



Open source platform and sustainability

Punam Gupta¹ and Sapna Kapoor^{1*}

ABSTRACT

Technology is necessary for development as it makes people have access to computer resources. It is not the software that does the development or that gives the access to knowledge rather it is a tool that makes the access to knowledge and resources much easier. By buying software, one gets linked to and becomes dependant on its developers. As the software evolutes, one has to buy upgrades, patches or new versions. And there may be a situation one would have to buy different software, which is often not compatible with the previous one. To reach sustainability, whether in an open source and closed source the project must make one of its initial objectives of implementing sustainability plan at a very early stage of the project's life. For developing countries like ours open standards offer clear advantages as compared to proprietary solutions. Government may play an important role in software development by bringing in the standards based on open source platforms. The happy ending success story of migrating RojgarWahini, from DB2 7.2 to PostgreSQL 8.1.4 (open source) is a live example of moving from proprietary software to Open Source platform for sustainability

Keywords: Open Source, Sustainability, RojgarWahini, Portal built on Open Source

1. Introduction

Open Source software generally is distributed under a license that guarantees the right to read, modify, redistribute and use the software freely. Open source software may be developed by community of programmers interested in developing a software application for a specific purpose. Companies may also develop open source software. These companies will distribute their software for free and make their money from support contracts and customized development. Much of the open source software is distributed under the GNU General Public License (GNU GPL). The GNU GPL allows you to copy, use, modify, re-distribute the software but prohibits companies or individuals from making modified versions proprietary. Richard Stallman, a McArthur genius award recipient, developed this license, in order to encourage the development of a software sharing community. Tens of thousands of developers and several large corporations such as IBM, Sun Microsystems, and Intel have chosen to participate in the open source software movement. Some of most successful and robust software on the web today has been developed under this license

Open Source Software for the Web

The web is dominated by open source software solutions.

¹ National Informatics Centre, Pune, India
Corresponding Author: (Email: Sapna.kapoor@nic.in)

- Over 58% of the web is using the Apache Web Server as compared to 28% for Microsoft's IIS Web Server.
- Linux and Free BSD, both open source flavors of UNIX, are the dominant operating system for web servers, not Microsoft Windows 2000 or Windows NT. Linux has a 34% market share for web server operating systems. Microsoft Windows has a 23% market share (<http://www.linuxtoday.com>)
- The most widely used language for web programming is Perl, not Microsoft's Active Server Pages. Perl has been often referred to as the "glue" that holds the Internet together.
- Over 60% to 80% of e-mail travels across the Internet using the open source program, Send Mail.
- Etoys.com was the third busiest e-commerce site during the 2000 Christmas season. They served over 2.5 million page views and processed 20,000 orders per hour. Etoys was built using Perl running the Apache server under the Linux operating system, all open source software tools. Yan et. al. (2006)

