Chapter 20
Web and Database Systems

Abstract In this chapter we discuss the use of open-source software to implement Internet software including web servers, virtualization and databases.

Contents

20.1 Web Servers ................................................................. 455
20.2 Hadoop ................................................................. 457
20.3 Content Management with Joomla ......................... 457
20.4 Virtualization and Cloud Computing .................... 458
20.5 PostgreSQL ............................................................ 460
20.6 MySQL ................................................................. 464
20.7 SQLite ................................................................. 467
20.8 CouchDB ............................................................... 469
20.9 Conclusion .............................................................. 471

20.1 Web Servers

The Internet has transformed the way information travels across the world, and it has indeed impacted the manner in which information is disseminated, shared, and accessed. Internet based technologies are also experiencing some of the most rapid advances in technology since the space age, and it is not foolhardy to think that this rapid rate of innovation will have significant impact on the disciplines of scientific computing and research. In this chapter we provide an overview of the current open-source solutions for deploying data and computation on the Web. If anything in this book is sure to get obsolete quickly, I fear, this chapter is on the top of the stack, and thus we cover only the most commonly used web technology, or the most promising “new thing”, what we expect will last for a technological node or so.
20.1.1 HTTP Server: Apache

Apache, is a web server software, which (according to some estimates) is responsible for serving more than 60% of the Internet as of 2010. In sheer volume, Apache is said to server more than 100 million websites on the Internet. Apache supports a number of features, including: (i) compiled module support, (ii) compression support, (iii) virtual hosting, and (iv) high performance. The Apache binary is called 'httpd' and runs as a daemon or service on GNU/Linux.

20.1.2 YAWS: Yet Another Web Server

Yaws is written in Erlang (see Section 1.6.10), and is a robust, efficient and fault-tolerant web server. It is highly scalable, in an experiment Yaws was able to sustain operations during a DDOS (distributed denial of service) simulation, serving 80,000 page requests. It is designed for dynamic web content generation, with arbitrary Erlang code inserted in Web pages, the Yaws framework performs runtime substitution of the Erlang function’s output into the web page.

Yaws also has an integrated web applications framework which can be used to deploy applications written in Erlang. The full power of the Erlang language, Mnesia database is available in Yaws.

20.1.3 LAMP Stack

LAMP (Linux, Apache, MySQL, and PHP) is an acronym of the above mentioned tools which are most often associated with a software solution stack to build a general purpose web server and hosting solution. GNU/Linux has already been covered in this book (see Section I), and above we discussed the Apache web server. MySQL is discussed in Section 20.6. Although PHP is a programming language we did not present it in the context of other programming languages as it is mostly (only?) used in the context of web application development. Some have suggested replacing the P in LAMP by Perl and Python as these languages are also used for web development.

20.1.3.1 PHP

PHP: hypertext processor is a scripting language designed for web development, including dynamic web content generation, and user interaction and validation. PHP code is embedded inside HTML documents and is executed by the browser with the help of a PHP module.