

Linux

- **Unix released from Bell Labs in 1971.**
- **Announced by Linus Torvalds on 25 Aug 91 on comp.os.minix newsgroup:**
 - * **“I'm doing a (free) operating system (just a hobby, won't be big and professional like gnu) for 386(486) AT clones.”**
 - * **First release in 9/91**
- **Declared “obsolete” by Andy Tenenbaum in Jan 92**
 - * **“My real job is a professor and researcher in the area of operating systems. As a result of my occupation, I think I know a bit about where operating are going in the next decade or so.”**
 - * **objected to monolithic kernel design**

- **Torvald's polite response**
 - * **“Re 2: your job is being a professor and researcher: That's one hell of a good excuse for some of the brain-damages of minix.” Torvalds later posted an apology.**
- **MINIX, started in 1987, designed as educational system, intentionally small.**
- **Linux was designed as replacement for UNIX on a 386 Intel machine.**

- **Tanenbaum now states:**
 - * **“I would like to take this opportunity to thank Linus for producing it. Before there was Linux there was MINIX, which had a 40,000-person newsgroup, most of whom were sending me email every day.”**
 - * **“My consistent refusal to add all these new features is what inspired Linus to write Linux. Both of us are now happy with the results. The only person who is perhaps not so happy is Bill Gates. I think this is a good thing.”**
- **1994 a NASA team started assembling Beowulf clusters -- high performance parallel processing systems using Linux on multiple PCs**

- **1998 Los Alamos laboratory assembled “Avalon” in 3 days**
 - 68 Dec Alpha processors**
 - * **ranked 315th of top supercomputing sites**
 - * **“The key to the success of these machines lies in their software, and the most important part of that software is the Linux operating system. Linux can be obtained at no cost through the Internet, but that is minor compared to its other advantages. In my experience, the reliability and performance of Linux has no peer. ” -Michael Warren**
- **1999 Red Hat (Linux distributor) went public**

- **Currently, 3rd fastest supercomputer in 1152 nodes running Linux**
 - * **2,304 Intel Xeon CPUs,**
 - * **11.2 Tflops**
Tflop: trillion floating point operations per second
 - * **4.6 TB memory, 138.2 TB disk**
TB: trillion bytes memory
- **Distributions now include**
 - * **Intel, AMD, Motorola 68000, Alpha, ARM, IBM mainframes, Vax, MIPS, Oracle Network Information Computer, HP PA-RISC, PowerPC, Sparc, diskless terminals, various embedded devices, and Compaq iPac handhelds.**
- **Stallman notes most systems are technically “GNU/Linux”**

Requirements

- **4 MB ram minimal system, 32 MB desktop, 100 MB compiling C++ programs**
 - * **special distribution for running in 2 MB**
- **Disk space**
 - * **10 MB disk minimal system**
 - * **Typical server 80 MB**
 - * **15 G for desktop system with office suite**
- **386 or better Intel processor**

Linux features

- **Multithreaded - heavyweight and lightweight threads**
- **Multitasking**
 - * **priority based round robin scheduler**
 - * **some support for realtime**
 - * **supports CPU affinity**
- **Virtual memory**
 - * **Processes have linear address space**
 - * **Three-level page table used**
 - * **Demand paging from disk**
 - * **Copy on write**
 - * **Configurable page size (4K typical)**

- **Supports FAT, VFAT, HPFS, Amiga file systems**
 - * **NTFS support under development (read part works)**
- **network support**
- **multiple device drivers, include some USB**
- **Can choose which modules to include when kernel compiled**

Distributions

- **Red Hat probably most common**
- **Others include Debian, Mandrake, Slackware**
- **Knoppix has version which runs directly from CD drive**
- **Many, many others**

Linux development

- **Linux is written in C / assembler**
 - * **Uses the Gnu C compiler (gcc)**
 - * **Some object techniques used**
 - * **Key developers don't see any advantage of switching to C++**
 - require porting large code base**
 - knowledge base of many developers is C, not C++**
- **Kernel modifications written in terms of patches to official distribution**

- **Typically submitted to linux-kernel mailing list**
 - * **cc'd to appropriate kernel developer**
 - * **FAQ indicates they're pretty picky**
- **High activity list**
 - * **1290 posts in week ending 31 July**
 - * **Torvalds made one post replying to a private email**
- **Linux source at *<http://www.kernel.org>***

