

O ANALIZĂ FUNCȚIONALĂ A DISTRIBUȚIILOR GNU/LINUX.
STUDIUL DE CAZ A DOUĂ DISTRIBUȚII (REDHAT ȘI DEBIAN)

NAPOLEON-ALEXANDRU SIRETEANU*

Functional Analysis of Gnu/Linux Distributions. A Case Study of Two Distributions
(Red Hat and Debian)

Abstract

GNU/Linux distributions are very well coordinated and assembled pieces of software. Each one is in some sense a snapshot of a large fraction of the libre software development landscape at the time of the release. The components of such software systems (about 50-100 millions of lines of source code) are built by hundreds of independent groups of developers, with no formal connection to the group releasing the whole system. In this article, we present some quantitative and comparative information about the software included in two such distributions (Red Hat and Debian) and insist especially on the number of lines of source code. Both distributions are doubling their size every two years and both present similar patterns in programming language usage and package size distributions.

Keywords: Linux, Red Hat, Debian, source lines of code

1 Introducere

În ultimii ani, software-ul gratuit și deschis a determinat apariția unui interes deosebit manifestat în mediul de afaceri, dar și în lumea academică. Deoarece proiecte din domenii de aplicativitate diferite (cum ar fi Linux împreună cu suita de utilitare GNU, GNOME, KDE, Apache, sendmail, bind și diferite limbaje de programare) au obținut un succes deosebit pe piață, s-a demonstrat că acest nou model (paradigmă) de dezvoltare poate duce la rezultate de o calitate remarcabilă. Aceasta a determinat un interes deosebit al lumii afacerilor și a permis apariția unor noi corporații ca RedHat sau VA Software și a forțat corporațiile mici și mari (IBM, Sun Microsystems sau Netscape) să investească într-un anumit mod în această nouă direcție.

Software-ul liber (software-ul gratuit și software-ul open-source) oferă o inginerie software cu oportunități unice în privința caracterizării detaliate a proiectelor software, care pot fi astfel complete, detaliate și reproductibile deoarece codul sursă poate fi citit de oricine. Aceasta face posibilă construirea unor modele complete bazate pe studii publice și repetabile. Conform acestei idei, pare rezonabil să colectăm datele din proiectele software libere, pentru a obține anumite date care pot fi ulterior utilizate în susținerea teoriilor privitoare la modul în care softul liber este dezvoltat.

* Lector doctorand, Catedra de Informatică Economică, Facultatea de Economie și Administrarea Afacerilor, Universitatea „Alexandru Ioan Cuza” Iași, e-mail: salexis@uaic.ro

Din acest motiv, cele două distribuții GNU/Linux sunt cele mai potrivite pentru a fi luate în discuție (termenul Linux se referă la nucleu, în timp ce distribuțiile software care se bazează pe acesta sunt cunoscute sub numele de GNU/Linux). În ultima jumătate a anilor '90, distribuțiile GNU/Linux au evoluat, astfel încât la începutul anilor 2000 includeau cele mai cuprinzătoare și bine coordonate compilații de software liber. Atunci când se studiază cele mai reprezentative distribuții, se analizează cel mai important și reprezentativ subset de software liber matur disponibil în momentul lansării unei astfel de distribuții. Descoperirea limbajelor care sunt cel mai des folosite în aceste distribuții și a modului în care evoluează dimensiunea pachetelor ne indică modul în care lucrează comunitatea software-ului liber și ne permite să anticipăm direcțiile viitoare (de exemplu, când limbajul C++ va fi superior limbajului C în ceea ce privește popularitatea limbajelor software libere).

Dimensiunea acestor distribuții este considerabilă și este dificilă găsirea unor colecții coordonate de software de dimensiunea distribuțiilor GNU/Linux, cu interdependențe complexe, formate din rezultatul a sute de proiecte software libere (coordonate de voluntari, companii sau de o mixtură a ambelor), care, când sunt distribuite satisfac cerințele a milioane de utilizatori din întreaga lume.