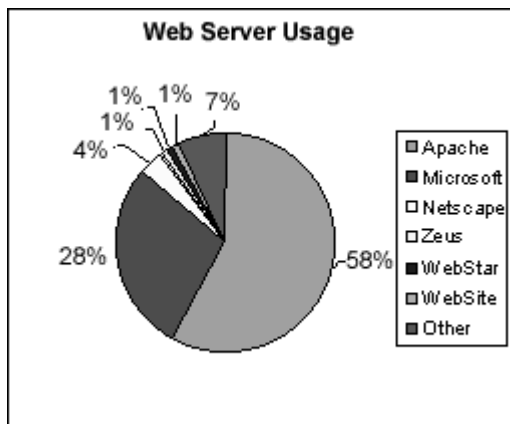


Figure 1: Web Server Usage at a glance

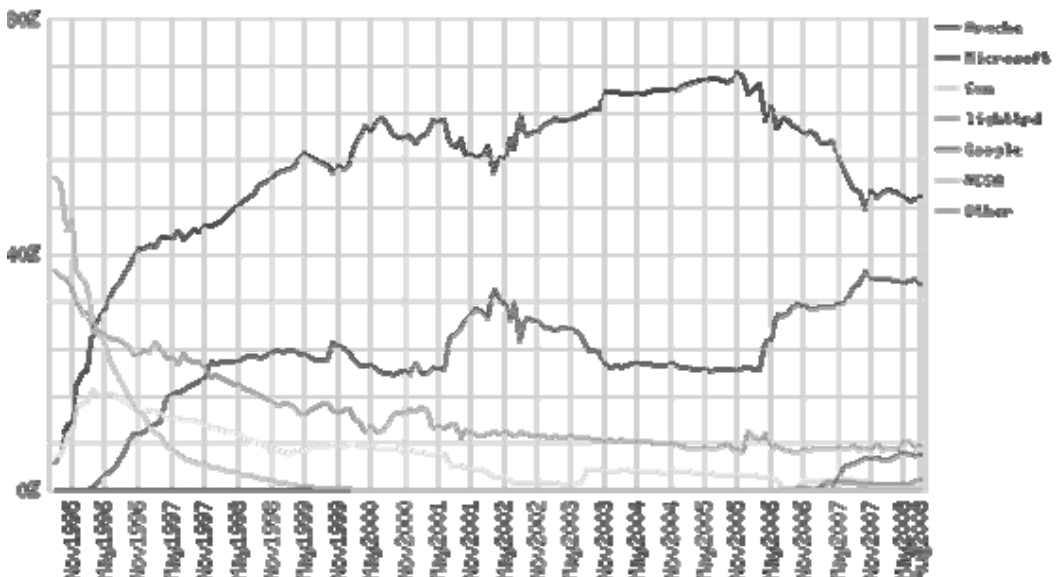


Figure 2: Market Share for Top Servers across All Domains August 1995 - August 2008

Web Server Survey

In the August 2008 survey from 176,748,506 sites showed overall growth of 1.3 million sites. It reflects Apache's growth of 1.2 million and Google's gain of half a million sites, but a loss of 760 thousand sites using Microsoft IIS. (<http://www.netcraft.com>)

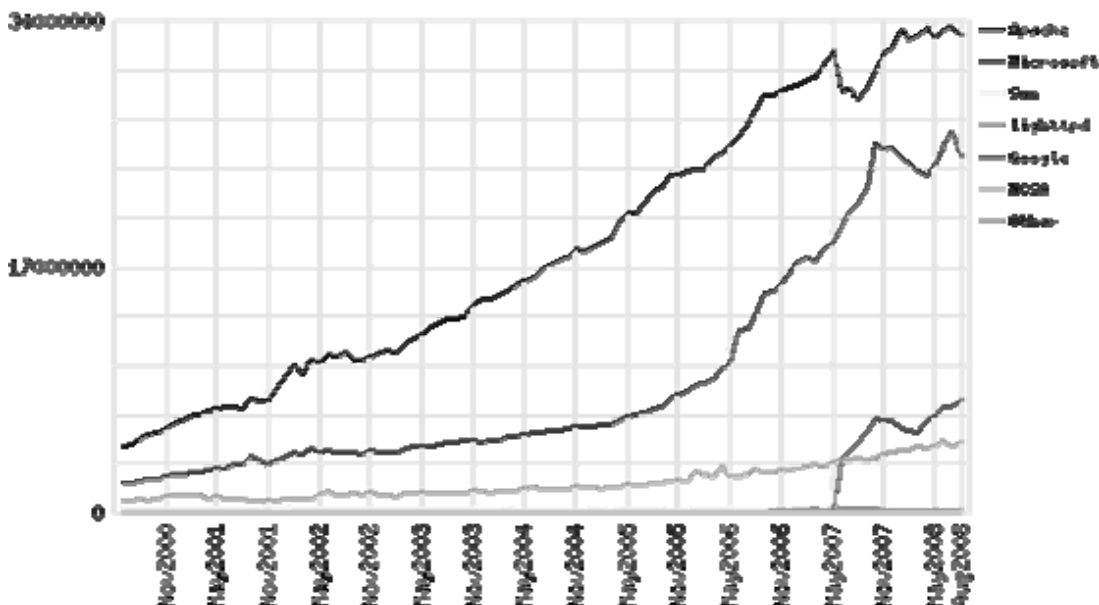


Figure 3: Totals for Active Servers across All Domains June 2000 - August 2008
Source: Gartner (<http://news.zdnet.co.uk>)

In a few years' time, almost all businesses will use open source, according to Gartner; even though IT managers may be unaware of it, and prefer to talk about fashions such as software as a service (SaaS). "By 2012, more than 90 percent of enterprises will use open source in direct or embedded forms," predicts a Gartner report, *The State of Open Source 2008*, which sees a "stealth" impact for the technology in embedded form: "Users who reject open source for technical, legal or business reasons might find themselves unintentionally using open source despite their opposition." Open-source promoters have welcomed the endorsement by what is seen as a conservative commentator, but predict the changes will go further than Gartner assumes.

