Open Source Business Intelligence and Data Warehousing

Seth Grimes

Alta Plana Corporation

+1 301-270-0795 -- http://altaplana.com

Technology Transfer

June 10, 2008

Alta Plana

Course goals and approach

The goal is that you understand:

- Open source business & technology.
- Open source BI & DW options.
- How to evaluate and implement.

Approach:

- There are many aspects of open source. I will try to be comprehensive and complete.
- I will maintain a practical focus.
- I provide a lot of reference material that I will not cover extensively.

Course goals and approach

Myself, I regularly use open-source software:

- Apache, MySQL, Python since the mid/late-'90s.
- Firefox, Thunderbird, OpenOffice, GIMP, etc.
- I occasionally use R, GATE, cygwin, TrueCrypt.
- On a project: Mondrian, JPivot.

Course goals and approach

I believe in open source as a philosophy centering on collaboration and sharing.

I see open source as a viable approach to software publishing.

I think that open-source products, if properly chosen and used, are reliable, performant, and safe.

Perspectives

You are:

1. A BI end user or manager.
2. IT staff or management, responsible for supporting BI/DW in your organization.
3. A software developer creating BI applications or products.
4. Someone who develops or works with open source, and you want to learn about BI/DW.
5. Other?

This class will address the needs of all roles.

Course sections

The Business of Open Source

- The business case for open source.
- History, developments, and trends.
- Licensing and support.

Open Source Technology and Solutions

- Infrastructure, applications, databases, and tools.
- BI, data warehousing, and integration.

Evaluating and implementing

- Best Practices, strategy, and resources.

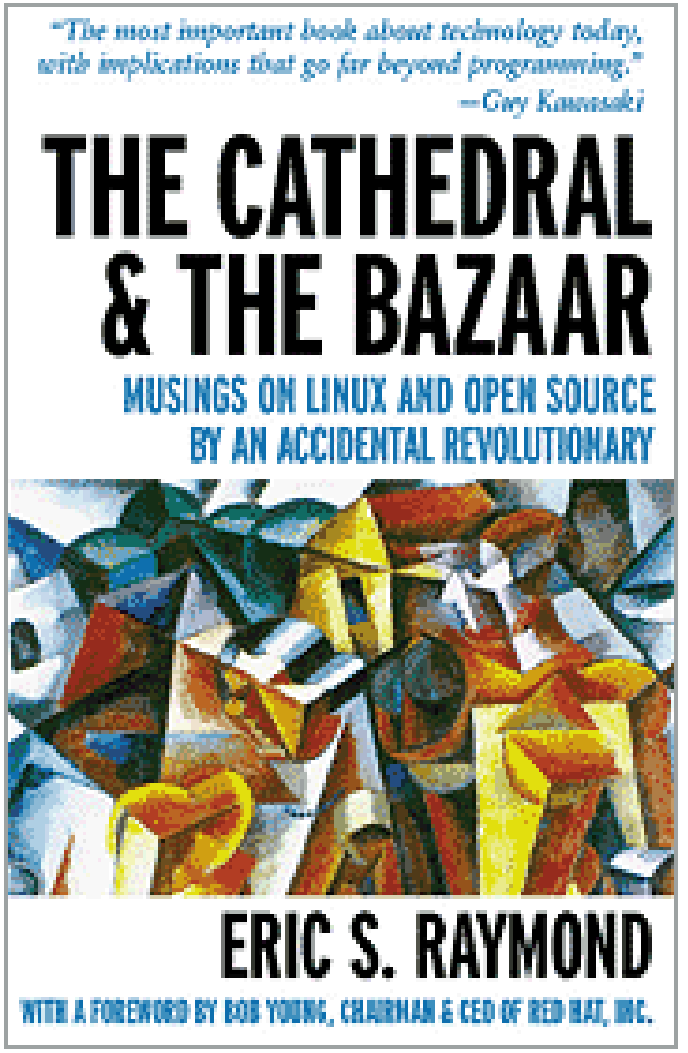
The Cathedral & the Bazaar

The Cathedral & the Bazaar

"I believed that the most important software (operating systems and really large tools like Emacs) needed to be built like cathedrals, carefully crafted by individual wizards or small bands of mages working in splendid isolation, with no beta to be released before its time.

"Linus Torvalds's style of development - release early and often, delegate everything you can, be open to the point of promiscuity - came as a surprise. No quiet, reverent cathedral-building here - rather, the Linux community seemed to resemble a great babbling bazaar of differing agendas and approaches (aptly symbolized by the Linux archive sites, who'd take submissions from *anyone*) out of which a coherent and stable system could seemingly emerge only by a succession of miracles.

"The fact that this bazaar style seemed to work, and work well, came as a distinct shock."



Agenda

The Business Case for Open Source

Licensing and support

Infrastructure, applications, databases, and tools

BI, data warehousing, and data integration

Evaluation and implementation best practices, strategy, and resources

Origins of Open Source

First Thesis:

Open Source is as old as computing.

- Modern computing originated in the WW II era.
  - New electronics, e.g., radar.
  - Need to support sophisticated logistics, artillery trajectories, etc.
  - Cryptography and code-breaking.
- Military-industry-university collaboration.
  - Continues to this day, e.g., DARPA, In-Q-Tel.

Origins of Open Source

Through the '60s, hardware vendors created and/or supplied most systems software.

- Logic design, "machine language" and assembler.
- Input/Output control.
- Batch/job control.
- Higher-level languages (Cobol, Fortran, Algol).
- Transaction management.
- Early database systems.

Software was bundled and was essentially free.

Origins of Open Source

The late '60s featured:

- ICs (integrated circuits) and lower costs.
- Commodity hardware with wider diffusion for business and science (e.g., the space program).

This led to independent software development, with a culture of openness and sharing, at universities and research centers.

- The real founding moment for open source was the creation of Unix at Bell Labs.

Commercialization changes the game.

Origins of Open Source

Open source is cooperative in essence, reflecting academic/research culture.

Closed source is competitive in essence, reflecting commercial culture.

… but most of the software world is not either/or.

Origins of Open Source

- Open source can be competitive in origins and intent and distribution.
  - Motivations include desire for recognition and influence and profit from added value, e.g., services and up-sell.
  - Open source developers are self-interested.
- Closed source can be open to…
  - Alliances.
  - Interoperation.

What is open source?

What is open source? For software:

- Source code is free and easily available.
- You can modify source code for your own purposes.
- You do not necessarily get an executable or documentation.
- It may not compile or run on every platform.
- You are not necessarily allowed to distribute alerted code.
- You may not be allowed to sell software that uses open source assets.

What is open source?

Not open source?

- Certain customers can see the source code for Microsoft Windows.
- Oracle owns InnoDB, a MySQL transactional engine, and Berkeley DB.

Open source?

- You must contribute any Linux kernel modifications that you distribute.
- Apache server may be used in commercial products.
- StarOffice is based on and extends OpenOffice.

Standards

Open source relies on standards.

- Windows API (application programming interface) and the .Net framework.
- SQL (Structured Query Language) and ODBC (open database connectivity).
- Java EE stack, e.g., JSR 168 portlet specification.
- UIMA (Unstructured Information Management Architecture), an text analytics interoperability framework.

Standards

Open source relies on standards.

- Some standards are *de-facto*, market standards:
  - Intel chip architectures and instruction sets.
  - Windows API (application programming interface) and the .Net framework.
  - ODBC (open database connectivity).
- Some standards are open, community owned:
  - * SPARC chip architectures and instruction sets.
  - * SQL (Structured Query Language).
  - JSR 168 portlet specification,* Java EE framework.

* = closed, proprietary origins

Standards

We have standards for:

- Hardware architecture.
- Operating systems.
- Software platform and applications stack.
- Programming tools.
- APIs and interoperability frameworks.

Open-source may run on closed-source hardware and software platforms and interoperate with closed-source software… and vice versa.

Standards

Development of open standards must be open –

- OASIS (Organization for the Advancement of Structured Information Standards) is a not-for-profit, international consortium that drives the development, convergence, and adoption of e-business standards.

  www.oasis-open.org/who/

- The Java Community Process (JCP) is the mechanism by which the Java community develops standard technical specifications for Java technology.

  jcp.org/en/home/index

More Open...

"Open" may also apply to or suggest:

- Algorithms and techniques.
- Processes.
- Information.
- Other intellectual property.

Open knowledge

We have the Open Access Initiative.

An "international effort to make research articles in all academic fields freely available on the Internet."

www.soros.org/openaccess/

"Our mission of disseminating knowledge is only half complete if the information is not made widely and readily available to society."

Berlin Declaration, October 2003; www.ec-petition.eu/

SPARC, the Scholarly Publishing and Academic Resources Coalition, is an international alliance of academic and research libraries that is backing this.

www.sparceurope.org/

Open designs

We have OpenSPARC (*www.opensparc.net/*).

- SPARC = Scalable Processor ARchitecture., a RISC architecture that dates to the late '80s.
  - Sun Microsystems, with SPARC, revolutionized workstation and network computing.
- Sun opened UltraSPARC T1 source code in 2006.
  - Lets "developers create innovative software applications faster… with a higher degree of hardware integration."
  - "Helps create an environment that will speed the development of new, thread-rich applications."
  - "Gives OEMs the opportunity to create unique solutions built on a proven architecture."

Open IP

Intellectual property.

- W3C patent policy (2004) requires royalty-free use of patents that are relied on by standards.

  www.w3.org/Consortium/Patent-Policy-20040205/

- IBM (January 2005) pledged open access to key innovations covered by 500 IBM software patents for use in open-source software.

  "While IP ownership is an essential driver of innovation, technological advances are often dependent on shared knowledge, standards, and collaborative innovation."

  www-03.ibm.com/press/us/en/pressrelease/7473.wss

  www.ibm.com/ibm/licensing/patents/pledgedpatents.pdf

Open IP

Intellectual property, continued.

- Sun Microsystems (January 2005) then released over 1,600 patents to open source.
  - Common Development and Distribution License.
  - Includes the OpenSolaris operating system platform.

    www.sun.com/smi/Press/sunflash/2005-01/sunflash.20050125.2.xml
- Open Invention Network created (2005).
  - IBM, NEC, Novell, Philips Red Hat, Sony.

    www.openinventionnetwork.com/index.php
- Patent commons initiative.
  - Started by the OSDL, now the Linux Foundation.

The Business Case

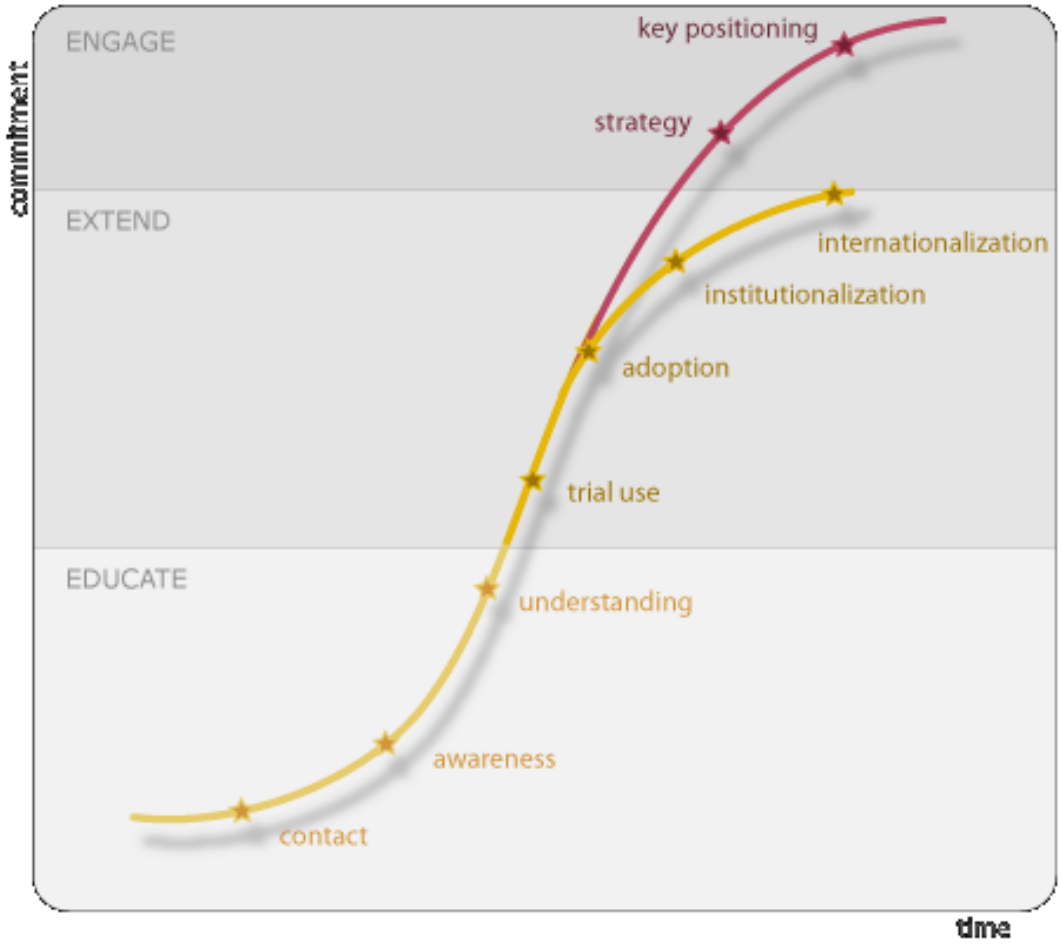
Second Thesis:

You are already using open-source software and services.

Why do people & organizations introduce OS?

- Everybody's using it, e.g., Apache Web server.
- Through a competitive evaluation, often against non-OS.
- Included in a distribution, that is, a packaging that includes a variety of software components.
- Unintentionally, not knowing it's OS.
- Platform-indicated choice, that is, it's the only option given your computing platform.
- IT mandate.

The Road to Open Source Adoption



"Technology adoption typically follows an S-Curve, going from first contact to complete institutionalization." – SourceSense

Is this true?

www.sourcesense.com/en/services/share/

The Business Case

From the blog of Sun Microsystems CEO Jonathan Schwartz, April 14, 2008 (*http://blogs.sun.com/jonathan/*):

"A few weeks ago, I was visiting… a large commercial institution… We had just closed the acquisition of MySQL, so before I wrapped up, I asked, "And would you like a quick update on the newest addition to our family, MySQL?"

"The CIO responded categorically with 'we don't run MySQL, we run [*name withheld to protect the proprietary*].' The CISO said, 'We can't just let developers download software off the net, you know, we've got regulation and security to worry about.' The CTO smiled. Everyone else appeared to be sitting on their hands. I was going to leave it at that. Thanks for the business.

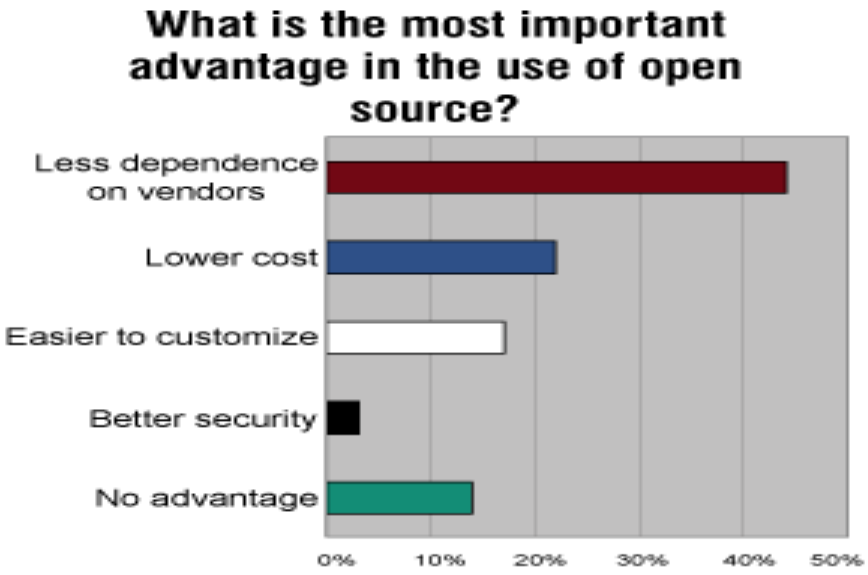
"Until a (diplomatically) assertive Sun sales rep piped up, 'Um... no, I connected with a buddy of mine over at MySQL, and had him check - you've downloaded MySQL more than 1,300 times in the last twelve months.'

"After a profoundly awkward silence, one of the individuals from their internal development team piped up, 'Actually, everybody uses it. Why bother hassling with license agreements when MySQL's got you covered. We're stoked you bought them.'"

The Business Case

Computer Economics surveyed visitors to its website regarding the perceived advantages in the use of open source software (May 2005).

www.computereconomics.com/article.cfm?id=1043



Evaluation criteria

If evaluation, what criteria?

1. Cost.
2. Ease of evaluation/introduction.
3. Flexibility.
4. Support.
5. Compatibility.
6. Security.
7. Ease of use.
8. Quality and capabilities.

Evaluation criteria

Criterion 1, Cost:

- Software is free, that is, no licensing cost.
  - Sometimes free, OS versions have limited capabilities or usage terms. More later.
- You must examine the Total Cost of Ownership (TCO), the cost of:
  - Supporting hardware.
  - User support and training.
  - Software maintenance, community participation.
  - Intellectual Property indemnification.

Evaluation criteria

Criterion 2, Ease of evaluation/introduction:

- You don't have to deal with vendor sales and their insistence on qualifying prospects.
- The vendor doesn't control the evaluation; they don't limit its duration, data volume, functions used, number of users taking part.
- By actually installing the software for evaluation, you can better assess the TCO.

Evaluation criteria

Criterion 3, Flexibility:

- You can modify the source code.
- You can use and install only those modules you need.
- You can often incorporate the software in commercial products or services at no royalty cost.
- You can move the software among machines without license hassles.

Evaluation criteria

Criterion 4, Support:

- Community support, often national/local, is readily available.
  - Forums, e-mail lists.
  - You can often reach developers directly.
- Commercial support is frequently available.
  - Sometimes provided by systems integrators.
  - Sometimes provided by companies that lead or "own" OS products such as Red Hat.