The Penguin



- **Named "Tux"**
- **Torvalds made a trip, saw a penguin**
 - * **It bit him.**

Installing Linux

- **Key options are whether to use dedicated machine or disk drives for Linux**
- **Possible to configure machine for dual boot with Windows**
- **Key issue is NTFS file system**
 - * **Optional for NT, 2000.**
 - * **Also used in XP (required?)**
 - * **NTFS preferred on Windows machine**
- **NTFS is read-only from Linux**
 - * **Requires a separate partition or disk drive on NTFS machine**

- **A separate machine obviously preferred for a server**
 - * **Can telnet in from Windows machine**
- **Can download and install yourself**
- **Can configure kernel to**
 - * **increase security**
 - * **tailor to minimize footprint, configure for specific use**
 - * **play around with operating system (e.g. CSCI6140 Operating System course)**
- **Initial install most commonly from CD from Red Hat, et. al.**

Using Linux

- **Just as easy as using Unix, or just as hard as using Unix**
- **Console oriented system**
- **Choice of command line interfaces**
 - * **bash - born again shell -- enhanced Bourne shell**
 - * **ksh - Korn shell**
 - * **tcsh - version of csh (C-like shell)**
 - * **zsh - enhanced Korn shell**
- **environment is set in various *filename* files. e.g. .profile, .cshrc, .login et. al.**

Linux file system

- **to Linux, pretty much everything is a file**
- **Linux file systems supports multiple names, or links, to a file**
 - * **“deleting” a file removes the link**
 - * **file actually removed when all links removed**
- **Linux uses / for path separator**
- **File system looks like one tree**
 - * **Head of the tree is /**
 - * **multiple physical devices or network drives may be mounted, or added to tree**
 - * **Can add anywhere in tree**
 - * **Use *mount* to get listing**

- **Linux permissions**
 - * **Group of three values**
owner, group, all
 - * **permissions are read, write, and/or execute**
 - * **for directory:**
 - read means list**
 - write means create/delete files**
 - execute means cd into directory and access by name**

- **Use *chmod* command to alter**
 - * **symbolic form: use o, g or a (owner, group or all)**
 - * **+perm adds permission, -perm removes**
 - * **chmod a+x *name* adds execute to all**
 - * **chmod g-r *name* removes read from group**
- **chmod octal mode**
 - * **uses three octal numbers for owner / group / all**
 - * **bit 1 is execute, 2 is write, 4 is read**
 - * **chmod 755 *name* -> rwxr-xr-x**

Common Linux commands

- ***ls*** - performs a listing
 - * **-a** (all) lists files including those starting with a .
 - * **-d** shows just directory name
 - * **-l** long listing: permissions, owner, modification times
 - * **-t** sort by modified time
- ***cd*** change directory
- ***mkdir*** make directory
- ***rm*** remove a file
 - * **-f** (force) remove readonly files
 - * **-r** (recursive) remove directories

- ***cat*** (concatenate) shows a file
 - * copying a file from a name to the console
 - * can also be used to combine files
- **editing**
 - * **vi** - visual interactive -- great editor after the first six months
 - modal -- typing mode / editing mode
 - * **emacs** - editing macros - another great editor (or so I'm told)
 - Uses control sequences to control
 - Extensive macro compatibility / lisp built in
 - Power users use it as complete development / work environment

* **pico - simpler editor**

Used in pine mail program

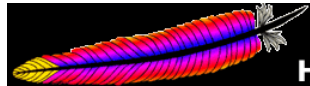
Control commands listed at bottom of screen

* **X-windows editors are also available**

Linux information

- ***sysconf* system library call can provide information about current Linux system**
- **See example program**

Apache



- **In 1994 most common HTTP server was HTTP daemon written by Rob McCool at NCSA (National Center for Super-computing Applications)**
 - * **Development stalled when McCool left NCSA**
 - * **Various webmasters made fixes, or patches, to HTTP daemon**
- **8 primary and 3 additional folks got together via email to combine patches**
 - * **Name is from “A patchy” server (feather came later)**
 - * **First release April 95**
 - * **Became most common server within a year**

- **Currently has about 64% of server market, according to netcraft.com** http://news.netcraft.com/archives/web_server_survey.html
 - * **Microsoft next at 24%**
- **Apache Software group formed in 1999**
 - * **Core members approved / reject changes**
requires 3 positive votes and no negative votes
 - * **Also nominate / approve new core members**
- **Apache 2.0**
 - * **Redesign / rearchitecture**
 - * **Discussed back in 96. Formal planning starting in 98**
 - * **Beta release March 2001**
 - * **General release Apr 2002**

