between applications. Jabber, an open source, peer-to-peer instant messaging system, relies heavily on XML as its transport language, and XSLT gives developers the ability to transform XML documents into output forms such as PostScript, PDF, ASCII text, and HTML.

Useful Books:
- *HTML & XHTML: The Definitive Guide* Chuck Musciano, Bill Kennedy
- *Cascading Style Sheets: The Definitive Guide* Eric A. Meyer
- *Dynamic HTML: The Definitive Reference* Danny Goodman
- *Programming PHP* Rasmus Lerdorf, Randy Jay Yarger, Andi Gutmans, Zeev Suraski, Stig Bakken, Shane Caraveo
- *Web Application Development with PHP 4.0* Tobias Ratschiller, Till Gerken (New Riders)
- *JavaScript: The Definitive Guide* David Flanagan
- *Designing with JavaScript* Nick Heinle, Martin Webb
- *Java and XML* Brett McLaughlin
- *Web Design in a Nutshell* Jennifer Niederst
- *HTML Pocket Reference* Jennifer Niederst
- *JavaScript Pocket Reference* David Flanagan
- *XML Pocket Reference* Robert Eckstein
- *PHP Pocket Reference* Rasmus Lerdorf

Key Web Sites:
www.xml.com
www.ibm.com/developer/sml
www.w3.org
www.webstandards.org
www.irt.org
www.jabber.org
www.php.net

Protocols, Standards, and Data Formats

There is a host of standards and protocols that we all rely on, and we often use them without even giving them a thought. Things like TCP/IP and Ethernet for network communication and data transmission; DHCP for doling out IP addresses within a host network; SMTP, POP3, and IMAP for sending and receiving email; HTTP for the Web, and FTP for transferring files over the Internet; and NNTP for accessing Usenet news. Then there are MIME-types for images, video, audio, and documents. We live and breathe these standards, often without knowing we're doing it. Open protocol standards are a key part of what makes the Internet work.

Useful Books:
- *TCP/IP Network Administration* Craig Hunt
- *Managing IP Networks with Cisco Routers* Scott M. Ballew
- *Managing IMAP* Dianna Mullet, Kevin Mullet
- *Using & Managing PPP* Andrew Sun
- *Internet Core Protocols: The Definitive Guide* Eric Hall
- *Ethernet: The Definitive Guide* Charles E. Spurgeon
- *Managing Usenet* Henry Spencer
- *PNG: The Definitive Guide* Greg Roelofs
- *DocBook: The Definitive Guide* Norman Walsh, Leonard Muellner
- *MP3: The Definitive Guide* Scot Hacker
- *HTTP Pocket Reference* Clinton Wong

Key Web Sites:
www.ietf.org
www.ieee.org
www.openldap.org

Office Applications

Sun's StarOffice and Corel's WordPerfect Office Suite are the most popular third-party office-type applications for Linux. They aren't strictly a part of Linux but are bundled with many distributions or available for download over the Internet. The GNOME team is currently developing Gnumeric, an Excel-like XML-based spreadsheet application, and Evolution, an Outlook Express-type groupware suite. The GNOME and KDE groups are also developing office application suites that will include word processors, presentation software, and image editing/viewing tools. Other companies, including ApplixWare and AbiSource, are also developing office applications for Linux for commercial distribution. The GIMP (GNU Image Manipulation Program) is an open source Photoshop® clone, and Ghostscript is a freely available PostScript interpreter.

Useful Books:
- *GIMP Pocket Reference* Sven Neumann
- *Grokking the GIMP* Carey Bunks (New Riders)

Key Web Sites:
www.openoffice.org
www.corel.com
koffice.kde.org
www.abisource.com
www.vistasource.com
www.gimp.org

Mail Servers and List Managers

Email is one of the must-have applications for any modern system. Users often rely on their ISP to handle the gory details of mail forwarding and delivery, but many administrators need to set up these services themselves. Eric Allman's sendmail,® developed in 1979 as part of Berkeley's Unix, is the granddaddy of Internet mail servers and still the most widely used, but Linux also includes Dan Bernstein's qmail and University of Cambridge's Exim.

One of the earliest forms of peer-to-peer communication over the Internet came in the form of email mailing lists. Whether the discussion takes place on an internal server, or in a public forum, mailing lists bring people with like interests together using list management software such as LISTSERV, Majordomo, Listproc, SmartList, Mailman, and ezmlm.

Useful Books:
- *sendmail* Bryan Costales, Eric Allman
- *Managing Mailing Lists* Alan Schwartz
- *Stopping Spam* Alan Schwartz, Simson Garfinkel

Key Web Sites:
www.sendmail.org
www.sendmail.net
www.sendmail.com
www.qmail.org
www.exim.org
www.postfix.org
LISTSERV: www.lsoft.com
Majordomo: www.greatcircle.com/majordomo
SmartList: www.procmail.org
Mailman: www.list.org
www.ezmlm.org

Samba

Samba allows the Linux or Unix system to act as a file and print server on a Windows® network. It's a high-profile application that is helping to drive Linux's acceptance in a corporate setting. Samba received support from Australian National University, SGI, and Linuxcare.