Evaluation criteria

Criterion 5, Compatibility:

- Open source applications are sometimes "free-standing," sometime platform dependent.
- OS tools are often part of a stack, an "ecosystem," or a distribution.
  - Java – Java EE is an "ecosystem" on which stacks are built. Perl, PHP, and Python are similar. So is Microsoft's non-OS .Net.
  - JBoss, Geronimo are OS platform; IBM's WebSphere is similarly a Java EE platform, albeit non-OS.
  - Eclipse is both a platform and a development environment.

Evaluation criteria

Criterion 6, Security:

- Is OS, with code transparency and community process, inherently more secure?
- With OS, there are no hidden back doors.
- With OS, you can close holes immediately, yourself.

Criterion 7, Ease of use:

- Varies by product and audience, for both open and closed source.

Evaluation criteria

Criterion 8, Quality and capabilities:

- Vary by component, context, and need.
  - For example, Linux is higher quality and far more scalable than Windows in the server context but lags in capabilities (namely applications software) for non-technical consumers.
- "Given enough eyeballs, all bugs are shallow."
  - – Eric Raymond in *The Cathedral and the Bazaar*.

Overcoming objections

Third Thesis:

Everyone is in the same boat, facing the same enterprise software challenges.

Points:

- Your enterprise is neither unique nor alone.
- You face cost and competitiveness pressures.
- You have limited resources and need to derive strategic value from information technology (IT).
- Security, support, and manageability are concerns.

Overcoming objections

Look for comparators:

- Other organizations in your line of business.
- Other organizations with a similar IT environment (size, complexity).
- Industry best practices for introduction and management of OS.

Overcoming objections

Thesis Four:

Open Source is viable: the movement and the products and the companies.

Overcoming objections

These Theses are the key to overcoming objections to OS:

- OS is well established. Products are often quite mature.
- Every organizations has OS. The question is not "if" but rather "how."

The business case starts with cost advantage.

It concludes with the realization that open source is not extraordinary. It is part of everyday IT.

Business/funding models

How can you make money from something free?

- IDC predicts a $3 billion 2009 OS software market.
- An EC study projects that OS will represent 32% of European software services by 2010.

ec.europa.eu/enterprise/ict/policy/doc/2006-11-20-flossimpact.pdf

How do open-source vendors survive?

- "Professional" and "enterprise" versions.
- Commercial-use and ISV licenses.
- Support and service revenues.
- Venture funding.

Business/funding models

Pentaho is a leading open source BI vendor.

- Venture Funding rounds:
  - $12 million Series C (February 2008).
  - $8 million Series B (July 2006).
  - $5 million Series A (December 2005).
- "The four-year-old company says it has had three million lifetime downloads, and has more than 20,000 registered community members. It says customers include Cox Communications, Monsanto Corporation, Savvion, Sun Microsystems and U.S. Naval Air Command."

http://venturebeat.com/2008/02/20/pentaho-raises-12m-for-open-source-business-intelligence-software/

Business/funding models

Venture funding totalled $204 million in the first quarter of 2008 according to the 451 Group.

What does a company do with venture funding?

- Developers, sales & marketing, support, admin.
- The same as a commercial-product company.

What justifies venture investments?

- Only "angel" funders do not prioritize returns.
- The company is itself a product. Value is created by:
  - Sales.
  - Company sale, merger, or IPO.

Open source sources

Organizations:

- Free Software Foundation.
- Apache.
- Mozilla.
- Open Source Initiative.

Community/multi-project hosting sites:

- *Sourceforge.net*.
- *ObjectWeb.org*.
- *Eclipse.org*.
- Microsoft's CodePlex (*codeplex.com/*)

Free Software Foundation

Founded by Richard Stallman in 1985.

Pioneered the notion of "copyleft."

- "All rights reversed."

Sponsors the GNU project.

- **G**NU's **N**ot **U**nix.
- A set of licenses: GPL and LGPL.
- An operating system: the Linux kernel (disputably) and systems and applications software.

Open Solutions Alliance

Open Solutions Alliance (2006) activities:

- Defining and promoting tools, frameworks and best practices that facilitate easy deployment and interoperability between member applications;
- Building "meta-communities" by partnering on projects that involve a variety of companies, communities and individuals to drive innovation and collaboration; and
- Coordinating joint marketing campaigns to raise awareness of business-hardened open applications and solution suites.

Questions?

Discussion?

Next: Licensing and Support

Agenda

The Business Case for Open Source

Licensing and support

Infrastructure, applications, databases, and tools

BI, data warehousing, and data integration

Evaluation and implementation best practices, strategy, and resources

Adoption challenges

The 451 Group lists challenges to adoption of open source:

* Commercial-grade support.

Deliberately spread fear, uncertainty and doubt.

FUD = a traditional Microsoft marketing tactic.

License proliferation.

* Security.

* Software quality.

www.the451group.com/caos/caos_community.php

* = a point I cited in the first session.

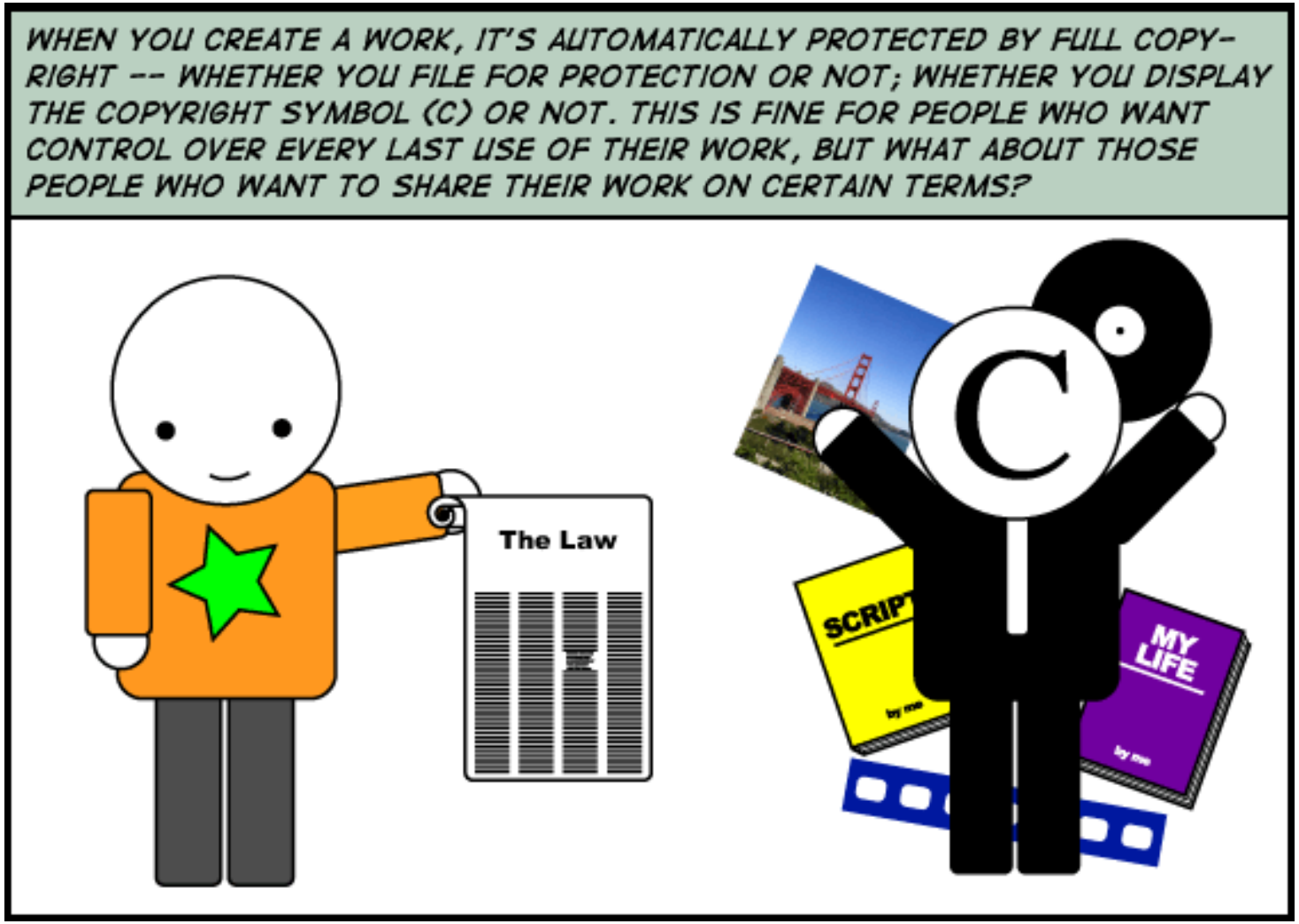
Licensing

A license is a contract that stipulates about (non-public domain) intellectual property:

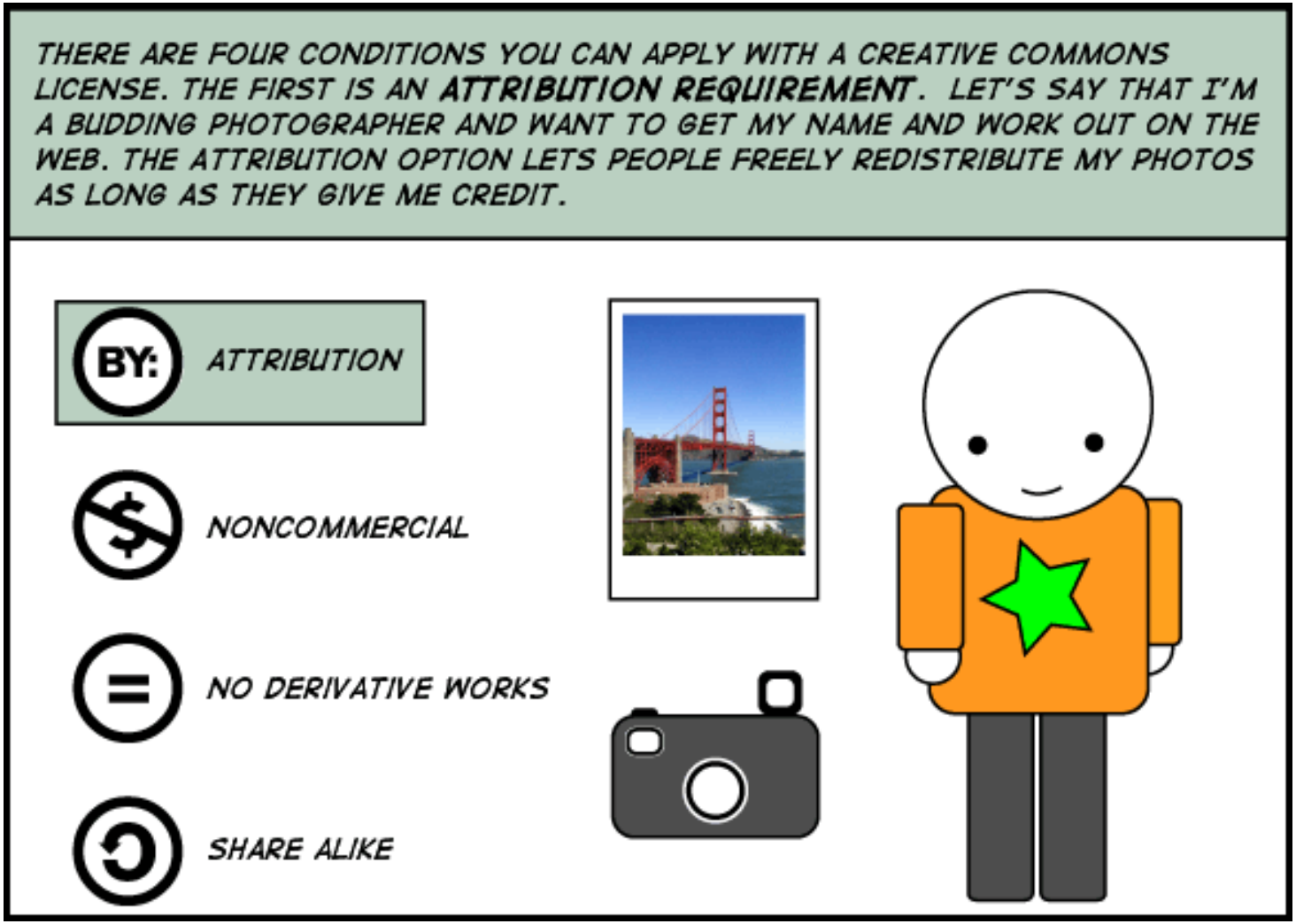
- Who may use it.
- When it may be used.
- Cost.
- Warranty, liability, disclosure, redistribution, derivation, and other licensee rights.
- For what purposes.
- (Software: On what machines.)

Creative Commons provides illuminating material…

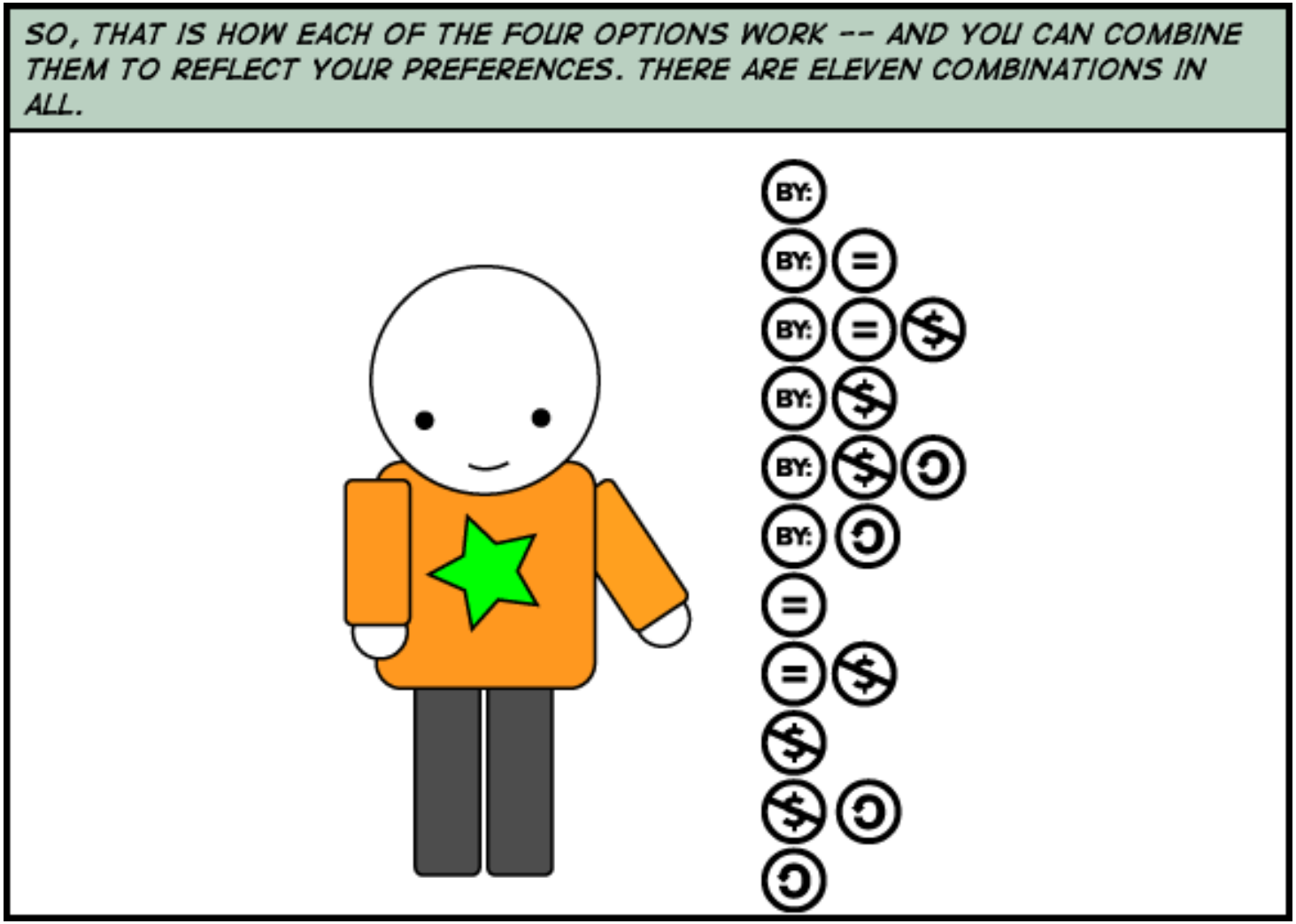
Licensing



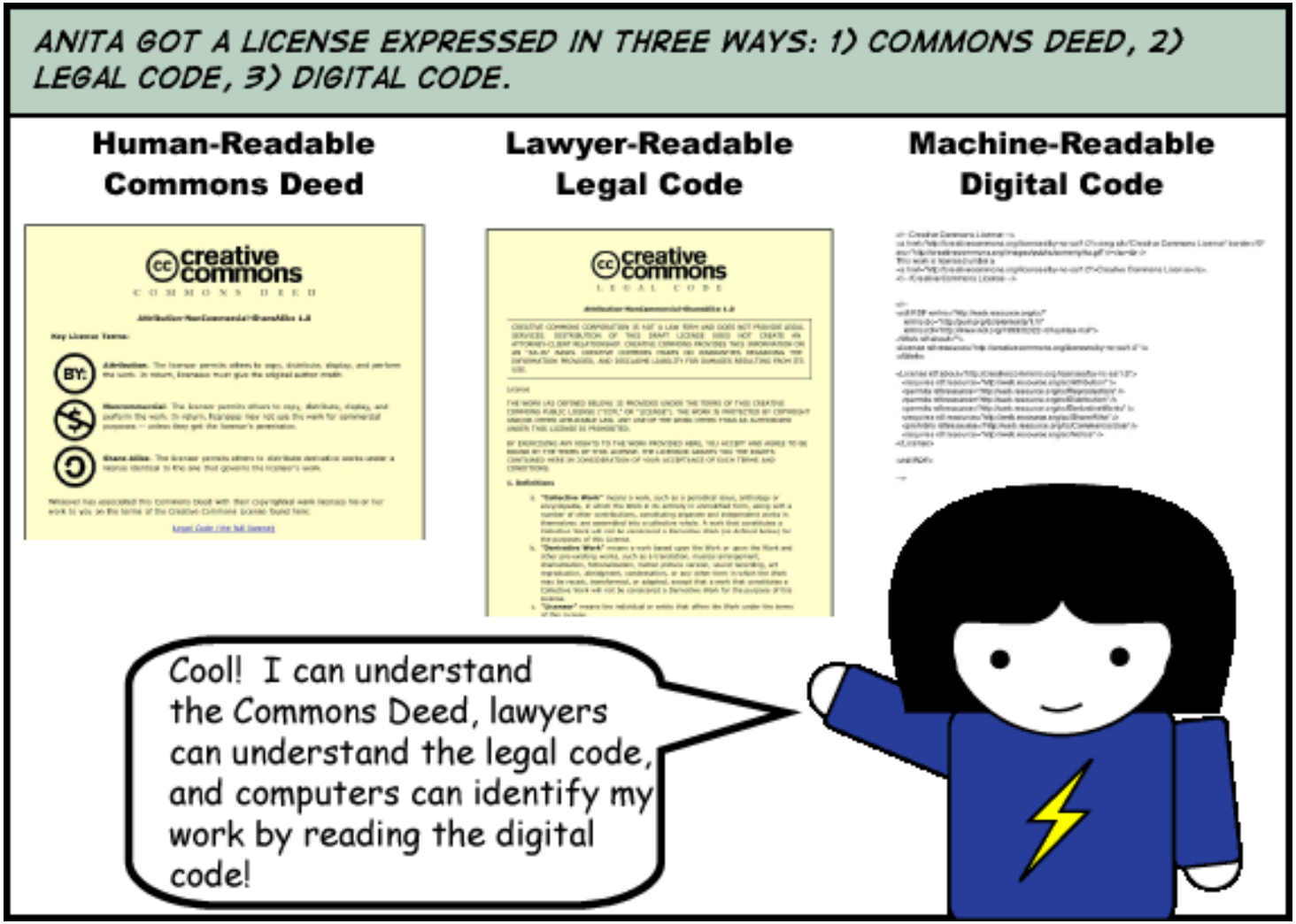
Licensing



Licensing



Licensing



Software licensing

If software is the *what*, licensing is the *how*.