David A. Wheeler [Wheeler, D.A., 2000], [Wheeler, D.A., 2001] a fost primul care a făcut public un studiu detaliat al distribuției GNU/Linux Red Hat (versiunile 6.2 și 7.1), în timp ce González-Barahona și alții [González-Barahona, 2001], [González-Barahona, 2003] au studiat două versiuni de Debian. Alți autori au studiat Mozilla și Apache (de ex. Mochus) și GNOME (de ex. Koch). În cele ce urmează ne vom focaliza atenția asupra colecțiilor de software care includ produsele a sute și mii de proiecte software libere, cu dimensiuni de zeci de milioane de linii de cod sursă.

2 Red Hat și Debian - două distribuții majore GNU/Linux

Red Hat Linux a fost una din primele distribuții comerciale de GNU/Linux. În prezent, este cea mai cunoscută dintre distribuțiile comerciale. Debian a fost prima distribuție GNU/Linux asamblată de un grup de voluntari și este în prezent cea mai cunoscută distribuție non-comercială. Ambele distribuții dispun de modalități diferite de colectare a software-ului pe care utilizatorii le solicită, fiecare având modele de dezvoltare, detalii tehnice, obiective și metode de consolidare bine definite. În ciuda diferențelor existente între acestea, ambele sunt distribuții în sensul strict al cuvântului; dezvoltatorii Red Hat și Debian s-au axat în principal pe integrare și nu pe dezvoltarea software-ului. Ca integratori, aceștia au preluat codul sursă furnizat de autorii de software și l-au împachetat într-un mod care îndeplinea criterii precise (tehnice și organizaționale). Rezultatul acestui proces îl reprezintă o distribuție (un set de pachete organizate în mod convenabil care permite utilizatorului să le instaleze, elimine și actualizeze cu ușurință). Cei care realizează distribuțiile sunt responsabili pentru asigurarea calității, stabilității și securității, majoritatea proiectelor software fiind gestionate de voluntari [Michlmayr & Hill, 2003].

Deși Debian și Red Hat sunt realizatori de distribuții și prezintă caracteristicile menționate, obiectivele și politicile acestora sunt foarte diferite. De exemplu, în timp ce Red Hat are în vedere elementele de marketing pentru a decide în ce condiții și ce anume intră în distribuțiile sale, proiectul Debian ia în calcul caracteristicile tehnice principale. Din acest punct de vedere, în timp ce distribuțiile Red Hat sunt anunțate în avans și lansate în mod regulat (de obicei, la un interval de 6 luni), Debian lansează o versiune "când este gata", ceea ce înseamnă că deficiențele software critice sunt foarte scăzute. Din punct de vedere al conținutului, ambele distribuții sunt diferite - distribuțiile Debian sunt mult mai mari (în numărul de pachete și numărul total de linii cod sursă), acest lucru fiind influențat de următorii factori:

- Red Hat trebuie să realizeze anumite calcule economice pentru a decide cantitatea de efort necesară pentru o nouă distribuție, care depinde în cele din urmă de numărul de pachete; Debian ia în calcul numărul de voluntari, factorul limită în distribuții fiind perioada de timp necesară pentru dezvoltare și posibilitățile de coordonare (nu resursele economice);

- Red Hat s-a orientat către un anumit segment de utilizatori și a manifestat un interes minim în întreținerea unei piese minore software; în Debian, interesul pentru un anumit pachet este atribuit unui dezvoltator cu aptitudini deosebite care îl va introduce în distribuție.