2. Issues with open source software development

Despite the growing success of the Open Source movement, most of the general public continues to feel that Open Source software is inaccessible to them. However, Open Source technology continues to remain foreign to the large majority of computer users. In order to accommodate these users, the following five issues must be seriously addressed and actively resolved. Due to the increases in Open Source usage, these changes should take place as soon as possible. They should be resolved before those sampling what Open Source has to offer decide to switch back. (Woods and Guliani, 2005)

Some most important flaws with Open Source software development to be as follows:

- *User interface design:* If the Open Source community wishes to truly prosper and have their tools used by the general public, it is fundamentally necessary for them to recognize that the majority of

the users will never know that they happened to invent a particularly clever algorithm for synchronizing the multi-threaded editing of their complex data structure. What the user will see and what they'll judge the project based on is the user interface. If it's inadequate, no one outside of other geeks will touch the program.

- *Documentation:* Open Source projects tend to have a major problem with providing decent documentation. Because they don't have a contractual responsibility to provide this documentation, it's usually intended to be a general guide rather than a complete manual that you could hand to a novice. If we can't understand it, we're not ready to install it, but then how are they expected to learn? Documentation should always cater to the lowest common denominator. Without adequate documentation, Open Source projects are inherently at a disadvantage.
- *Feature-centric development or Programming for the self :* The result is that Open Source projects are made by programmers for programmers, who then can't understand why the general public would bother with proprietary software when this Open Source tool is working so well for them. Meanwhile, the rest of the world begins to associate "Open Source" with software that's only accessible to the technocratic elite. Before the start of every Open Source programming project, a conscious decision should be reached about whether this project's target audience is other programmers or the general public. If it's the latter, there should be a regular effort to ensure that all elements of the project are accessible to this target audience. Programming for the self is an easy trap to fall into, but one that needs to be avoided at all costs when it's not applicable.
- *Learning what proprietary software:* While this has the advantage of increasing Open Source software usage amongst programmers themselves, unfortunately it has the side effect of preventing the Open Source community from learning what proprietary software has to teach. Concepts invented in the world of proprietary software are automatically rejected on the assumption that there's nothing that could possibly be learned from those who are competing with their movement.

3. Case Study of RojgarWahini

The web portal RojgarWahini, (<http://ese.mah.nic.in>) has been developed for the Department of Employment and Self Employment (DE&SE), Government of Maharashtra. DE&SE is a state govt. organization providing free services like vocational guidance, job opportunities and self employment guidance to the job seekers. It also collects, compiles and provides the statistical information to the planning commissions and planning corporations to be used for manpower planning. It functions as per the CNV Act 1959. The portal design has six major sub-sites, Candidates' Corner, Employers' Corner, Self-Employment, About Us, Right to Information (RTI) and Kamgar Katta. It is a single point contact for the services provided by the Department. Data of 'RojgarWahini' is stored in 500 tables. It has accumulated data of all 35 lakh candidates registered across 45 Employment Exchanges. Initially, the portal was designed with PHP as front end and DB2 Ver 7.2 as backend database on the Linux platform.

Platforms are Used By NIC for RojgarWahini

NIC has chosen all Open Source Software development tools (LAMP) to develop the best and most cost effective web site designs for DE&SE. Of course, there are plenty of excellent open source variants for any of the pieces of LAMP. The L stand for Linux, FreeBSD, NetBSD, OpenBSD, and Darwin/Mac OS X, all of which are open source operating systems and all but the latter have open source GUI layers. A stands for Apache web server. Let the M stand for MySQL and PostgreSQL. Let the P stand for PHP, Perl, Python, and Ruby.

- *Web Site Development Tools:* We used Macromedia Dreamweaver, EditPlus and Flash. Dreamweaver is used by over 70% of web design professionals. (<http://www.wscsd.org/2004/01/04/open-source-software-fair-trade-for-software/>)

- **Web Programming Language:** We used PHP for Server Side application development, as PHP is an open source solution built from the ground up specifically for web applications. PHP enjoys a large developer's community. PHP runs under Linux and can be compiled as an Apache module which makes PHP very fast. PHP is one of the fastest growing languages on the web. Over 5 million domains (about 800,000 IP addresses) are running PHP (<http://www.php.net/usage.php>). We choose PHP because it has the flexibility of Perl but is built from the ground up for web application development. PHP is fast, robust and scalable.
- **Database (Migration from DB2 7.2 to PostgreSQL):** RojgarWahini was developed with DB2 as database at the backend. DB2 was chosen as per the policy decision of Department of Information Technology, Government Of Maharashtra. Initially, the server was stable, but as the volume of data grew the Server became unstable. Problems like stopping of DB2 service, tables getting corrupted, refusal of connection were faced. All means were tried to make the server stable but there was no success. It included changing hardware configuration, tuning of parameters, loading less data. There was no online and offline support for DB2 7.2 available to bail us out of this difficult situation. Finally, we could not do the diagnosis and analysis of the reasons for this instability. As DB2 is proprietary RDBMS, and the only solution suggested was to Upgrade to the latest version of the database. But upgradation could not be done because of the following two reason.
- Lack of Support for DB2 7.2 under REDHAT Linux and
- Exorbitant cost of upgradation of DB2 from 7.2 to 9.x: Then, NIC did Proof of Concept for porting of database from DB2 to an open source RDBMS. The choice PostgreSQL was obvious as it is an open source high performance RDMS which is widely used on the web and has Interfaces (APIs) to most programming languages including C++, Java, PHP, Perl as well as ODBC.