Some licenses are *shrink-wrap*:

- Non-negotiable.
- Usually for less expensive, single-user software
- Often unread.

Some vendors offer multiple licensing schemes depending on:

- Use, e.g., commercial or non-commercial; trial/evaluation; academic.
- Rights, e.g., open source or proprietary.

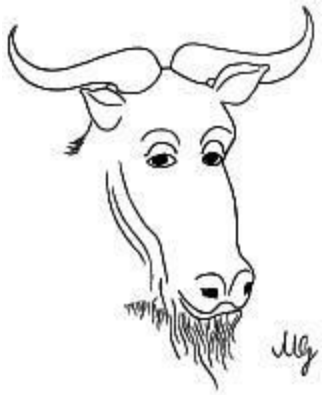
Software licensing models

There are many license models in the open-source world.

www.gnu.org/philosophy/categories.html

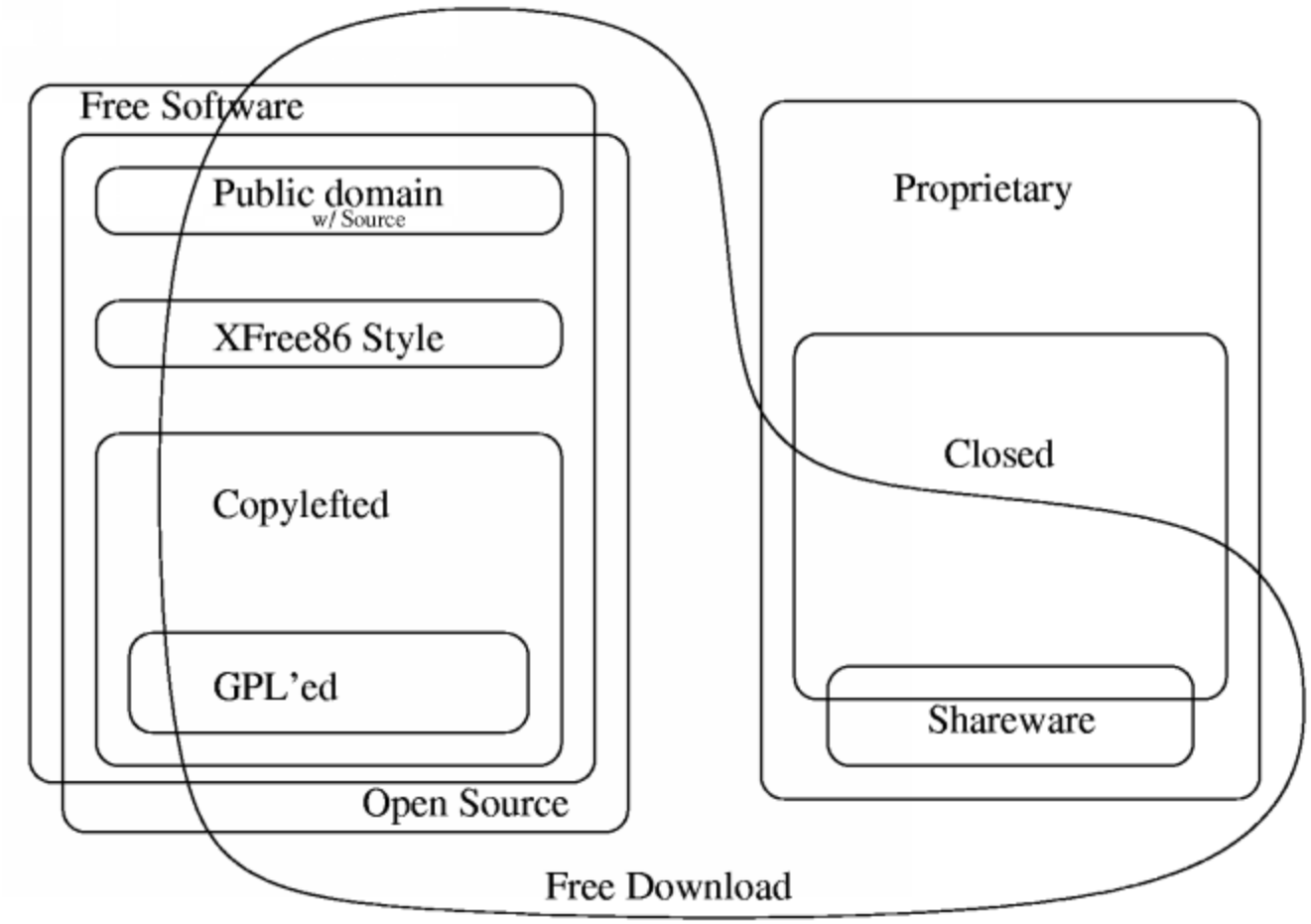
The GNU project has a list of models.

www.gnu.org/licenses/license-list.html



"Questo GNU appare particolarmente elegante con sopracciglia arcuate."

Let's start with Free…



www.gnu.org/philosophy/categories.html

Free software

According to the GNU project:

"Free software" is a matter of liberty, not price. To understand the concept, you should think of *"free"* as in *"free speech,"* not as in *"free beer."*

Free software is a matter of the users' freedom to run, copy, distribute, study, change and improve the software. More precisely, it refers to four kinds of freedom, for the users of the software:

- The freedom to run the program, for any purpose (freedom 0).
- The freedom to study how the program works, and adapt it to your needs (freedom 1). Access to the source code is a precondition for this.
- The freedom to redistribute copies so you can help your neighbor (freedom 2).
- The freedom to improve the program, and release your improvements to the public, so that the whole community benefits (freedom 3). Access to the source code is a precondition for this.

Free software

- Consider PostgreSQL.
  - Free, open source RDBMS.
  - BSD license: allows proprietary modification and licensee-restricted redistribution.
- Consider EnterpriseDB.
  - Postgres Plus is layered on PostgreSQL and includes FOSS extensions to PostgreSQL.
  - Postgres Plus Advanced Server is not FOSS.

Open source licenses

What is an open source solution?

- Free?
- Source code available for free?
- Source code may be redistributed?
- Anyone can contribute?

The Open Source Initiative (OSI) is

- involved in OS community-building and education and
- provides a formal definition.

www.opensource.org/docs/osd

OSI definition of open source

Distribution terms must comply with these criteria:

1. Free Redistribution.
2. Source Code.
3. Derived Works.
4. Integrity of The Author's Source Code.
5. No Discrimination Against Persons or Groups.
6. No Discrimination Against Fields of Endeavor.
7. Distribution of License.
8. License Must Not Be Specific to a Product.
9. License Must Not Restrict Other Software.

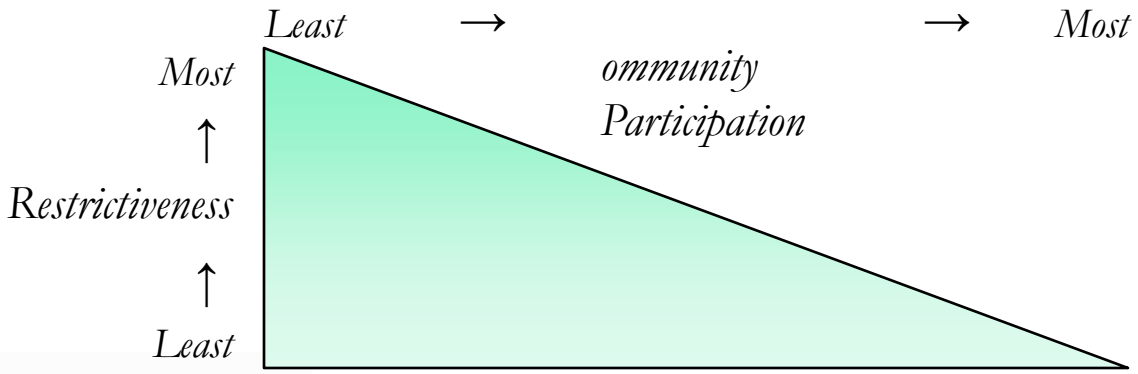
*10. License Must Be Technology-Neutral.

Open source (licensing) incline

Dana Blankenhorn talks about the "open source incline" for licensing strategy:

> "Newcomers start with BSD licenses which protect their right to profit but are eventually pushed toward GPL licenses… to secure the benefits of community participation in their projects."

blogs.zdnet.com/open-source/?p=756



Open source (licensing) incline

There's a point missing.

- The Ingres DBMS started as closed source, effectively a version of Michael Stonebraker's university work that evolved to Postgres.
- Ingres landed in the hands of Computer Associates, which released it to open source in 2004.
- Original CA-TOSL (T=Trusted), a Common Public License derivative, wasn't friendly.
- Then CA spun off the company; now GPL.

… so some OS projects start as commercial.

Commercial licensing

Many variations –

- Per machine/CPU/core.
- Per "seat."
- Per named user.
- Value based.

And possible conditions –

- Fee for support, maintenance (upgrades).
- Fee for transfer between machines or named users.
- No resale or transfer of rights.
- No reverse engineering, access to source code.

Shared source licensing

Microsoft shared-source are most restrictive.

- Microsoft Permissive License (Ms-PL).
  - View, modify, and redistribute the source code for either commercial or non-commercial purposes.
- Microsoft Community License (Ms-CL).
  - Modification and redistribution of licensed software with a per-file reciprocal term.
- Microsoft Reference License (Ms-RL).
  - View source code, no modification or redistribution.

www.microsoft.com/resources/sharedsource/licensingbasics/sharedsourcelicenses.mspx

BSD licensing

BSD = Berkeley Software Distribution, which includes a fork of the Unix operating system.

Allows proprietary commercial use and for the software released under the license to be incorporated into proprietary commercial products.

www.opensource.org/licenses/bsd-license.php

BSD license, 1

<OWNER> = Regents of the University of California
<ORGANIZATION> = University of California, Berkeley
<YEAR> = 1998

In the original BSD license, both occurrences of the phrase "COPYRIGHT HOLDERS AND CONTRIBUTORS" in the disclaimer read "REGENTS AND CONTRIBUTORS".

Here is the license template:

Copyright (c) <YEAR>, <OWNER> All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

•Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

•Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

•Neither the name of the <ORGANIZATION> nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

…

BSD license, 2

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

GNU licenses

General Public License, GPL.

- Copyleft! *www.gnu.org/licenses/gpl.html*
- Used for a majority of OS projects including Linux, MySQL, Samba, Alfresco.
- GPL 3, released last year, was controversial , including provisions redistribution of software when used in devices such as TIVO.

GNU Lesser General Public License, LGPL.

- Designed to permit linking OS libraries into non-free programs. *www.gnu.org/licenses/lgpl.html*

License case study: Mozilla

Mozilla as a case study:

- Very successful software.
- Time-tested, widely emulated licensing.
- An exemplary community-based, commercially backed open source project.
- Software prizes standards adherence, innovation.

License case study: Mozilla

Core-product source code –

MPL/GPL/LGPL tri-license or a license compatible with all three of those (e.g. the BSD license).

www.mozilla.org/MPL/

To understand MPL better, review the Annotated Mozilla Public License, version 1.1.

Official binary releases –

Mozilla End-User Licensing Agreements (EULAs).

Vary by product, release version.

www.mozilla.com/en-US/legal/eula/

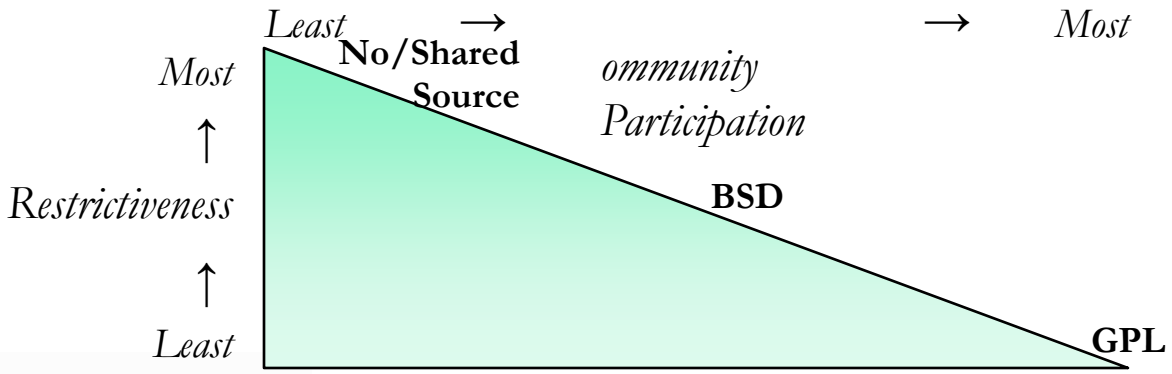
Licensing summary

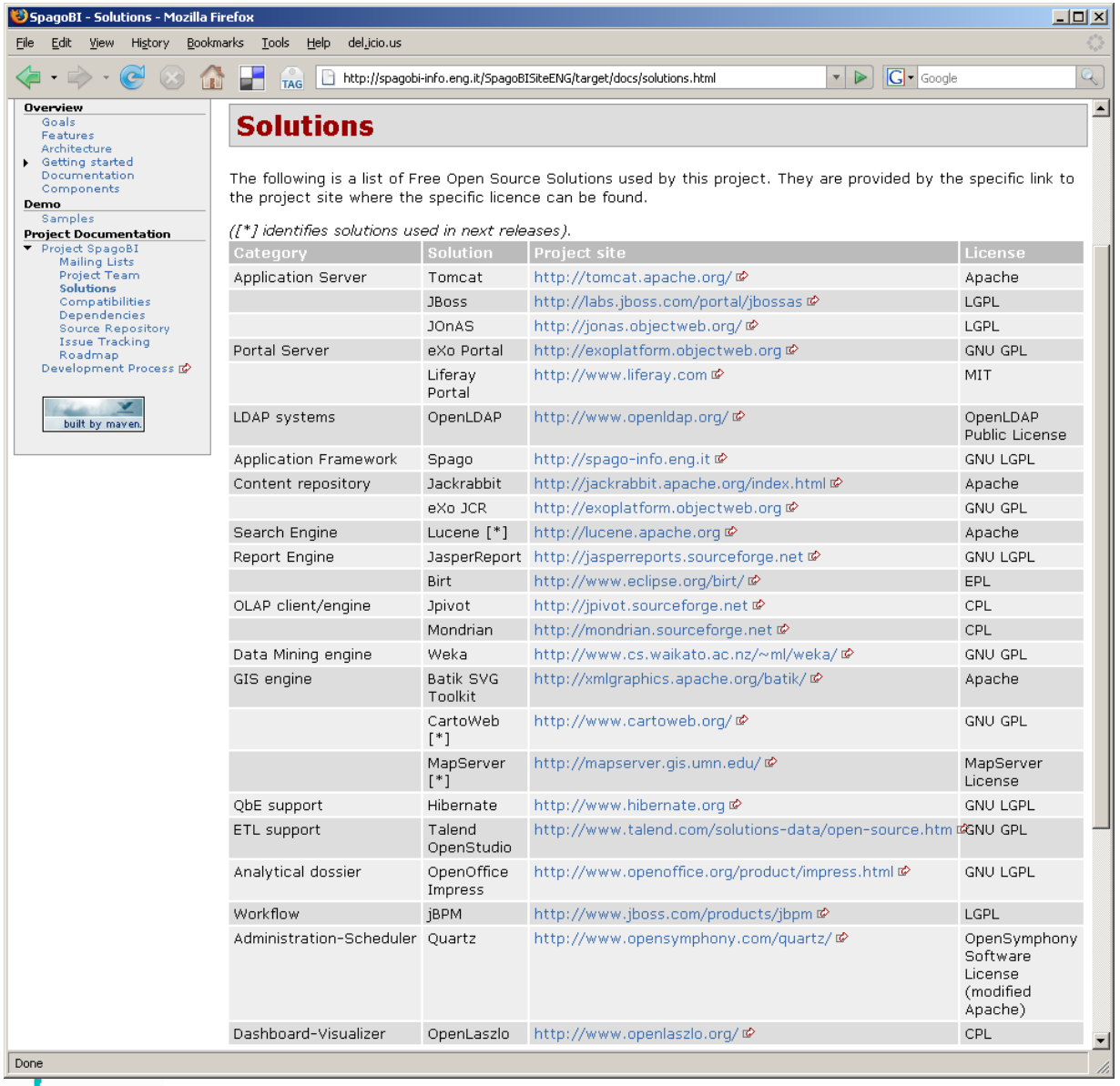
No-source & shared-source are most restrictive.