Other Apache projects

- **ant - automated build system (like *make*, et. al.). Uses XML for directive files**
- **php**
- **tcl - supports mod_tcl. Tcl is tool command language, an open source scripting language**
- **perl - supports mod_perl. Runs perl within Apache process space and allows Perl access to Apache interiors**
- **web services - includes implementation of SOAP (Simple Object Access Protocol). Object oriented, XML based remote procedure call mechanism**
- **XML - xml parsers/generators for multiple languages**

- **Jakarta - open source Java projects**
 - * **tomcat - reference implementation for servlets/Java Server Pages (JSP)**
 - * **taglibs - JSP tag library**
 - * **struts - model/view controller framework for servlets/JSP**
 - * **ORO - Java library for Perl like regular expression processing**
 - * **POI - Write Microsoft compound documents using Java. (xls format implemented)**
- **Web site is <http://www.apache.org>**

Apache HTTP Server features

- **Serves up web pages**
- **Include data from other pages in returned pages**
- **Executes CGI (common gateway interface) scripts**
- **Multiple languages support**
- **Virtual hosting -- support multiple domain names from same machine**
- **Authentication**
 - * **local password files, dbm or Berkeley db type files**
 - * **can allow / deny based on client's IP address or name**
- **Logging**
 - * **custom formatting available**

- **Custom error handling**
 - * **Specify pages to be shown for the various errors**
- **Automatic indexing of directories**
 - * **Customizable**
- **Extendable via module system**

Apache 2.0 features

- **Hybrid multiprocess / multithreaded support**
- **Supports multiple protocols (port numbers)**
- **Revamped to support non Unix platforms better**
- **Revamped API for modules**
 - * **Simplify dependency issues between modules**
 - * **Modules can act as filters or translators**
- **Support for IPv6**
- **Error messages in multiple languages**
- **Updated regular expression processing to use Perl 5 syntax**

Apache architecture

- **Main server process listens for requests**
 - * **On Unix process must be root to access port 80**
 - * **On Windows NT/2000/XP runs as *LocalSystem* account**
 - Significant access to local machine, no network access by default**
 - Can configure service to run as less privileged account**
- **Child processes are spawned to service requests**
 - * **On Unix, run as accounts specified in *User* and *Group* directives**
- **Typically put in UNIX startup file, */etc/rcxx***

- **On Windows NT/2000/XP, run as system service**
- **To control after login, use *apache -k start|stop|restart|graceful***
 - * **graceful allows child processes servicing requests to exit**
 - * **stop / restart terminates immediately**
- **Apache must be restarted for changes in configuration to take effect**

Apache configuration

- **Main control file is *httpd.conf***
- **Location compiled in, or may be passed to apache upon start-up**
 - * **Include directive allows configuration to be broken into multiple files**
- ***ServerType* controls whether server is standalone or inetd based**
- ***PidFile* specifies file Apache uses to store process id of current server**
- ***Timeout* specifies seconds until server gives up on client. e.g. 300**

- ***KeepAlive* (on/off) controls whether persistent connections are allowed e.g. On**
 - * **faster throughput between client and server**
 - * **requires resources to keep connection open**
- ***KeepAliveTimeout* number of seconds idle before keep alive connections closed e.g. 15**
- ***MaxRequestsPerChild* number of requests 1 child process allowed to handle. 0 means unlimited. Used on systems where there are memory leaks in libraries. e.g. 0**
- ***ThreadsPerChild* number of threads child process is allowed to run. e.g. 50 More threads means**
 - * **greater responsiveness to clients**
 - * **more resources consumed**

- ***AddModule*** directives specify which modules should dynamically added to server. e.g. `AddModule mod_perl.c`
 - * Many modules typically compiled into server. These don't require a `AddModule` directive
- ***Port*** specifies listening port. e.g. `80`
- ***ServerAdmin*** specifies email address for Apache generated responses
- ***ServerName*** name or IP address of web server. e.g. `65.113.124.81`
- ***BrowserMatch*** specifies specific responses for certain browsers e.g.