Dacă luăm în discuție versiunile de pachete incluse, Debian și Red Hat prezintă, de asemenea, diferențe. Debian include, de obicei, în distribuții versiuni vechi ale pachetelor. Din acest motiv, este interesant să comparăm câteva din versiunile stabile din cele două distribuții: Red Hat 6.2, Debian 2.2 și Red Hat 7.1 (versiunile anumitor pachete incluse în aceste distribuții sunt prezentate în Tabelul nr. 1). Intervalul dintre cele două versiuni de Red Hat este de un an, în timp ce Debian 2.2 a fost lansat la mijlocul acestui interval. Se observă că majoritatea pachetelor Red Hat 6.2 au aceeași versiune ca pachetele din Debian 2.2 (care a fost lansată cinci luni mai târziu); există și situații când Debian 2.2 conținea versiuni mai vechi decât cele din Red Hat 6.2. Explicația este simplă. Deoarece Red Hat devenise orientat mult mai mult spre comercial, avea o motivație mai mare să dispună de ultima versiune a unui pachet, în timp ce în Debian presiunea pentru o astfel de politică nu era atât de ridicată. Procesul de lansare a Debianului include o fază a "înghețului"; în această perioadă nici o nouă versiune nu intră în ansamblul de pachete care vor fi în cele din urmă în versiunea stabilă, fiind permise doar programele care rezolvă deficiențele pachetelor din acest ansamblu. Această fază poate dura multe luni, ceea ce înseamnă că noile versiuni de pachete nu vor fi luate în considerare în versiunea finală.

Tabel nr. 1 - Versiuni de pachete în Red Hat 6.2, Debian 2.2 și Red Hat 7.1

Pachet	Versiune în Red Hat 6.2	Versiune în Debian 2.2	Versiune în Red Hat 7.1
Linux	2.2.14	2.2.19.1	2.4.2
Xfree86	3.3.6	3.3.6	4.0.3
Gdb	1991004	1990928	20010316
Python	1.5.2	1.5.2	1.5.2
Perl	5.005	5.0005	5.6.0
Gnome-libs	1.0.55	1.0.56	1.2.8
Apache	1.3.12	1.3.9	1.3.19
Glibc	2.1.3	2.1.3	2.2.2

Red Hat Software Inc. a fost înființată în 1994 de Bob Young și Marc Ewing, obiectivul său principal fiind compilarea și punerea în vânzare a unei distribuții GNU/Linux, care se numea și încă se numește Red Hat Linux [Young, 1999]. Aceasta consta, în principal, dintr-o versiune împachetată a programelor existente pe rețea în acel moment, inclusiv documentație și suport. Versiunea 1.0 a fost lansată în vara anului 1995, iar versiunea 2.0 a fost lansată în toamna anului 1995 și includea managerul de pachete Red Hat (RPM), acesta devenind mai târziu standard *de facto* pentru pachetele GNU/Linux. În 1998, Red Hat a lansat versiunea 5.2, fiind prima distribuție GNU/Linux care a avut o cerere mare pe piață.

Înainte să apară sistemul RPM, aproape toate distribuțiile GNU/Linux puteau fi instalate folosind o procedură bazată pe un meniu, modificările și adăugirile de software după prima instalare nefiind deloc ușoare. RPM oferea un management al pachetelor, care permitea cu ușurință eliminarea, instalarea sau actualizarea oricărui pachet software

disponibil în distribuție. RPM este și în prezent cel mai utilizat sistem de management al pachetelor în distribuțiile GNU/Linux. Din statisticile de pe situl DistroWatch din mai 2003, 118 distribuții utilizau RPM (55%), în timp ce managerul de pachete Debian (DEB) era utilizat doar de 16 distribuții (14%). În prezent, conform statisticilor din aprilie 2006, 129 distribuții utilizează DEB, în timp ce 112 distribuții folosesc RPM [DistroWatch, 2006].