BSD-type licenses are next.

- Allowing for commercial use without give-back.
- There are many variants including Apache and Mozilla.

GPL is most open.





Category	Solution	Project site	License
Application Server	Tomcat	http://tomcat.apache.org/	Apache
	JBoss	http://labs.jboss.com/portal/jbossas	LGPL
	JOnAS	http://jonas.objectweb.org/	LGPL
Portal Server	eXo Portal	http://exoplatform.objectweb.org	GNU GPL
	Liferay Portal	http://www.liferay.com	MIT
LDAP systems	OpenLDAP	http://www.openldap.org/	OpenLDAP Public License
Application Framework	Spago	http://spago-info.eng.it	GNU LGPL
Content repository	Jackrabbit	http://jackrabbit.apache.org/index.html	Apache
	eXo JCR	http://exoplatform.objectweb.org	GNU GPL
Search Engine	Lucene [*]	http://lucene.apache.org	Apache
Report Engine	JasperReport	http://jasperreports.sourceforge.net	GNU LGPL
	Birt	http://www.eclipse.org/birt/	EPL
OLAP client/engine	Jpivot	http://jpivot.sourceforge.net	CPL
	Mondrian	http://mondrian.sourceforge.net	CPL
Data Mining engine	Weka	http://www.cs.waikato.ac.nz/~ml/weka/	GNU GPL
GIS engine	Batik SVG Toolkit	http://xmlgraphics.apache.org/batik/	Apache
	CartoWeb [*]	http://www.cartoweb.org/	GNU GPL
	MapServer [*]	http://mapserver.gis.umn.edu/	MapServer License
QbE support	Hibernate	http://www.hibernate.org	GNU LGPL
ETL support	Talend OpenStudio	http://www.talend.com/solutions-data/open-source.htm	GNU GPL
Analytical dossier	OpenOffice Impress	http://www.openoffice.org/product/impress.html	GNU LGPL
Workflow	jBPM	http://www.jboss.com/products/jbpm	LGPL
Administration-Scheduler	Quartz	http://www.opensymphony.com/quartz/	OpenSymphony Software License (modified Apache)
Dashboard-Visualizer	OpenLaszlo	http://www.openlaszlo.org/	CPL

Open source support

Support may mean a number of things:

- Help understanding enterprise needs.
- Installation and configuration assurance.
- End-user training and on-going help.
- Developer assistance.
- Software maintenance via bug tracking, patches
- Feature prioritization, software upgrades.
- Indemnification, i.e., assurance of intellectual-property rights.

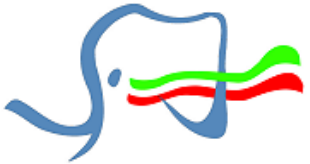
Community

- Major projects are community and/or commercially supported.
  - Take RHEL: community supported Linux core, commercially supported distro packaging.
- Some projects don't welcome community contributions, e.g., MySQL.
- Some projects, especially smaller ones, have no community.
  - Are they sustainable?

Community

Comunità Italiana Utenti e Sviluppatori PostgreSQL

www.psql.it/



Italian Postgres Day, in Prato

http://www.pgday.org/it/

Italian Linux Society

www.linux.it/

Conferenza Italiana sul Software Libero

http://www.confsl.org/

Pubblica Amministrazione Aperta e Libera

http://www.paal2008.it/

Questions?

Discussion?

Next: Open Source Infrastructure, applications, databases, and tools

Agenda

The Business Case for Open Source

Licensing and support

Infrastructure, applications, databases, and tools

BI, data warehousing, and data integration

Evaluation and implementation best practices, strategy, and resources

Unix

Created in 1969 AT&T's Bell Laboratories.

> *"When BTL withdrew from the project, they needed to rewrite an operating system (OS) in order to play space war on another smaller machine (a DEC PDP-7 [Programmed Data Processor] with 4K memory for user programs). The result was a system which a punning colleague called UNICS (UNiplexed Information and Computing Service)--an 'emasculated Multics'; no one recalls whose idea the change to UNIX was."*

BSD code fork occurred in 1975; 4.2 in 1984.

www.unix.org/what_is_unix/history_timeline.html

Unix

- Unix *per se* is not open source although some variants are:
  - FreeBSD, NetBSD, OpenBSD.
  - OpenSolaris! (*www.opensolaris.org/os/*)
    - The only OS variant of Unix System V Release 4.
- Why has BSD spawned more open source projects than SVR4? Consider university –
  - Culture and community.
  - Intellectual property.
  - Applications.

OpenSolaris

Sun Microsystems –

- Early versions of SunOS were based on BSD Unix.
- Started using the Solaris branding in 1991.
- Solaris 5 in 1994 was based on System V Release 4.
- After Solaris 10 (January 2005), Sun incrementally released Solaris source code to open source.

OpenSolaris = community-supported code base, build tools, and development infrastructure.

Solaris OS is branded, tested, maintained, and supported as a Sun product.

Linux

Linux is a Unix clone.

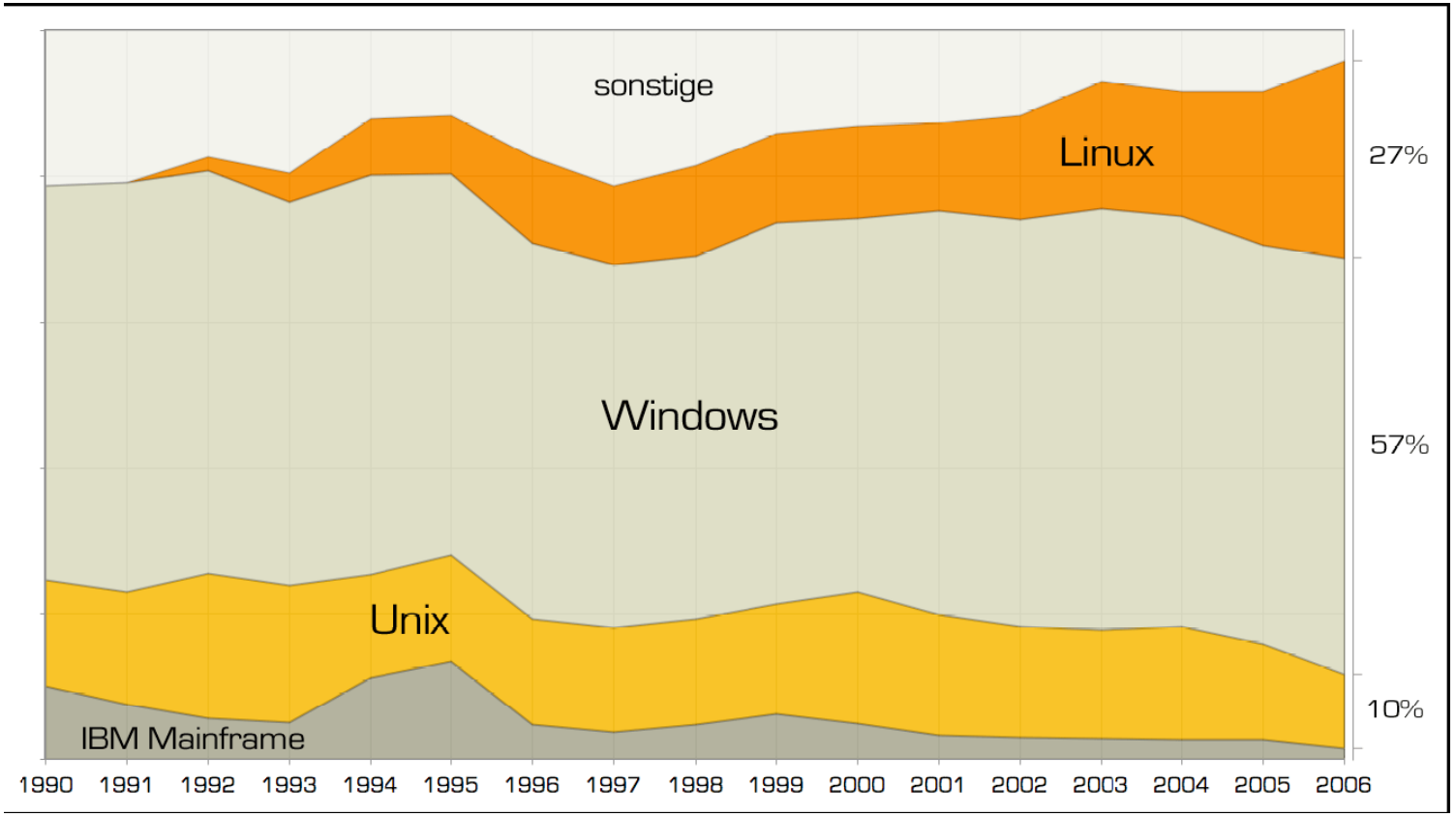
www.unix.org/what_is_unix/history_timeline.html

Created in 1991 by Linus Torvalds, who –

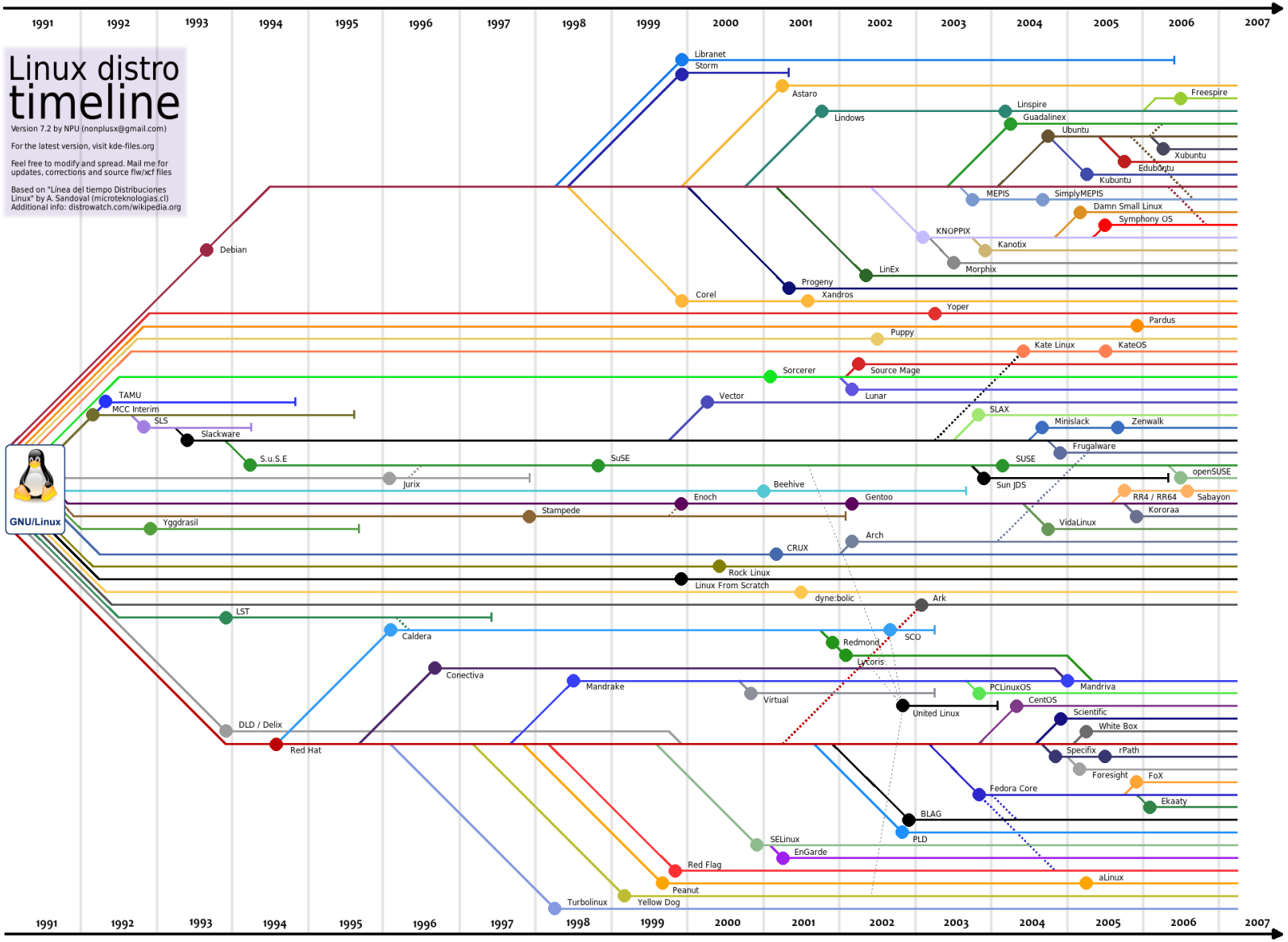
- Owns the Linux trademark, which is managed by the Linux Mark Institute (*www.linuxmark.org/*).

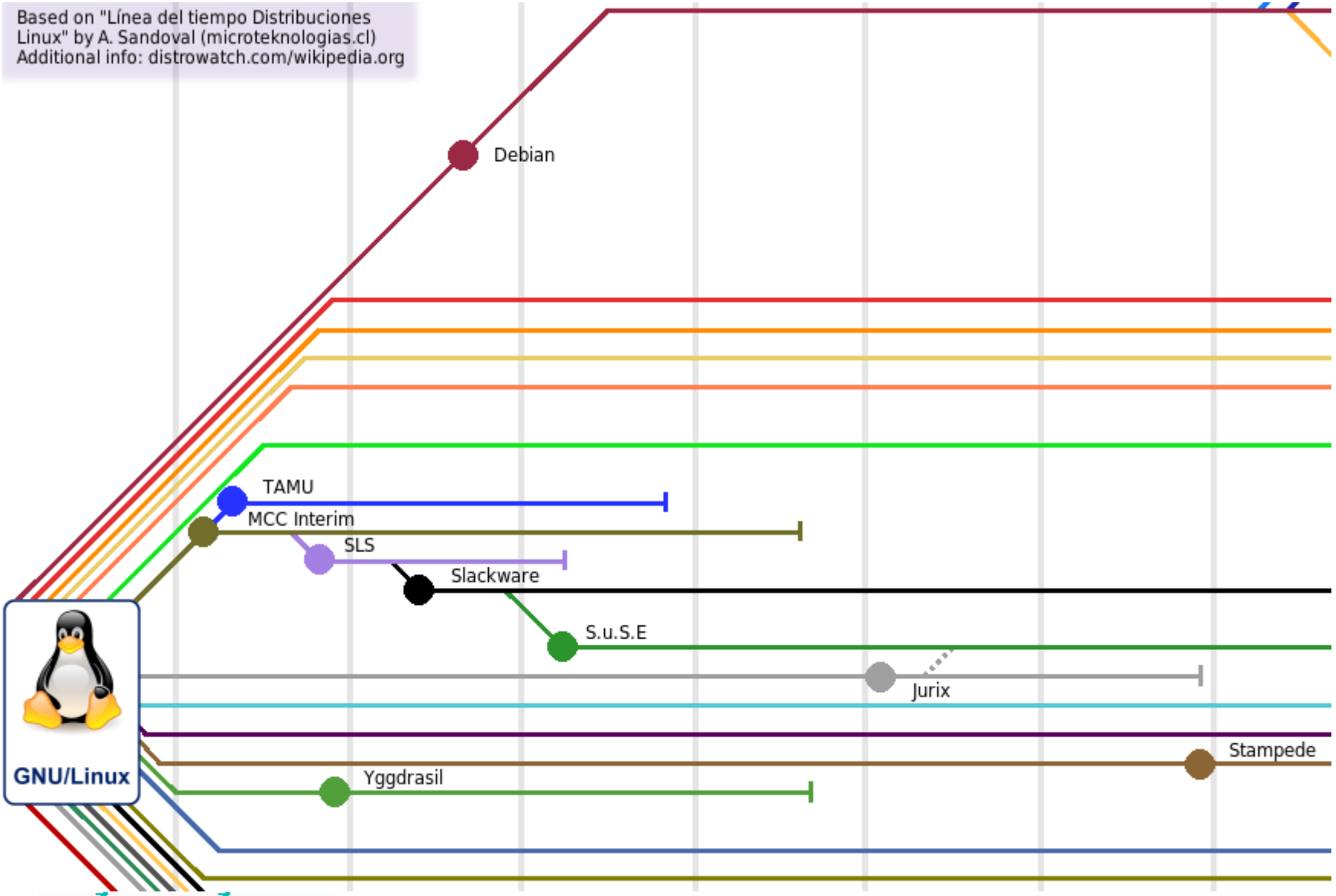
The kernel is managed by the Linux Kernel Archives (*www.kernel.org/*).

Linux



Source: JoinVision E-Services GmbH, July 2006, survey of the **JoinVision community**, *www.joinvision.com/jv/ext/infow/itfacts/200606/itfacts200606.pdf*





Red Hat

Linux is the kernel.

Fedora Core is Red Hat's OS distribution.

Red Hat Enterprise is a supported version.

> *Red Hat's fiscal year, which ended Feb. 29, was highlighted by total revenue of $523 million, an increase of 31% over revenue posted in fiscal 2007. Its subscription revenue, which is amassed from support licenses, was $449.8 million, a 32% increase over 2007. Net income for the fiscal year was up 28%, from $59.9 million to $76.7 million.*
>
> *http://www.linuxworld.com/news/2008/032808-red-hat.html*

Red Hat as a company owns JBoss.

Oracle Unbreakable Linux

Oracle's (*www.oracle.com/database/feature_db_dbleadership.html*) –

- Linux market share was 82.6% in 2006, up from 80.6% in 2005.
- Linux revenues were up 95%.
- RDBMS business on Linux was $1.94 billion in 2006, up from $1 billion.

Oracle Unbreakable Linux is a support program.

- Supports RHEL.
- … but aimed at Windows?

www.oracle.com/technologies/linux/index.html

Novell - SuSE

Novell is a long-time Microsoft rival.

SuSE focuses on both desktops and servers.

- SLED=SuSE Linux Enterprise Desktop.
  - Targeting Windows Vista.
- OpenSuSE is free.
- Xen virtualization.