`BrowserMatch "MSIE 4\0b2;" nokeepalive downgrade-1.0 force-response-1.0`

- ***Directory*** multiple line directive used to configure access to files

```
<Directory />
  Order Deny,Allow
  Deny from All
  AllowOverride None
</Directory>
```

- * ***Order*** specifies processing of `Deny` and `Allow` directives
- * ***Deny*** lists clients denied access
- * ***AllowOverride*** (`All|None|directive`) determines whether `.htaccess` files can be used on a directory by directory basis to alter access

Allowing hinders performance because must look in directory of html file for a `.htaccess` file

- **Directory *Options* directive includes**
 - * **All (everything except MultiViews)**
 - * **ExecCGI - allows execution of CGI scripts**
 - * **FollowSymLinks - follows symbolic links**
 - * **Includes - allow server side includes**
file which typically ends in.shtml
allows directives like

```
<!--#include virtual="/footer.html" -->
```

```
This document last modified <!--#lastmod file="index.html" -->
```

- * **IncludesNOEXEC - includes can't execute programs**
- * **Indexes - generates directory listing if no index file (e.g. index.html) present**
- * **MultiViews - allows multiple (language specific) versions of page to be served up based on browser request**

- ***Files* - similar to Directory, but based on filename instead of path. Following example prohibits files starting with .ht to be served.**

```
<Files ~ "^\.ht">
```

```
Order allow,deny
```

```
Deny from all
```

```
</Files>
```

- ***DirectoryIndex* - name of file generated for directory views e.g. index.html**
- ***AccessFileName* - name of file used for access file. Typically .htaccess**
- ***ErrorLog* - name of file used for error messages. e.g. logs/error.log**

- ***HostNameLookups (on/off)*** - whether server should look up domain names of IP addresses for logging. e.g. off
 - * Turning on hinders performance because server has to query DNS server for all client requests
- ***LogFormat*** assigns a name to a specific logging format

LogFormat "%h %l %u %t \"%r\" %>s %b" common

%h - remote host

%l - remote logname

%u - remote username (if logged in)

%t - time

%r - first line of HTTP request

%>s - status

%b bytes sent

- ***CustomLog*** - specifies log name and format to use.
 - e.g. logs/pto.log common

* Sample log entry:

68.64.87.94 - - [04/Aug/2003:11:39:37 -0400] "GET /bowplatform/thbulwagga.jpg HTTP/1.1" 304 -

* Sample error log:

[Fri Aug 01 23:10:27 2003] [error] [client 213.156.39.4] File does not exist: d:/web/catsmeow/formmail.php

[Sun Aug 03 21:27:16 2003] [info] [client 63.196.198.171] (10054)Unknown error: client stopped connection before send body completed

- ***ServerSignature*** - (on|off) whether server will report Apache and Operating System version info

- **Indexing directives**

- * *IndexOptions FancyIndexing*

- * *AddIconByType (TXT,/icons/text.gif) text/**

- Specify icon to use by type of file

- * *AddIcon /icons/binary.gif .bin .exe*

- Specify icon to use by file extension

- * *IndexIgnore .??.* *~ *# HEADER* RCS CVS *,v *,t*

- Specify files not to list

Virtual Hosting

- **Virtual Hosting allows the same Apache server to service several different websites.**

- www.rockyhillteach.org and www.charlesconsulting.com are hosted by same server**

- **Saves on hardware and IP addresses**

- **Example specification**

```
<VirtualHost *>  
  ServerAdmin webmaster@rockyhillteach.org  
  ServerName www.rockyhillteach.org  
  DocumentRoot "g:/rhhs/web"  
  ErrorLog logs/rockyhillteach.log  
  CustomLog logs/rockyhillteach.log common  
  ScriptAlias /scripts g:/rhhs/web/scripts  
  ErrorDocument 500 /serverError.html  
  ErrorDocument 404 /notFoundError.html  
  ErrorDocument 403 /forbidden.html  
  <Directory "g:/rhhs/web">  
    Options Indexes FollowSymLinks MultiViews  
    Order allow,deny  
    Allow from all  
  </Directory>
```

```
<Directory "g:/rhhs/web/private">  
  AuthType Basic  
  AuthName "Access prohibited"  
  AuthUserFile c:/passworfileword.txt  
  Require user gerardw  
</Directory>  
</VirtualHost>
```

Performance tuning

<http://httpd.apache.org/docs/misc/perf-tuning.html>

- **Ensure sufficient RAM so child processes don't swap.**
 - * **Add more RAM**
 - * **Set *MaxClients***
- **Decent CPU, NIC, Internet connection, disks**
- **Install latest OS system patches**
- **Set *FollowSymLinks* on and *FollowSymLinksIfOwnerMatch* off**
- ***AllowOverride* None**
- **Use *mod_perl*, *mod_tcl* to speed up scripts**

- **Set *MaxRequestsPerChild* high (10,000) for static pages, lower if using *mod_perl* (30)**