Debian este un sistem de operare liber, care utilizează nucleul Linux pentru a asambla distribuția software Debian GNU/Linux. Distribuția este disponibilă pentru diferite arhitecturi, inclusiv Intel x86, Motorola 680x0, Power PC, Alpha și SPARC. Distribuția Debian este alcătuită doar din software liber și este conformă Ghidului Debian pentru Software-ul Gratuit (DFSG) [Debian Project, 2003]; este disponibilă pe Internet pentru descărcare, mulți redistribuitori oferind-o pe CD sau alt suport. Dezvoltatorii Debian adaptează și împachetează întregul software inclus în distribuție, întrețin serviciile bazate pe Internet (site web, arhive online, sistem de urmărire a deficiențelor software, liste de e-mail pentru suport și dezvoltare), realizează operații de traducere și internaționalizare, dezvoltă utilitare specifice Debian și, în ansamblu, susțin toată infrastructura care permite lansarea distribuției Debian. Distribuțiile Debian oferă codul sursă al pachetelor pentru o compilare ulterioară și reconstruire a surselor originale.

3 Situația liniilor de cod

Numărul de linii fizice de cod este una din cele mai simple și larg răspândite tehnici folosite pentru compararea pieselor software. În analiza care va urma am luat în calcul numărul de linii de cod a cinci distribuții Red Hat (5.2 din 1998, 6.0 din 1999, 6.2 din 2000, 7.1 din 2001 și 8.0 din 2002) și patru distribuții Debian GNU/Linux (2.0 din 1998, 2.1 din 1999, 2.2 din 2000 și 3.0 din 2002). Aceste versiuni reprezintă un reper important pentru ambele distribuții pe o perioadă de 4 ani, cele specificate fiind cele mai stabile Debian și cele mai reprezentative Red Hat. În tabelul și graficul următor (vezi Tabel nr. 2 și Fig. 1) am analizat numărul de linii și de pachete sursă pentru aceste distribuții.

Tabel nr. 2 - Dimensiunea distribuțiilor Red Hat și Debian GNU/Linux

Nume	Data lansării	Pachete sursă	Dimensiune (milioane linii de cod sursă)	Dimensiunea pachetelor (linii de cod sursă)
Red Hat 5.2	1998-10	403	12	31200
Red Hat 6.0	1999-04	433	15	35500
Red Hat 6.2	2000-03	470	18	39900
Red Hat 7.1	2001-04	605	32	52800
Red Hat 8.0	2002-09	792	50	63000
Debian 2.0	1998-07	1096	25	22850
Debian 2.1	1999-03	1551	37	23910
Debian 2.2	2000-08	2611	59	22650
Debian 3.0	2002-07	4579	105	22860

Red Hat Linux 5.2, lansat la sfârșitul anului 1998, includea 403 pachete sursă, echivalentul a 12.5 milioane linii de cod sursă. Red Hat 6.0 a fost lansat șase luni mai târziu și avea cu 30 de pachete mai mult și aproximativ 15 milioane de linii de cod sursă. Versiunea 6.2, apărută un an mai târziu, includea 470 de pachete sursă, cu aproximativ 18 milioane de linii de cod sursă. Red Hat 7.1, lansat cu an mai târziu, includea 605 pachete sursă, cu mai mult de 32 milioane de linii de cod sursă. În cele din urmă, versiunea 8.0, apărută după un an și jumătate, includea 800 de pachete sursă și aproximativ 50 milioane de

linii de cod sursă. În ansamblu, numărul de linii de cod sursă existent în 2002 în Red Hat a crescut de 4 ori față de 1998, ceea ce înseamnă că s-a dublat în dimensiune la aproximativ doi ani; numărul de pachete (deci și dimensiunea acestora) s-a dublat după o perioadă de 4 ani.