Software frameworks

A software framework is a reusable design for a software system (or subsystem).

en.wikipedia.org/wiki/Software_framework

Software frameworks provide –

- reusable software components.
- a means of plugging in your own components.
- a mechanism for component interoperability.

Software frameworks are the foundation of modern enterprise application development.

Java – Java EE

Java –

- originated as a programming language and run-time environment for device-embedded software.
- got attention as a tool for developing rich Internet client applications that run in the Web browser.
  - Plain-old HTML and JavaScript won that role however, supplemented by Flash and some Microsoft ASP.
- became a tool of choice for server-side programming.
- has only recently been released as OS by Sun Microsystems.

Java – Java EE

Java specification development is –

governed by the Java Community Process (JCP).

See *jcp.org/en/home/index*

determined by JSRs, Java Specification Requests.

Java application development may be performed within a number of OS frameworks.

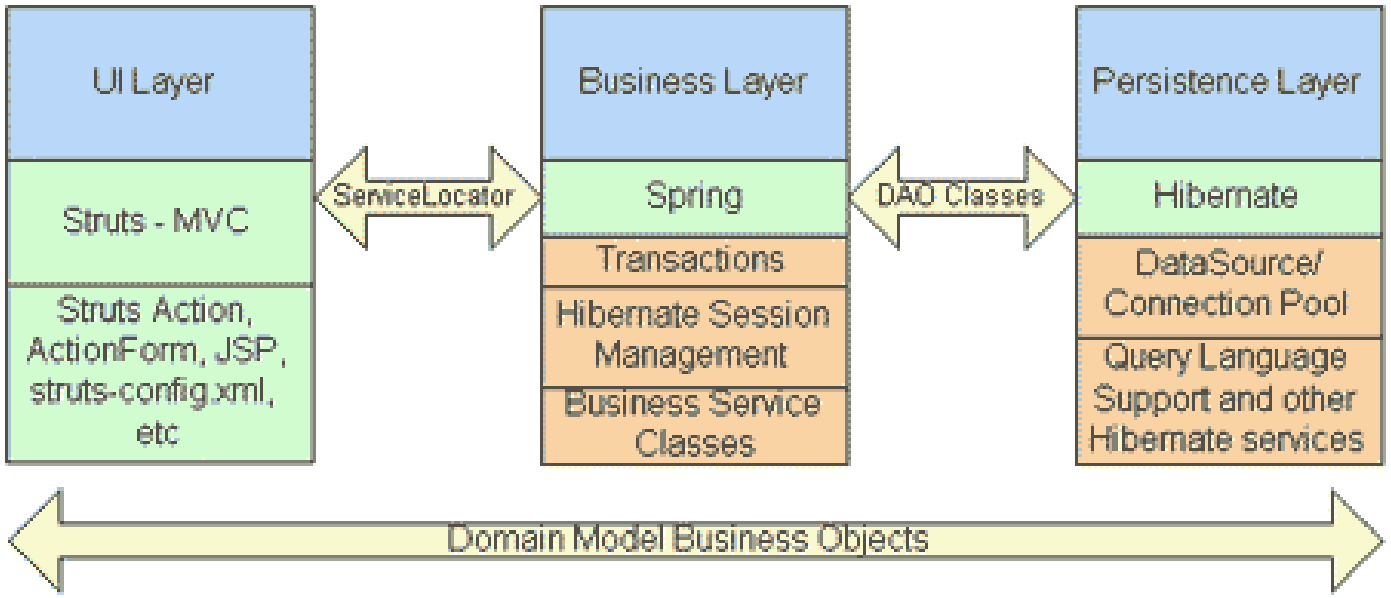
MVC oriented: Struts-Spring-Hibernate.

Component oriented: JavaServer Faces, Tapestry.

Eclipse.

Java – Java EE

One can combine frameworks, for instance for Web development –



www.onjava.com/pub/a/onjava/2004/04/07/wiringwebapps.html

Java – Java EE

Java EE is the Java Platform, Enterprise Edition.

- Java EE was formerly known as Java 2 Enterprise Edition (J2EE).
- Java EE 6 is the current version and includes:
  - Enterprise JavaBeans 3.0 with Plain Old Java Objects (POJOs).
  - [Object] Persistence API.
  - Enhanced web services and support for Service Oriented Architectures (SOAs).
  - JavaServer Faces (JSF); Java Server Pages (JSP), and the JSP Standard Tag Library (JSTL).

See *java.sun.com/javaee/technologies/*

Spago

J2EE framework.

GNU LGPL license.

Hosted by ObjectWeb Consortium.

Sponsored by Engineering Ingengnerla Informatica (EII) in Rome.

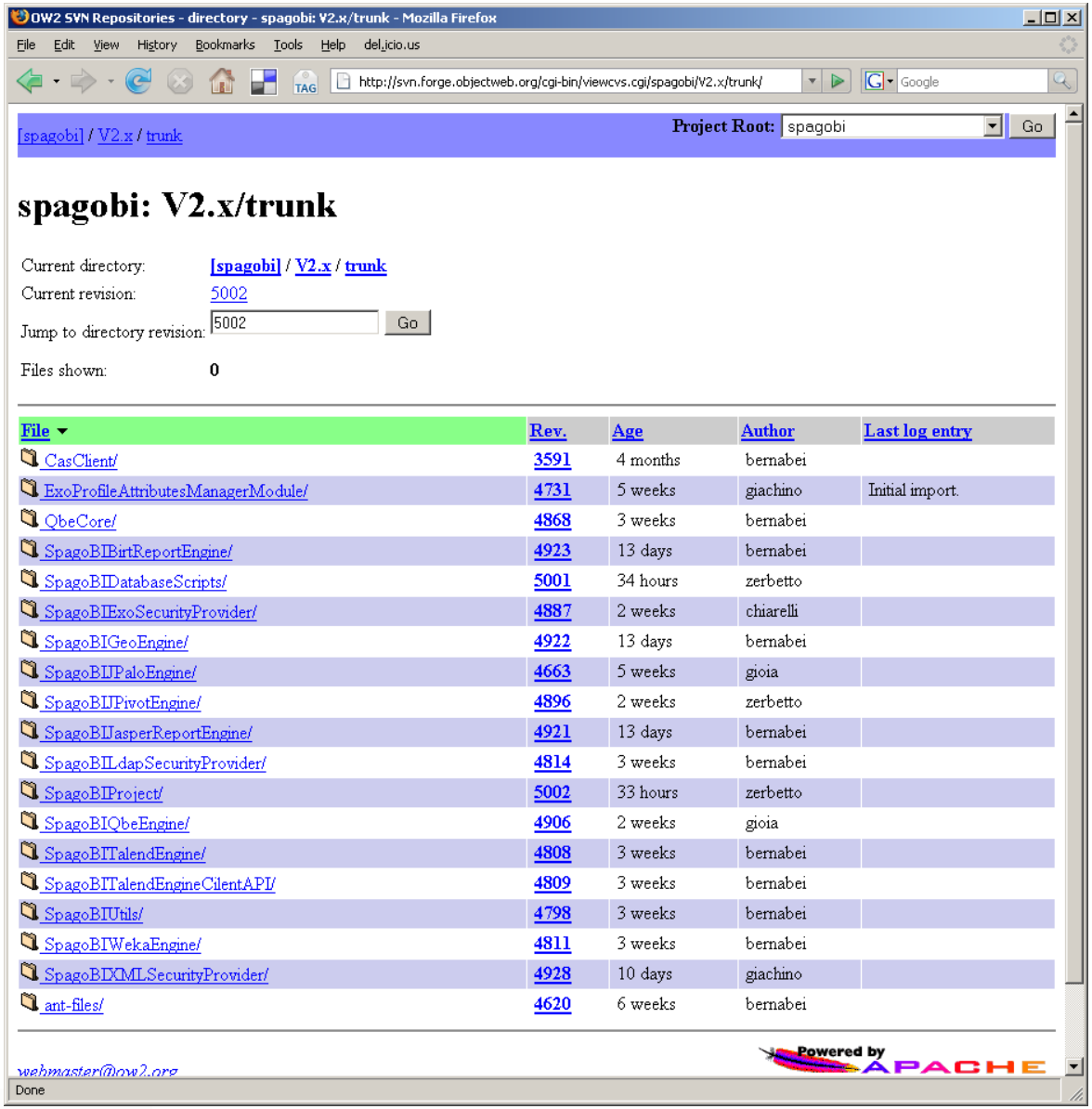
spago.eng.it/

Spago World community.

spagoworld.com/

SpagoBI business intelligence project.

spagobi-info.eng.it/



.Net

.Net is Microsoft's application development framework. .Net is –

Currently out in version 3.0.

"Microsoft software for connecting information, people, systems, and devices. .NET provides XML-based interoperability and is being incorporated across Microsoft clients, servers, services, and tools. For example, products like Microsoft Windows and Microsoft Office will use .NET to connect with other systems and applications. For developers, .NET is manifested in the programming model delivered in the Microsoft .NET Framework."

See *www.microsoft.com/net/default.aspx*

.Net

The open source Mono project –

"Provides the necessary software to develop and run .NET client and server applications on Linux, Solaris, Mac OS X, Windows, and Unix."

Commercially supported by Novell.

See *www.mono-project.com/Main_Page*

LAMP

LAMP is an alternative, open-source stack consisting of –

1. Linux operating system.
2. Apache Web server.
3. MySQL database.
4. Perl/Python/PHP programming.

The programming layer includes application frameworks.

LAMP

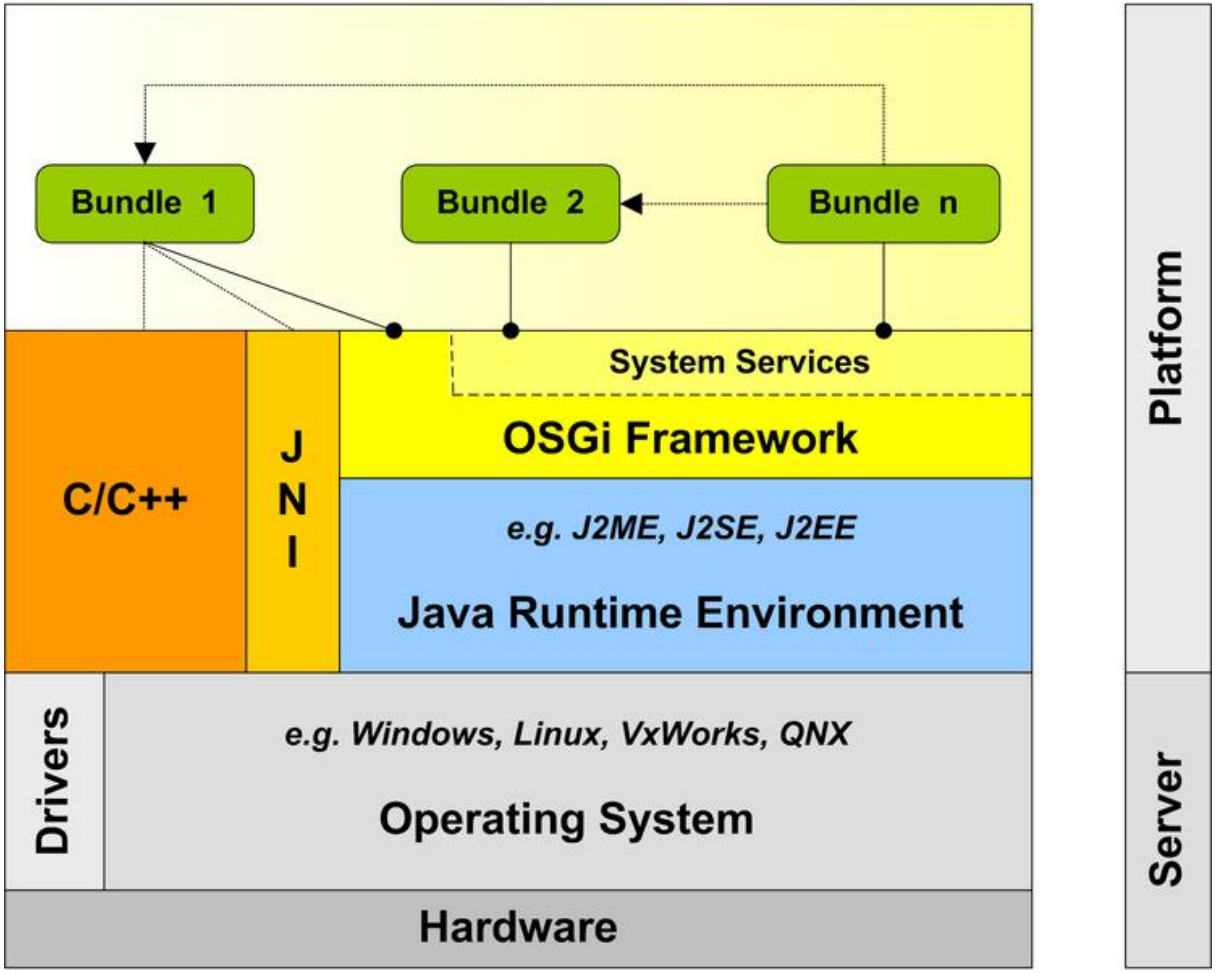
- LAMP essentially provides an alternative Web development/delivery platform.
  - PostgreSQL easily substitutes for MySQL.
  - Ruby substitutes for the programming-layer Ps.
- Each of the programming-layer alternatives may be packaged within one or more software frameworks.
  - Zend is one of many frameworks for PHP.
  - Rails is a noted framework for Ruby development.
  - Django is one of many frameworks for Python.

Eclipse

Eclipse–

- is a "universal tool platform."
- offers an integrated development environment (IDE) for software development.
- provides a framework, a "rich-client platform" (RCP), for applications.
- implements the Open Services Gateway initiative (OSGi) framework for service/component interoperations.

Framework integration



en.wikipedia.org/wiki/Image:Osgi_layer.png

Application servers

An application server is/runs middleware serving as a component container for –

- Clients/presentation.
- Business logic.
- Data access.

Open source examples include –

- Apache Tomcat.
- ObjectWeb JOnAS.
- Sun GlassFish.
- Zope (Python).

Apache

Started as a HTTP (Web) server.

Became a nexus for Apache-licensed projects that adhere to "The Apache Way":

- collaborative software development.
- commercial-friendly standard license.
- consistently high quality software.
- respectful, honest, technical-based interaction.
- faithful implementation of standards.
- security as a mandatory feature.

Apache Tomcat & Geronimo

Tomcat is a container for –

- Java servlets.
- Java Server Pages.

Geronimo supports the entire Java EE stack –

- Servlet container.
- Enterprise Java Beans (EJB) container.
- Messaging via Java Message Service (JMS) API.
- Java Connector Architecture (JCA) container.

IBM WebSphere Application Server Community Edition = Geronimo.

JBoss

The leading commercial-OS app server; LGPL.

JEMS = JBoss Enterprise Middleware.

- Provides applications services within an SOA.

Supports many projects including –

- Enterprise Service Bus.
- Portal.
- jBPM business process management engine.
- Seam application framework for POJOs/EJBs.

Architecture trends

Service Oriented Architecture (SOA) –

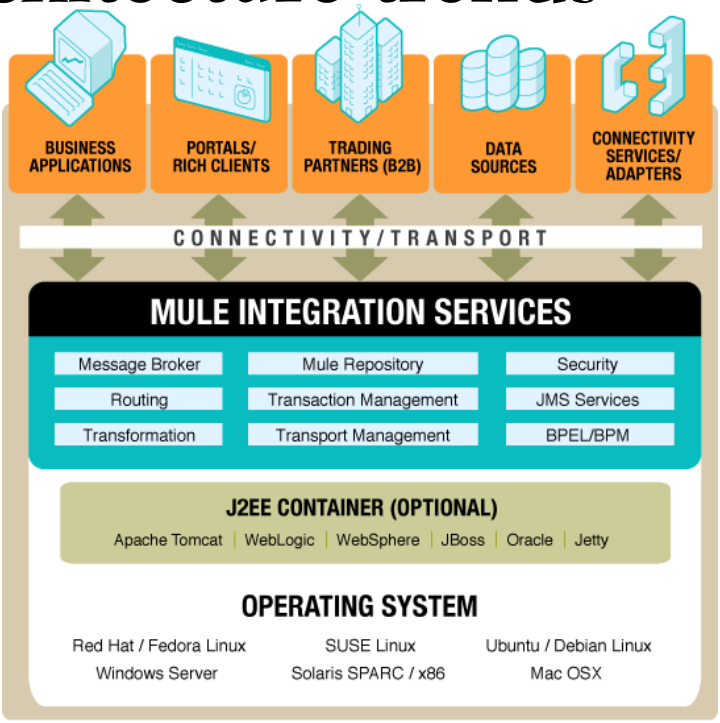
- Prizes interoperability and therefore standards.
- Typically includes process orchestration.
- Offers an Enterprise Service Bus (ESB) as one interconnection option.

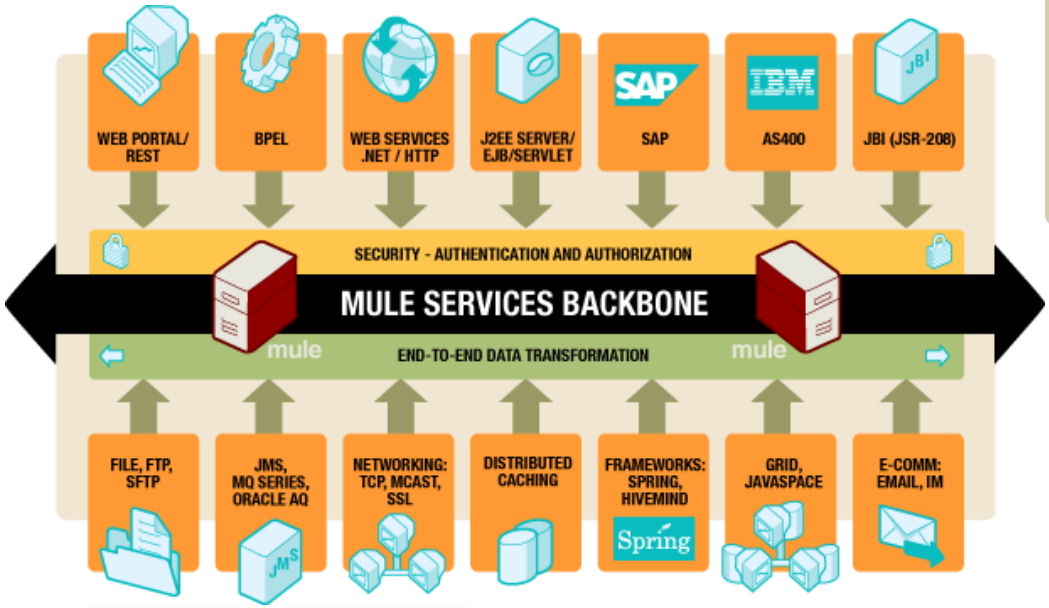
Interoperation –

- A software development framework supports interoperation within an application.
- An app server supports application interoperation.
- SOA adds an abstraction layer.