Change over from DB2 7.2 to PostgreSQL

Issues involve into change over from DB2 7.2 to PostgreSQL are as under

- There was no migration tool available which could port data directly from DB2 7.2 to PostgreSQL.
- The Part of data which was in Devnagari script could not be ported directly to PostgreSQL as the IXF format of data in DB2 (for devnagari) was not recognized by PostgreSQL during importing.
- The Date time format in DB2 is different from PostgreSQL, so the direct porting of data was not possible.
- Some tools had provision of exporting data form DB2 8.0 to PostgreSQL, but the department did not have DB2 8.0 version.

After doing extensive search of different ways of porting the data, the following methodology was adopted

- AquaStudio 4.7.2, a freeware tool was used to port Devnagari data directly to PostgreSQL. Prior to this DB2 7.2 data was migrated to DB2 8.0. This migration was done on the DB28.0 of some other Department.
- Exported around 400 tables from DB2 and then imported them to PostgreSQL using Aqua Studio. The date time format was masked as per DB2 format as shown in Figure 4.

Tuning of the database, webserver and the operating system parameters:

Since all three belonged to open source community there was lot of documentation available on the net to fine tune them. The calculations of various parameters were done depending on the hardware configuration of the server. It was found that the server improved the performance when we were able to tune the following critical parameters.

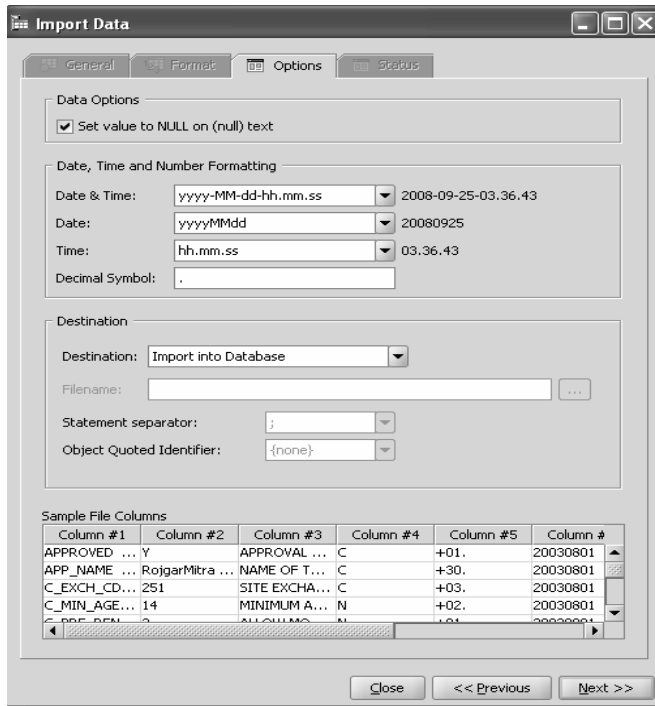


Figure 4 : Porting of exported data from DB2 to PostgreSQL after masking date time format

Table 1 : Values of Parameters for tuning of PostgreSQL

Parameter	Value
Max_connection	100
Shared_buffer	25000 (195 MB)
work_mem	2048 K
Maintenance_memory	128384 (125 MB)
max_fsm_pages	20000
Max_fsm_relations	1000

- Hardware Configuration
 - HP RX 2620 Itanium 64 bit Server
 - HP Dual Core 1.6 GHZ Itan2 Processor
 - 3 X 300 GB Hard disk with Raid 5 Technology
 - 16 GB RAM
 - DAT Drive Capacity – 72 GB
- Software Installed
 - Red Hat Linux Version 5.0
 - PostgreSQL Version 8.1.4
 - PHP Version 5.1.6
 - Apache 2.0 Web Server

4. Road Ahead

Today the application is built on open source platform. There are no issues of interoperability and porting

of the data. The department has no headache buying upgrades, patches or new versions of the front end and backend. In case while implementing this system for some other state/location, if the data needs to be ported to some proprietary RDBMS there are plenty of free tools available to do this task. In terms of raw speed, PostgreSQL benchmarks faster than many other databases such as Microsoft's SQL Server and performs favorably against industry heavyweights such as Oracle. PostgreSQL is fast because it was designed primarily as a web based relational database management system. Its features like legendary reliability and stability, scalability and extensibility, cross platform adaptability and easy administration made the life of developers and database administrator easy.