Useful Books:
- *Using Samba* Robert Eckstein, David Collier-Brown, Peter Kelly
- *Managing NFS and NIS* Hal Stern

Key Web Sites:
Samba Home: www.samba.org
SWAT: anu.samba.org/cgi-bin/swat
KSamba: www.kneschke.de/projekte/ksamba
GnoSamba: www.open-systems.com/gnosamba.html

Peer-to-Peer Communication

Increasingly, instant messaging and other peer-to-peer technologies are looking to be the foundation of the next revolution in Internet technologies. Gnutella and Freenet are peer-to-peer file-sharing tools. Jabber® is an open source instant messaging system with a client-server architecture that allows people to communicate with one another over different IM systems, including AOL's Instant Messenger℠ (AIM) and ICQ.

Useful Book:
- *Peer-to-Peer: The Disruptive Potential Behind Collaborative Networking* Gene Kan, Jeremie Miller (forthcoming from O'Reilly)

Key Web Sites:
www.jabber.org (developers)
www.jabber.com (users)
gnutella.wego.com
freenet.sourceforge.net

Databases

Databases are critical for even small organizations; they store the content for many web sites with interfaces through Perl DBI, PHP, or other languages. The Berkeley DB, originally developed and maintained by Margo Selzer and Keith Bostic, is now maintained by their company, Sleepycat Software. The Berkeley DB is quite flexible, supporting C, C++, Java, Tcl, Perl, and Python APIs. There are also a variety of SQL-based database applications available for Linux, the most popular of which is MySQL, developed by Michael Widenius. Other database projects include DB2 for Linux from IBM, PostgreSQL, and mSQL.

Useful Books:
- *MySQL & mSQL* Randy Jay Yarger, George Reese, Tim King
- *MySQL* Paul DuBois (New Riders)
- *Mastering PostgreSQL* Stephen J. Lombardo (forthcoming from O'Reilly)

Key Web Sites:
Berkeley DB: www.sleepycat.com
www.mysql.com
www.postgresql.org
DB2: www.ibm.com/db2/linux

Security

Linux offers traditional Unix security plus a sophisticated filtering and network address translation (NAT) mechanism. In addition, there are dozens of useful network and security tools in a typical Linux distribution or downloadable from the net.

Network Analysis: tcpd, tcpdmatch, tcpdump, traceroute, Hummer, Snort

Security Auditing: COPS, ISS, SATAN, Tripwire

Firewalls: ipfilter, portmap, socks, wrappers

Encryption Tools: PGP, Kerberos

Remote Access: SSH, Radius

Security Management: swatch, watcher, etc.

Useful Books:
- *Practical Unix & Internet Security* Simson Garfinkel, Gene Spafford
- *Building Internet Firewalls* Elizabeth D. Zwicky, Simon Cooper, D. Brent Chapman
- *SSH, The Secure Shell: The Definitive Guide* Daniel J. Barrett, Richard Silverman
- *Linux® Firewalls* Robert L. Ziegler (New Riders)
- *Firewalls and Internet Security* William R. Cheswick, Steven M. Bellovin (Addison-Wesley)
- *Applied Cryptography* Bruce Schneier (Addison-Wesley)
- *PGP* Simson Garfinkel
- *Maximum Linux Security* Anonymous (SAMS)

Key Web Sites:
CERIAS: www.cerias.purdue.edu
CIAC: ciac.llnl.gov
Gene Spafford's Home Page: www.cerias.purdue.edu/homes/spaf
security.oreilly.com
www.linuxsecurity.com
www.first.org
www.alw.nih.gov/Security
www.rootshell.org
www.securityfocus.com
www.freeswan.org

KEY CONTRIBUTORS:
Linux Kernel: Linus Torvalds, Alan Cox
BIND: Paul Vixie
Exim: Philip Hazel
glibc: Roland McGrath, Ulrich Drepper
GNU tools (Emacs and vision of the free operating system) (gcc) (*bash*): Richard M. Stallman
GNOME: Miguel de Icaza
KDE: Matthias Ettrich, Torben Weis
MySQL: Michael "Monty" Widenius
PHP: Rasmus Lerdorf
Perl: Larry Wall, Chip Salzenburg, Tom Christiansen, Tom Bunce, Gurusamy Sarathy
Python: Guido van Rossum

O'REILLY®
THE O'REILLY OPEN SOURCE SOFTWARE CONVENTION
July 23-26, 2001 • San Diego, California

Qmail: Dan Bernstein
sendmail: Eric Allman
Tcl/Tk: John Ousterhout
XFree86: Dirk Hohndel

Created by Tim O'Reilly and Chuck Toporek with review and input from Andy Oram, Frank Pohlmann, Laurie Petrycki, and attendees of the 2000 Open Source Convention

Designed by Kathryn Heflin and David Bacigalupi with input from the O'Reilly Marketing Design Group
Illustration by Jeff Reynolds Design