Architecture trends

OA/ESB architecture illustration from the open-source Mule platform.





The Mule license is based on Mozilla's. The company sells support subscriptions.

www.mulesource.com/products/,
www.mulesource.com/solutions/

Alta Plana

Software development

To create software you need –

- Team management and development approach.
- Programming tools.
- Development environment.
- Repository with version and build control.
- Test tools.

Software development

Traditional software development focuses on –

- Quality assurance via attention to –
  - Processes and procedures.
  - Knowledge management and learning.
  - Management practices.
- Quality control via –
  - Verification and validation.
  - Structured testing processes: unit, integration-system, acceptance.

OS development deviates from the traditional model given the nature of teams, coordination.

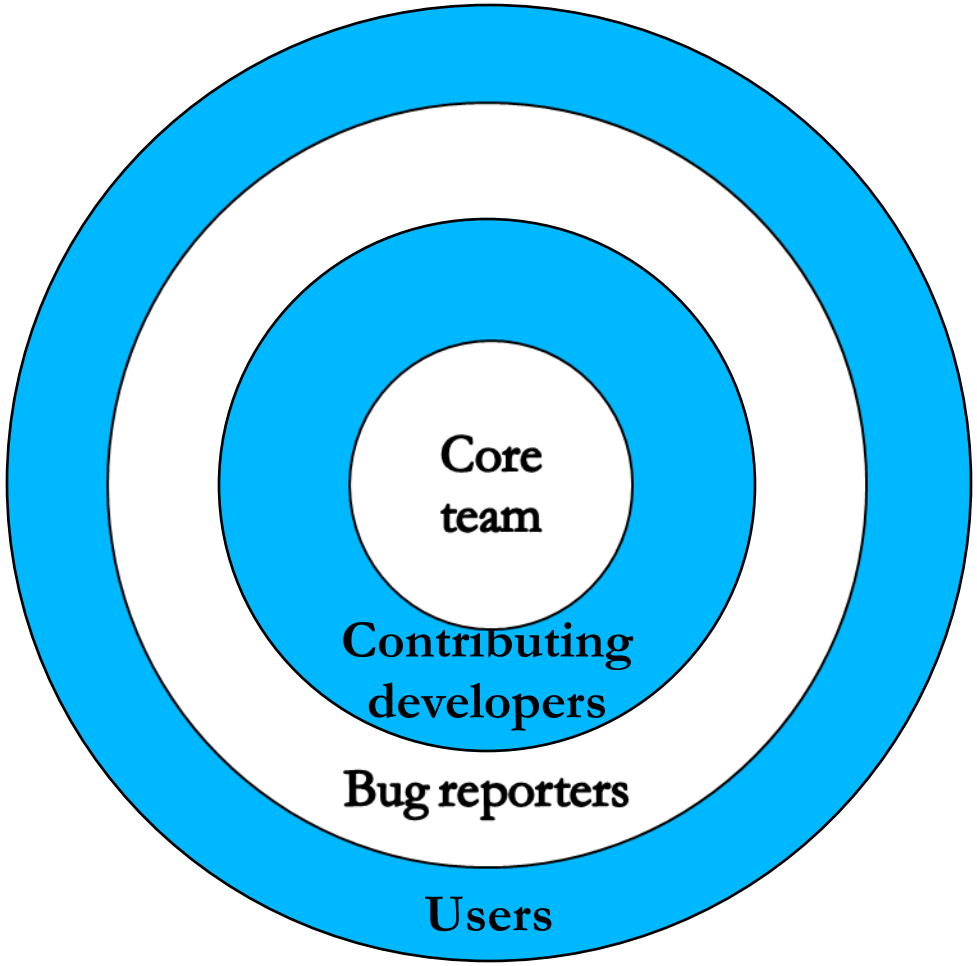
Software development

Closed source	Open source
Well defined development methodology	Development methodology often not defined or documented
Extensive project documentation	Little project documentation
Formal, structured testing and quality assurance methodology	Unstructured and informal testing and quality assurance methodology
Analysts define requirements	Programmers define requirements
Formal risk assessment produces – monitored and management throughout project	No formal risk assessment process
Measurable goals used throughout project	Few measurable goals
Defect discovery from black-box testing as early as possible	Defect discovery from black-box testing late in the process
Empirical evidence regarding quality used routinely to aid decision making	Empirical evidence regarding quality isn't collected
Team members are assigned work	Team members choose work
Formal design phase is carried out and signed off before programming starts	Projects often go straight to programming
Much effort put into project planning and scheduling	Little project planning and scheduling

Mark Aberdour, "Achieving Quality in Open Source Software," *IEEE Software*, vol. 24, no. 1, pp. 58-64, Jan/Feb, 2007.

Software development

"The onion model of a sustainable software development community."



Mark Aberdour, "Achieving Quality in Open Source Software," *IEEE Software*, vol. 24, no. 1, pp. 58-64, Jan/Feb, 2007.

Programming languages

GCC (GNU Compiler Collection)

Includes C++, Fortran, Java.

Perl

Python

Ruby

Popular Rails Web-development framework.

Java

Tcl (Tool Control Language)

Eclipse development environment

There are many development environments.

Our primary interest will be in Eclipse.

Originated as an IBM integrated development environment (IDE) designed to replace the Visual Age toolset.

Released to open source in 2001.

Managed by the Eclipse Foundation.

www.eclipse.org/

Eclipse development environment

Eclipse projects include –

- Business Intelligence and Reporting Tools (BIRT)
- Data Tools Platform
- Device Software Development Platform
- Eclipse Modeling Project
- SOA Tools
- … and technology, programming and Web tools, and more.

Database systems

"From 2005 to 2006, open source vendor revenue grew 36.3% to $140 million, compared to the overall market growth of 12.2%... This growth in revenue will continue during the next five years at more than 40%, exceeding $1 billion by 2012"

– Donald Feinberg, Gartner principal analyst, in *Open Source in Database Management Systems, 2008*.

Database systems

Four data-management modes:

- Content publishing.
- Transactional.
- Analytical.
- Operational (embedded).

Database systems

"Traditional":

- Ingres 

- PostgreSQL 

- MySQL 

- Firebird 

Java:

- Apache Derby 

- HSQLDB 

Embedded + Java:

- Oracle Berkeley DB (Sleepycat)

Database systems

Analytical:

(C-Store

(Commercialized as Vertica. 2006 0.2 version available.)

MonetDB (*monetdb.cwi.nl/*)

DBMS for high-performance applications in data mining, OLAP, GIS, XML Query, text and multimedia retrieval.

LucidDB (*luciddb.org/*)

Designed for LucidEra hosted (SaaS) BI.

Eigenbase (*www.eigenbase.org/*)

DBMS platform. Used by SQLStream and LucidDB.

Release 0.7.3 in March.

Database systems

MonetDB:

Monet Public License based on Mozilla PL.

Academic/research project of CWI Netherlands.

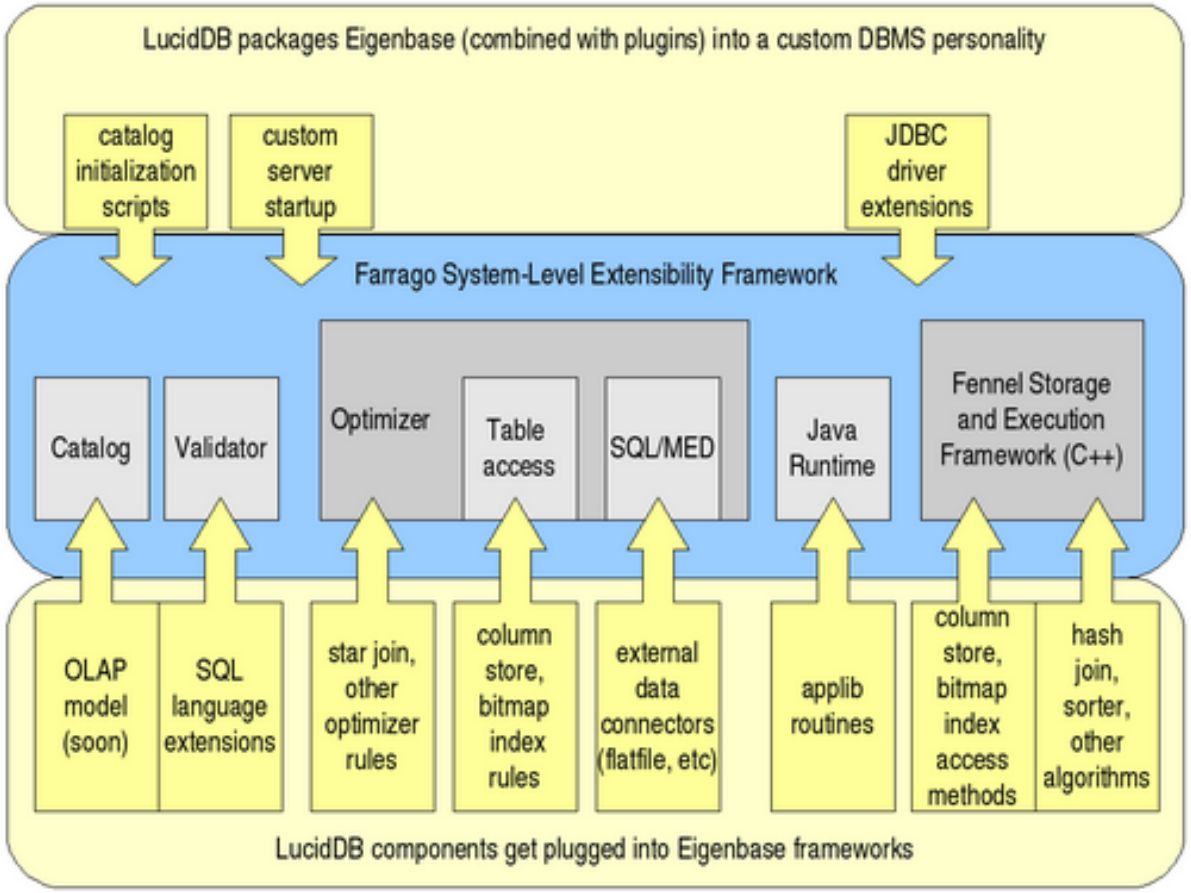
Supports SQL, Xquery, APIs & application bindings.

LucidDB:

GPL v2 licensing; LGPL v2.1 for the client.

"The first and only open-source RDBMS purpose-built entirely for DW and BI. It is based on architectural cornerstones such as column-store, bitmap indexing, hash join/aggregation, and page-level multiversioning."

Database systems



http://www.luciddb.org/

LucidDB is the DBMS for LucidEra SaaS BI.

Alta Plana

Firebird

Open source continuation of the Interbase RDBMS, *www.firebirdsql.org/*.

- (1984) Groton Database Systems → (1986-91) Ashton Tate → (1991) Borland → (1998) Inprise / Borland → (2000) code fork, open source.
- Innovative multi-generational architecture.
- InterBase Public Licence, a variant of Mozilla Public Licence (MPL).
- Supported by the Firebird Foundation.
- IBPhoenix development portal.

www.ibphoenix.com

PostgreSQL

Robust, object-relational DBMS.

Dates to 1986 Postgres U.C. Berkeley project.

Many prominent users.

Very extensive developer community.

BSD license.

- Use, modify and distribute PostgreSQL in any form, open or closed source.
- Commercialized by Stonebraker as Illustra in the mid-'90s.

PostgreSQL

Press kit in Italian:

www.postgresql.org/about/press/presskit82.html.it

PostgreSQL foundry, *pgfoundry.org/*

The PostgreSQL Development Group's site for developing and publishing PostgreSQL-related software that is not part of the core product.

Other tools development at *gborg.postgresql.org/*.

PostgreSQL

Features –

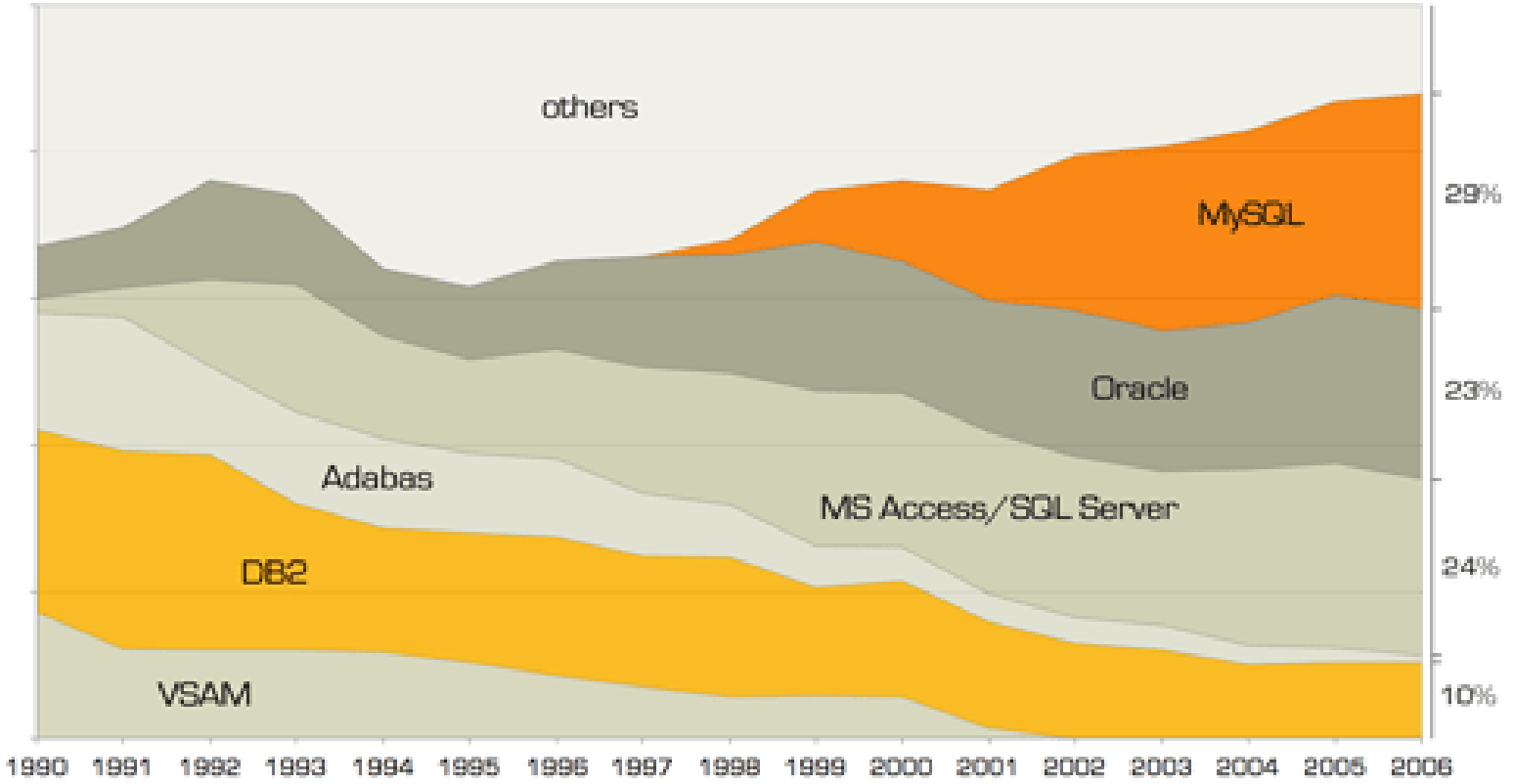
- Runs stored procedures in more than a dozen programming languages, including Java, Perl, Python, Ruby, Tcl, C/C++, and its own PL/pgSQL, which is similar to Oracle's PL/SQL.
- Interfaces for Java (JDBC), ODBC, Perl, Python, Ruby, C, C++, PHP, Lisp, Scheme, and Qt.
- Triggers and stored procedures can be written in C and loaded into the database as a library.
- Framework for custom data types along with supporting functions and operators.

MySQL

MySQL is the most popular open-source DBMS.

- Early success for Web content as part of the LAMP stack.
- Later success for transactional, analytical, and operational (embedded) systems.
- Targets developers, ISVs, VARs, hardware vendors, and network appliance.