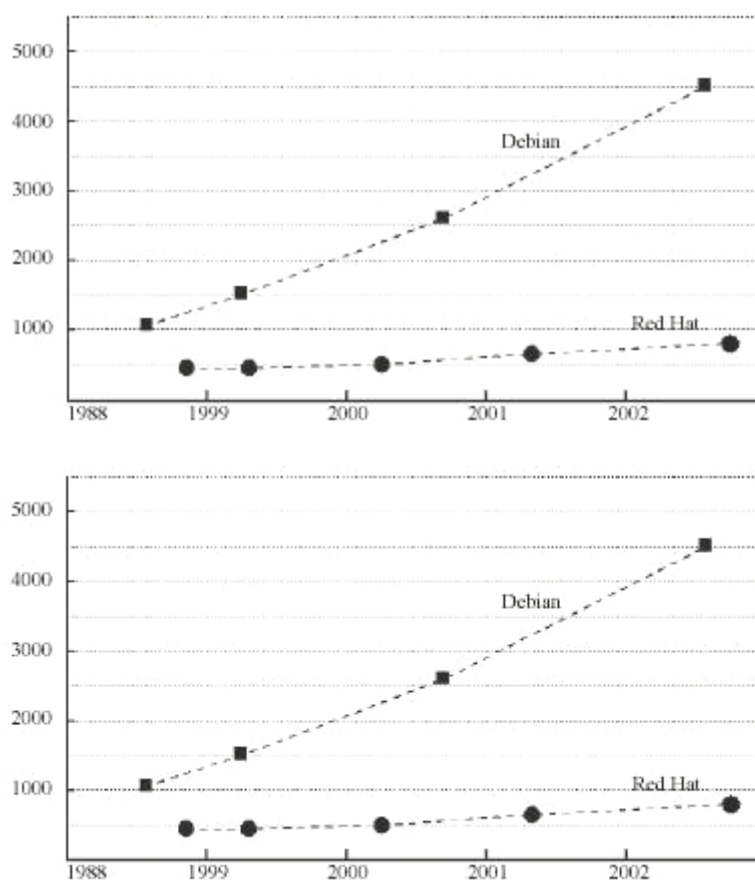


Fig. 1 Dimensiunea - în număr de linii de cod sursă și număr de pachete sursă - pentru fiecare versiune a distribuțiilor studiate

Debian 2.0 includea 1096 pachete sursă, echivalentul a mai mult de 25 milioane de linii de cod sursă. Următoarea versiune, Debian 2.1 (apărută după 9 luni) avea mai mult de 37 milioane de linii de cod sursă în 1551 pachete. Debian 2.2, apărută după 15 luni, conținea 59 milioane de linii de cod sursă distribuite în 2611 pachete. Ultima versiune stabilă a distribuției Debian GNU/Linux, Debian 3.0 (apărută la 2 ani după Debian 2.2) dispunea de 4579 pachete și aproape 105 milioane de linii de cod sursă. Aceasta înseamnă că în 4 ani, dimensiunea în număr de linii de cod sursă din Debian a crescut de 4 ori, dublându-se la fiecare 24 de luni, în timp ce numărul de pachete din distribuție a avut aceeași tendință de creștere. În concluzie, la fiecare doi ani, proiectul Debian a încorporat în distribuțiile sale mai mult cod decât în trecut.

Debian 2.2 este de două ori mai mare decât Red Hat 7.1 și de 3 ori mai mare decât Red Hat 6.2. Debian 3.0 este de 3 ori mai mare decât Red Hat 7.1 și de două ori mai mare față de Red Hat 8.0. Perioada de "îngheț" a Debian-ului, care în cazul lui Debian 3.0 a durat 6 luni, s-a datorat practic întârzierii pachetelor din distribuție.

4 Dimensiunea pachetelor

Echipele care lucrează la o distribuție decid când o piesă software este considerată pachet, în majoritatea cazurilor acestea urmând procedurile proiectului de dezvoltare original.

În cazul Debian-ului, există situații în care documentația este separată într-un alt pachet, acest aspect având un impact minim asupra pachetului (în privința liniilor de cod sursă), deoarece aceste pachete conțin de obicei foarte puțin cod sau deloc. În continuare vom face referiri la pachetele sursă, care includ codul sursă din care programele binare și librăriile pot fi construite.