Benefits of using Open Source

Cost: Open Source software available under the GNU GPL license is free. In some cases, you may choose to pay for the distribution. The cost of the distribution is generally trivial compared to the cost of many enterprise level commercial offerings. In addition, the developers of many of the open source solutions offer support contracts that are suitable to all levels of business or organization.

Software Source Code: When one purchases a license to use a commercial software, then one is dependent on the software designer to add features or customize the software for the needs of business or organization. The software manufacturer provides only the executable program and not the source code. With open source software, one is free to modify the software and customize it in order to suit your application.

Scalability and Robustness: A large community of highly skilled software developers has created open source solutions, such as Linux, Perl, and Apache. Open source UNIX based operating systems such as Linux and FreeBSD are extremely robust and efficient as they are suitable for both small and large organizations. Open source software is used across a full spectrum of web sites.

Large Support Community: A large community of developers that communicate through on-line discussion groups supports many open source offerings. This allows common problems to be easily solved and bugs to be quickly exposed and fixed.

Security and Protection of Proprietary Data: There is a myth that open source software is more vulnerable to attack than proprietary solutions. Actually, the opposite is often true. Because the source code is exposed, it is often easier for a security minded software community to close security holes or breaches.

5. Concluding Remarks

RojgarWahini is built on Open Source technology. The system is stable and is handling large volume of data. The Department of Employment & SelfEmployment need not buy any licences or worry about the costly upgradation charges. Though the journey of migration to open source was not very simple but now we have a very stable, cost effective, and secure system in place. There are no doubts about the sustainability of the system as it has following three pillars to ensure that there will be no looking back, the pillars are Use of Open Source Technology, Unicode based and Data base driven solution. Now, NIC as an organisation has built a helpdesks for open source to help it's officers to start working on Open Source or migrate to Open Source. Team members of RojgarWahini are key members of this helpdesk to promote open source technology.

Acknowledgement: We would like to thank Dr B. K. Gairola, Director General, NIC to give us the opportunity to work on this project. We also want to thank the Mrs. P. P. Joag, DDG, NIC, whose help, stimulating suggestions and encouragement help us in doing innovative works and do things differently. We want to thank her for all her help, support, interest and valuable hints. We are obliged to our colleagues from Appln Group 8, NIC, Pune, who supported us in this work. We would like to give our special thanks to Mr. Gorakh Megh, Principal Secretary, Department of Employment and Self-Employment for encouraging us and letting us use the project information and data.

References

1. Boulanger, A. (2005), Open-source versus proprietary software: Is one more reliable and secure than the other?
2. N. Yan, D. Leip, and K. Gupta (2006), The use of open-source software in the IBM corporate portal
3. Dan Woods, Gautam Guliani (2005), *Open source for the enterprise*, Published by O'Reilly
4. Mikko Valimaki (2005), *The Rise of Open Source Licensing: A Challenge to the Use of Intellectual Property in the Software Industry* Published by Turre Publishing
5. Siva Vaidhyathan (2004) *The Anarchist in the Library: How the Clash Between Freedom and Control Is Hacking the Real World and Crashing the System*, Published by Basic Books
6. Linux Today, available at: <http://www.linuxtoday.com> , accessed during may – July 2008
7. Netcraft, available at : <http://www.netcraft.com> , accessed during may – July 2008
8. Free Open Source Software: Pros and Cons from a Development Perspective by Simon Schneebeli, available at <http://www.wscsd.org/2004/01/04/open-source-software-fair-trade-for-software/> , accessed during may – July 2008
9. Usage Stats for PHP, available at: <http://www.php.net/usage.php> , accessed during may – July 2008
10. Gartner: Open source will quietly take over, Peter Judge available at: <http://news.zdnet.co.uk> Published: 04 Apr 2008, accessed during may – July 2008
11. How to use open source (and how not to) available at: <http://pingv.com/blog> , Published 2007, accessed during may – July 2008

About the Authors

Punam Gupta is Senior Technical Director and Head, Training Division & Application Division 8, National Informatics Centre, Government of India

Sapna Kapoor is Principal Systems Analyst, Training Division & Application Group 8, National Informatics Centre, Pune