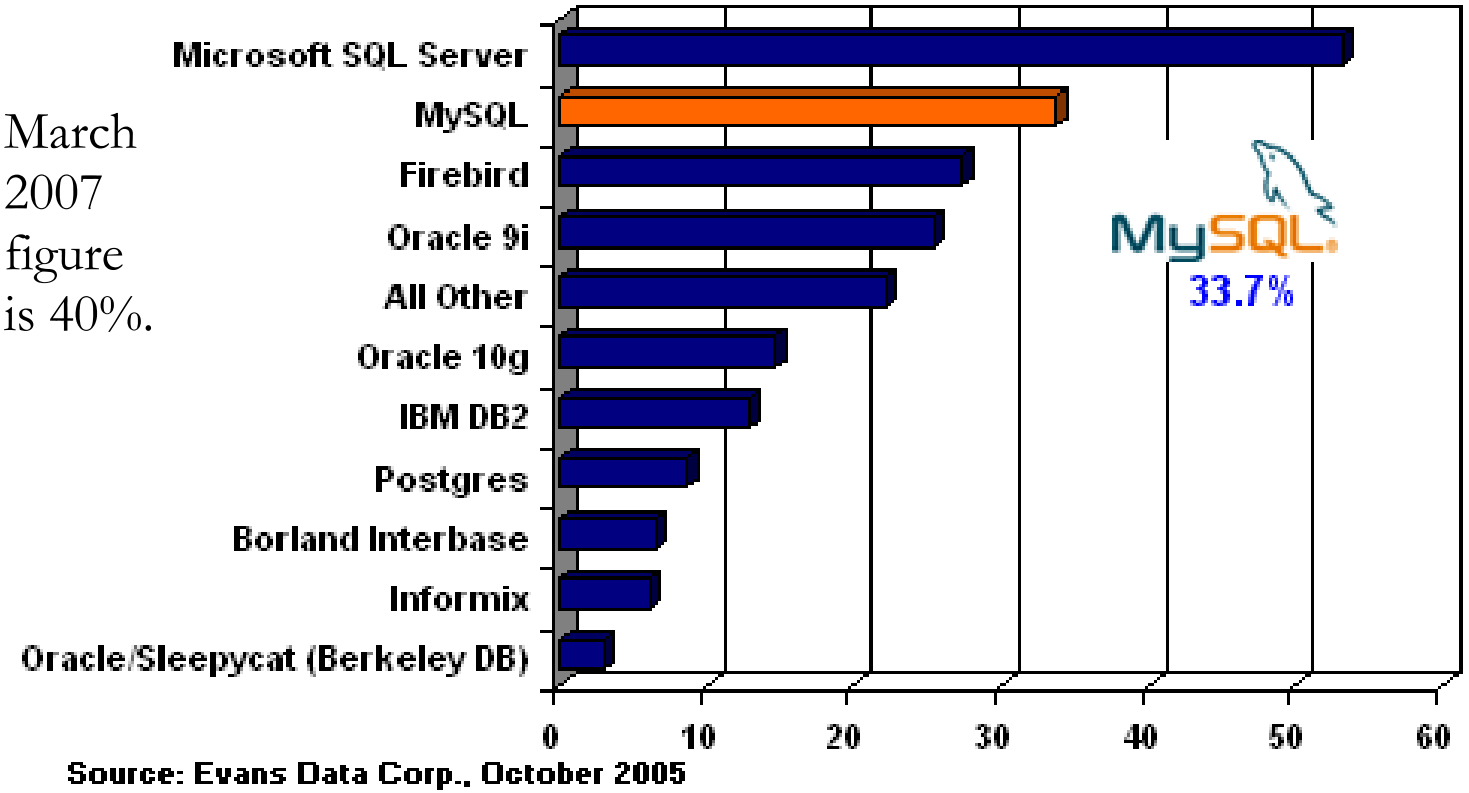
MySQL market share



Source: JoinVision E-Services GmbH, July 2006, survey of the **JoinVision community**, *http://www.mysql.com/why-mysql/marketshare/*

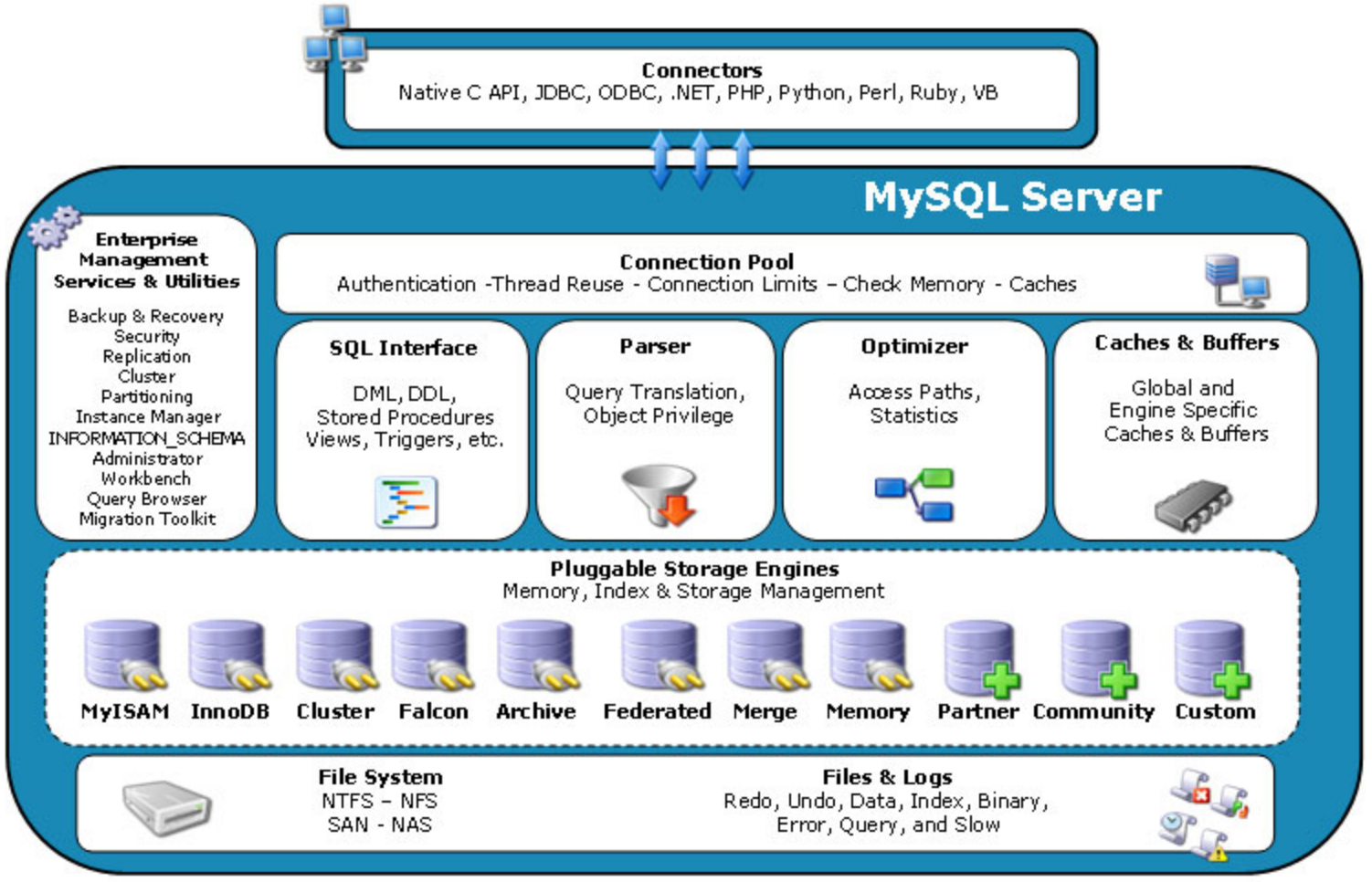
MySQL market share

March 2007 figure is 40%.



Evans Data Database Development Survey, October, 2005, examining the usage patterns and attitudes of **database developers**, *www.evansdata.com/survey_database_topical.shtml*

MySQL architecture



MySQL resources

Italian site:

www-it.mysql.com/

Italian white papers at:

www-it.mysql.com/why-mysql/white-papers/

My white paper, "MySQL 5.0 - Ready for Prime Time Business Intelligence with Pentaho" is at:

dev.mysql.com/tech-resources/articles/mysql_5.0_pentaho.html

PostgreSQL enhancements

EnterpriseDB:

- PostgresPlus open source.
- Company acquired ExtenDB MPP technology in 2007, releasing it as GridSQL to open source.
- PostgresPlus Advanced Server is not open source.

Greenplum database:

- Parallelized, "shared nothing" MPP architecture, designed for DW, but not open source.
- Supports very large DWs – 100s of TB.
- Appliance with Sun Microsystems.

DBMS tools

MySQL GPL tools include –

Administrator, Query browser, Migration utility.

SQuirrel SQL Client (*http://www.squirrelsql.org/*) –

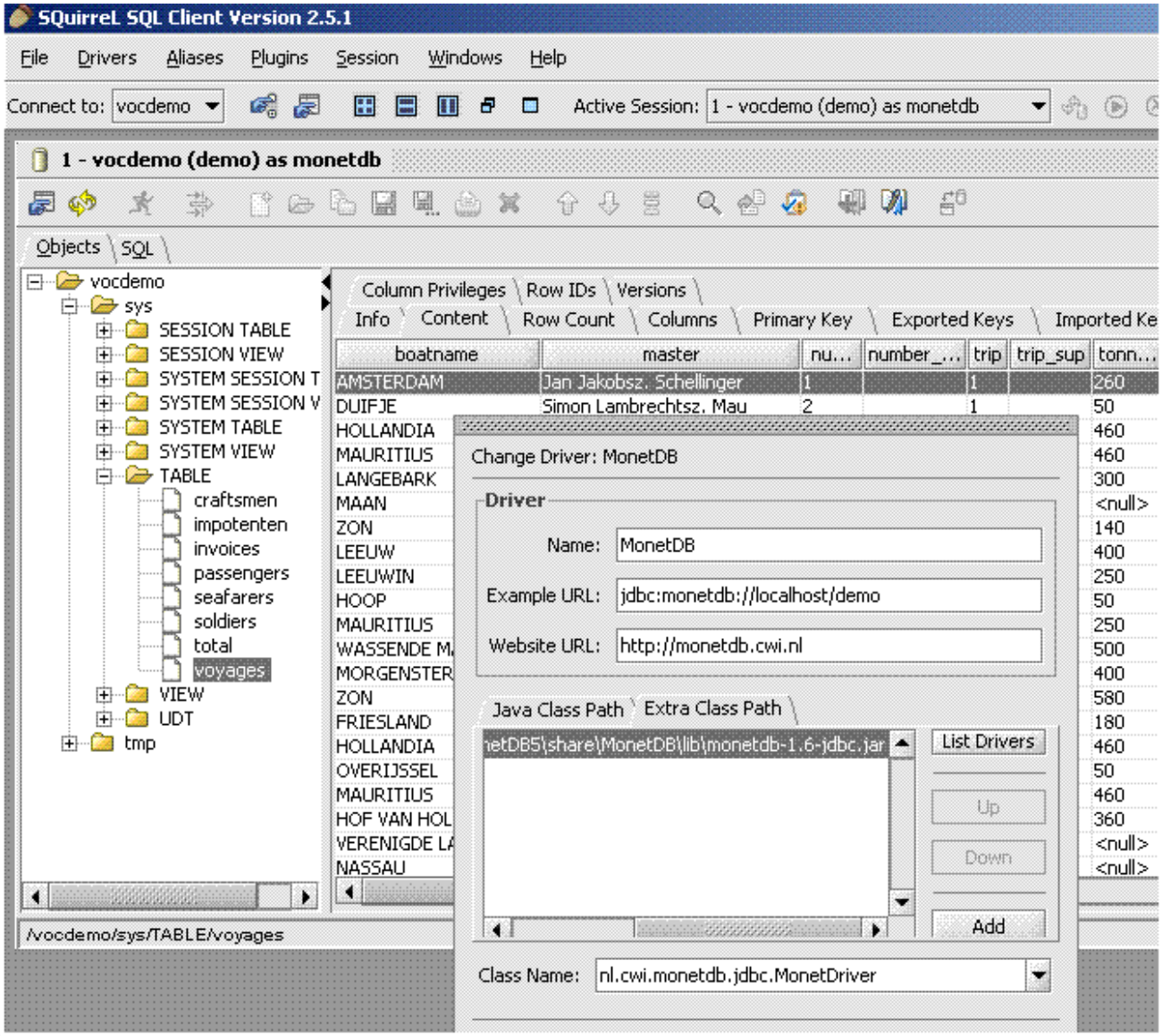
GPL+LGPL licensing; Java w/ JDBC DBMS access.

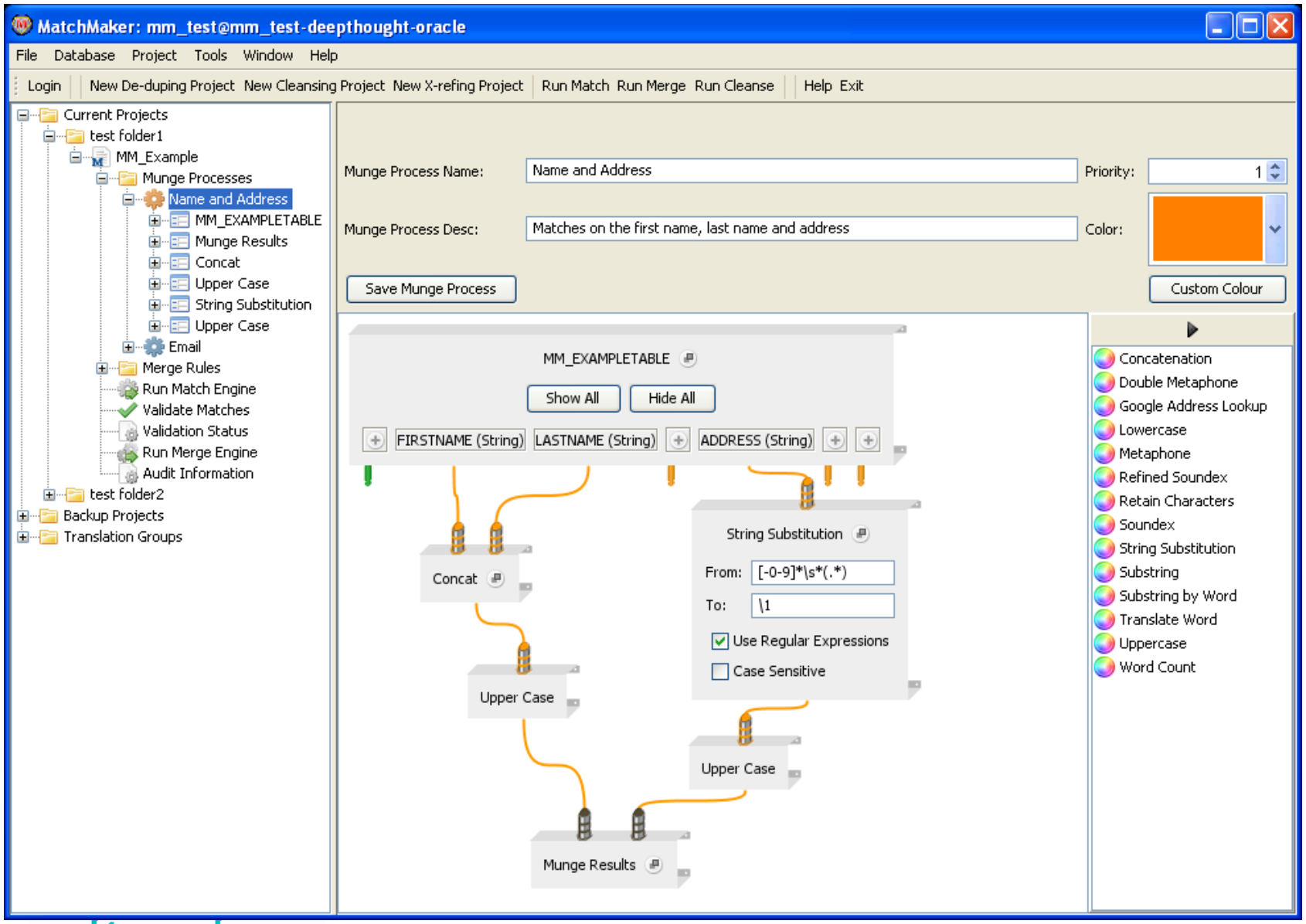
From SQL Power (*http://www.sqlpower.ca/*) –

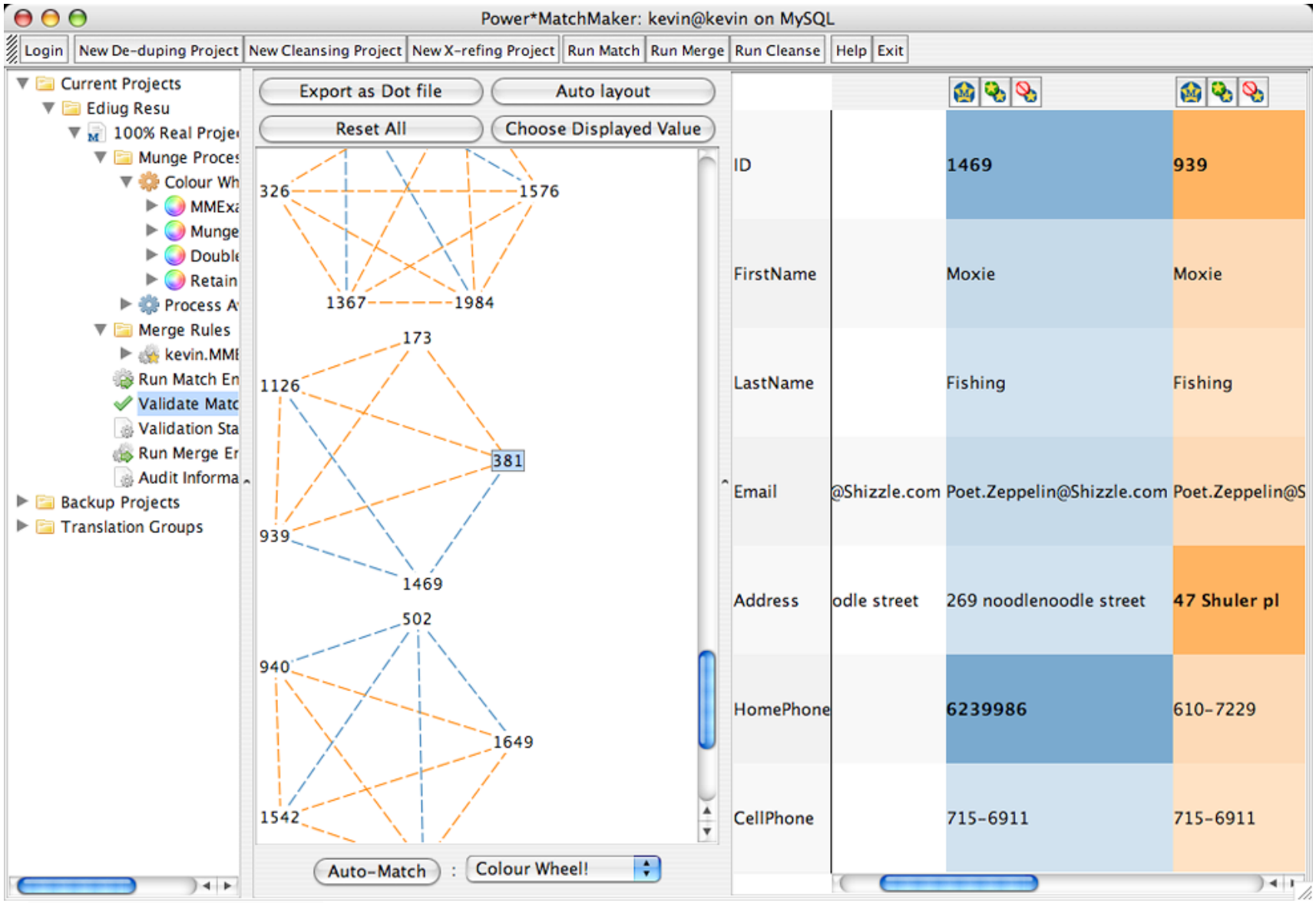
Power*Architect data modeling tool, BSD license.