În esență, când se studiază dimensiunea pachetelor sursă (numărul total al liniilor de cod sursă din distribuție împărțit la numărul de pachete sursă), există două diferențe semnificative între Debian și RedHat:

- dimensiunea pachetelor în distribuția Debian este deosebit de uniformă (aproximativ 23.000 de linii de cod sursă pentru Debian 2.0, 2.1, 2.2, și 3.0), în timp ce în RedHat s-a manifestat o creștere repetată (de la 31.000 de linii de cod sursă în RedHat 5.2 la 62.000 de linii de cod sursă în RedHat 8.0), în aceeași perioadă de timp. Aceasta se datorează faptului că, de-a lungul timpului, Red Hat nu s-a focalizat pe un set de pachete și numărul acestora a crescut în timp;

- dimensiunea pachetelor în distribuția Red Hat este de două ori mai mare comparativ cu pachetele Debian. Chiar și în 1998, când dimensiunea pachetelor RedHat a fost mult mai mică comparativ cu cea care a urmat după patru ani, era cu 40% mai mare față de pachetele Debian. În 2002, această proporție a fost de aproximativ unu la trei, fiind într-o continuă creștere și în prezent. Explicația este legată și de ceea ce am specificat anterior dar și de diferențele privind politicile de selectare a pachetelor care intră în distribuție.

5 Concluzii

Ambele distribuții (Red Hat și Debian) și-au dublat dimensiunea la fiecare doi ani, incluzând de două ori mai mult cod sursă în perioada 1998 – 2000 față de perioada anterioară. Același lucru poate fi spus și pentru perioada 2000 - 2002 și după aceasta. În viitor anticipăm o creștere a cantității de cod din Debian pentru software-ul liber la 400 milioane de linii de cod sursă (în 2006), ajungând chiar la miliarde în 2009; în cazul Red Hat-ului va fi de 200 milioane de linii de cod în 2006 și va atinge miliardul în 2011. Doar o viitoare analiză ne va edifica dacă presupunerea făcută va fi adevărată sau nu; este posibil ca software-ul liber să atingă o cotă maximă în viitorul apropiat. Studiul realizat poate fi folosit pentru a anticipa dezvoltarea întregului ecosistem de software liber pentru sistemele bazate pe UNIX și cele GNU/Linux.

Bibliografie

Debian Project, *Debian free software guidelines (part of the Debian social contract)*, 1999, la http://www.debian.org/social_contract, accesat pe 6 martie 2006.

DistroWatch, *Linux Distributions - Facts and Figures*, 2006, la <http://distrowatch.com/stats.php?section=packagemanagement>, accesat pe 2 aprilie 2006.

González, J., & Matellán-Olivera, V., "Counting potatoes: The size of Debian 2.2", în *Upgrade Magazine*. 2(6), 2001, la <http://upgrade-cepis.org/issues/2001/6/up2-6Gonzalez.pdf>, accesat pe 10 ianuarie 2006.

González-Barahona, J.M., Robles, G., Ortuño-Pérez, M., Roderio-Merino, L., Centeno-González, *Measuring Woody: The size of Debian 3.0.*, 2003, nepublicată, la

<http://people.debian.org/~jgb/debian-counting/>, accesat pe 12 ianuarie 2006.

Michlmayr, M. & Hill, B.M., "Quality and the reliance on individuals in free software projects", în *Proceedings of the 3rd Workshop on Open Source Software Engineering*, Portland, 2003, la <http://opensource.ucc.ie/icse2003/3rd-WS-on-OSS-Engineering.pdf>, accesat pe 11 ianuarie 2006.

Wheeler, D.A., *Estimating Linux's size*, 2000, la <http://www.dwheeler.com/sloc/>, accesat pe 15 ianuarie 2006.

Wheeler, D.A., *More than a gigabuck: Estimating GNU/Linux's size*, 2001, la <http://www.dwheeler.com/sloc/>, accesat pe 10 ianuarie 2006.

Young, R., *Giving it away. How Red Hat Software stumbled across a new economic model and helped improve an industry*, la <http://www.oreilly.com/catalog/opensources/book/young.html>, accesat pe 23 februarie 2006.