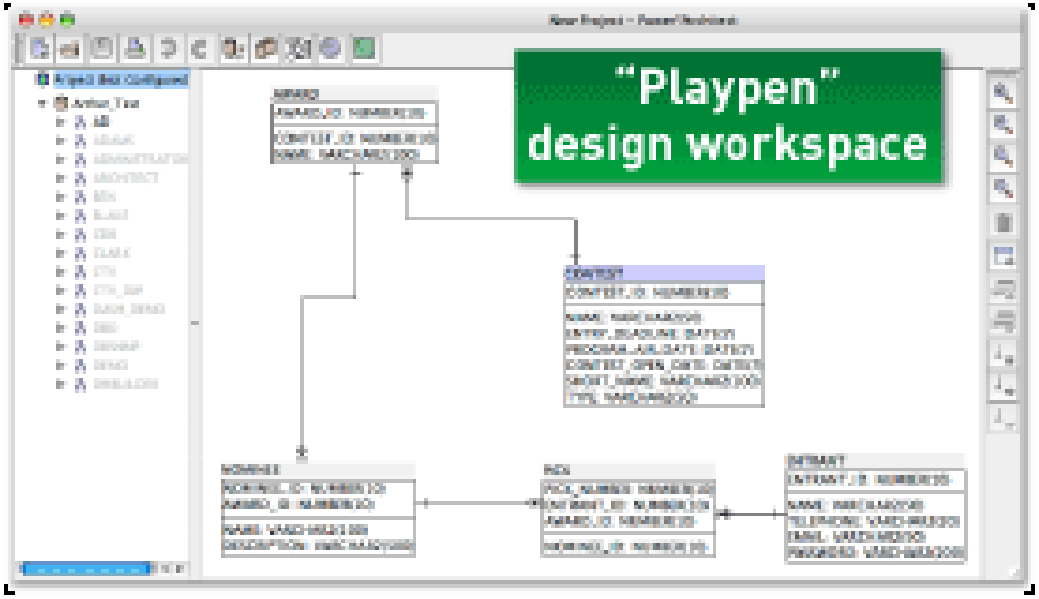
Power*MatchMaker data cleansing and deduplication, GPL v3 license.

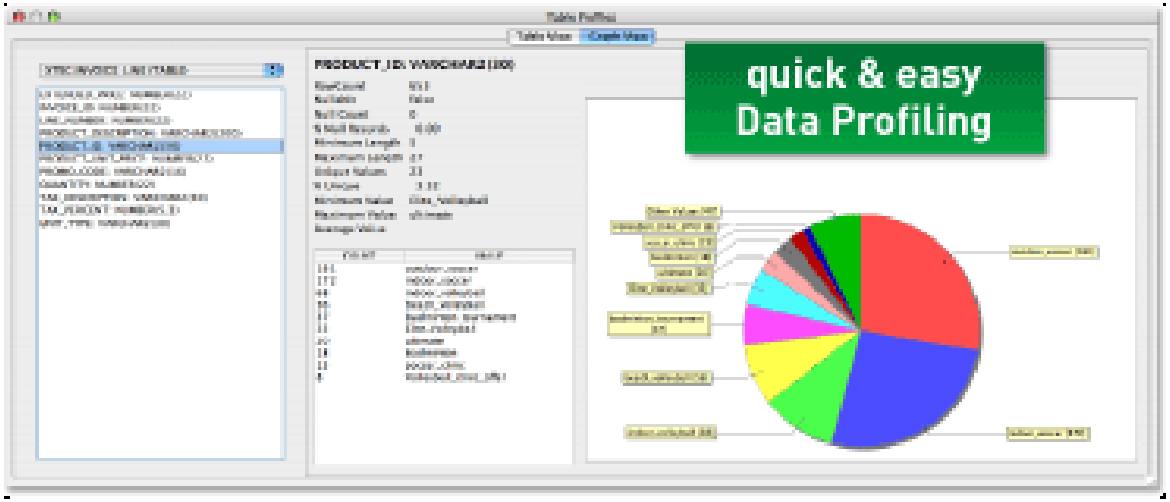
JDBC drivers for PostgreSQL, MS SQL Server 2005, MySQL, HSQLDB











DBMS tools

Free but not open source:

- Toad for MySQL.

  www.quest.com/toad-for-mysql/

Not free, not open source:

- MicroOLAP tools for MySQL & PostgreSQL.

  microolap.com/

- MySQL Maestro.

  www.sqlmaestro.com/products/mysql/maestro/download/

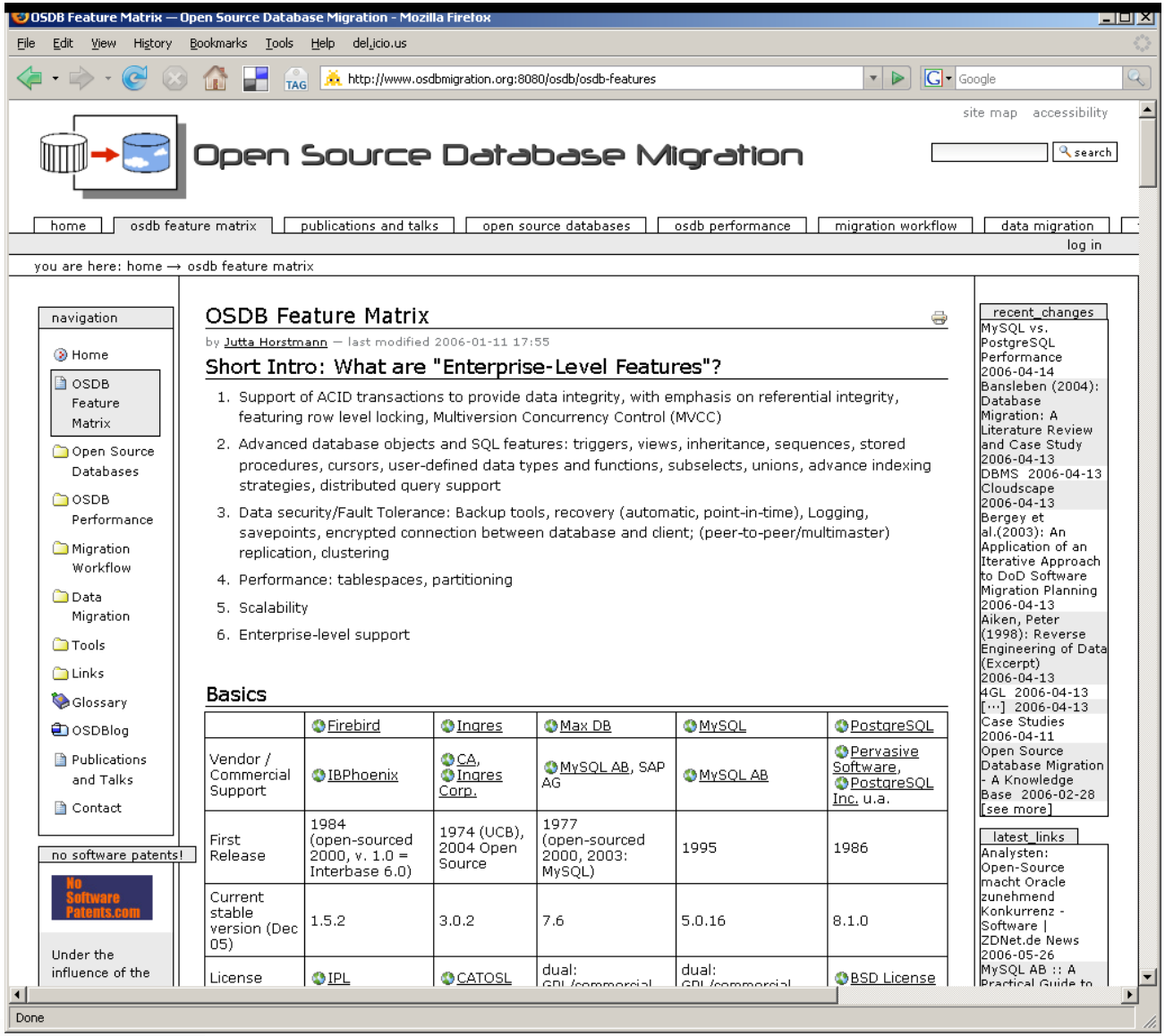
- PostgreSQL Maestro.

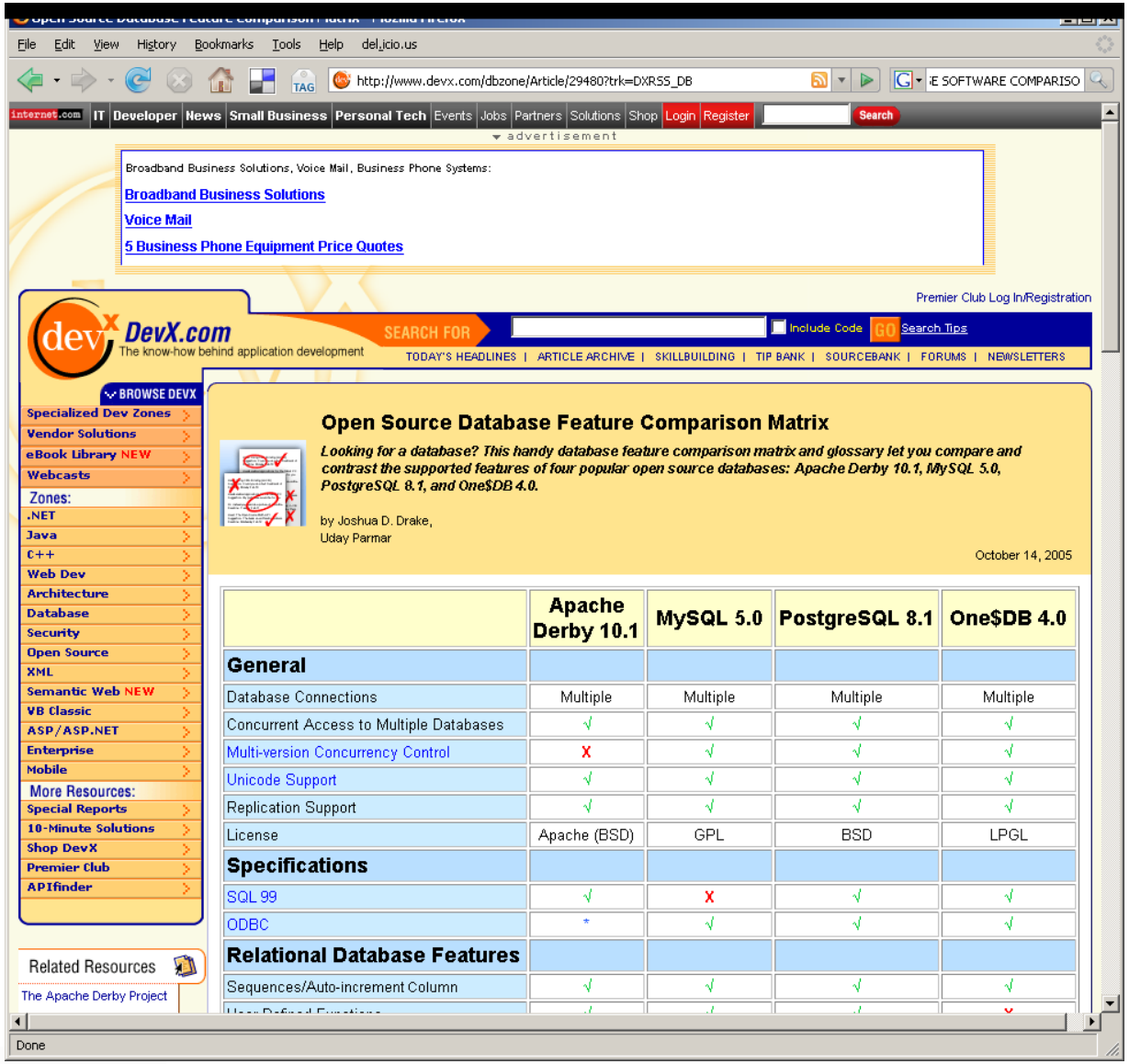
  www.sqlmaestro.com/products/postgresql/maestro/

Database systems

A variety of sources can help you decide which open-source DBMS is for you.

- Open Source Database Migration is a great start although a bit out of date:
  www.osdbmigration.org:8080/osdb/osdb-features
- DevX site is more recent but less comprehensive:
  www.devx.com/dbzone/Article/29480?trk=DXRSS_DB





File Edit View History Bookmarks Tools Help del.icio.us

http://www.devx.com/dbzone/Article/29480?trk=DXRSS_DB

internet.com | IT | Developer | News | Small Business | Personal Tech | Events | Jobs | Partners | Solutions | Shop | Login | Register | Search

advertisement

Broadband Business Solutions, Voice Mail, Business Phone Systems:

Broadband Business Solutions

Voice Mail

5 Business Phone Equipment Price Quotes

Premier Club Log In/Registration

DevX.com
The know-how behind application development

SEARCH FOR | Include Code | GO | Search Tips

TODAY'S HEADLINES | ARTICLE ARCHIVE | SKILLBUILDING | TIP BANK | SOURCEBANK | FORUMS | NEWSLETTERS

BROWSE DEVX

- Specialized Dev Zones
- Vendor Solutions
- eBook Library NEW
- Webcasts

Zones:

- .NET
- Java
- C++
- Web Dev
- Architecture
- Database
- Security
- Open Source
- XML
- Semantic Web NEW
- VB Classic
- ASP/ASP.NET
- Enterprise
- Mobile

More Resources:

- Special Reports
- 10-Minute Solutions
- Shop DevX
- Premier Club
- APIfinder

Related Resources

The Apache Derby Project

Open Source Database Feature Comparison Matrix

Looking for a database? This handy database feature comparison matrix and glossary let you compare and contrast the supported features of four popular open source databases: Apache Derby 10.1, MySQL 5.0, PostgreSQL 8.1, and One$DB 4.0.

by Joshua D. Drake,
Uday Parmar

October 14, 2005

	Apache Derby 10.1	MySQL 5.0	PostgreSQL 8.1	One$DB 4.0
General				
Database Connections	Multiple	Multiple	Multiple	Multiple
Concurrent Access to Multiple Databases	√	√	√	√
Multi-version Concurrency Control	X	√	√	√
Unicode Support	√	√	√	√
Replication Support	√	√	√	√
License	Apache (BSD)	GPL	BSD	LPGL
Specifications				
SQL 99	√	X	√	√
ODBC	*	√	√	√
Relational Database Features				
Sequences/Auto-increment Column	√	√	√	√

Done

Office productivity tools

Office productivity suites consist of desktop:

- Word processing.
- Spreadsheet.
- Presentation.
- E-mail, directory, calendar.
- (Database.)

Office productivity tools

The market:

- Microsoft achieved and maintains a dominant position by –
  - Beating Lotus & WordPerfect from DOS to Windows.
  - Strong Windows and inter-tool integration.
  - Effective (and predatory) relationship with PC makers.
- Competitors (WordPerfect, etc.) hang on by offering compatibility and low prices.
- OpenOffice.org from Sun is open source.
- Google docs & spreadsheets are free; Zoho is free for individuals: hosted but not OS.

Office productivity tools

OpenOffice.org:

"In Italy, where I have the updated numbers, we are hitting today - maybe while I'm writing this post - one million downloads since January 1st, 2008 (over 350.000 since the announcement of OOo 2.4 in late March). Although we don't have Microsoft figures for Office 2007, we estimate a maximum of 1.8 million licenses sold in 2008."

www.italovignoli.org/

Hosted office tools

- Hosted versions provide an alternative delivery model but do not redefine the concept.
  - Also known as Software as a Service (SaaS) and On Demand software.
  - They foster collaboration, location independence.
  - An emerging business model is to host OS software for a fee.
    - Reason for the Affero General Public License.
- If your primary reason for considering OS is cost, consider hosted/SaaS instead.

Google spread-sheet: hosted, not open source.

ft1dv10se22ma221mi221100yr2000 - Google Docs & Spreadsheets - Mozilla Firefox

File Edit View History Bookmarks Tools Help del.icio.us

http://2b8.spreadsheets.google.com/ccc?id=o0710115388917721994

Google Docs & Spreadsheets

seth.grimes@gmail.com | Docs Home | Help | Sign Out

ft1dv10se22ma221mi221100yr2000 Autosaved at Mar 20, 2007 2:52:41 PM EDT

Automatically Saved Save & close

File Edit Sort Formulas Revisions Preview Print Discuss Collaborate Publish

Format Align Insert Delete Wrap Text Merge across

	A	B	C	D	E	F	G
18	**Income Tax Returns of Active Corporations with Accounting Periods Ended July 2000 Through June 20001 --**						
19							
20	**Balance Sheet, Income Statement, and Selected Items, by Minor Industry, by Size of Total Assets**						
21							
22	**(All figures are estimates based on samples--money amounts and size of total assets are in thousands of dollars)**						
23							
24							
25		**Total**	**Zero Assets**	**1 to under 100**	**250 to under 500**	**500 to under 1,000**	**1,000 to u 5,000**
26	**1 Number of returns**	885	9	*247	*200	*121	*108
27	**2 Total assets**	519,234,712	0	*5,393	*66,721	*76,834	*264,151
28	**3 Cash**	9,451,713	0	*2,530	*56,581	*14,053	*14,804
29	**4 Notes and accounts receivable**	42,609,963	0	0	86	*21,963	*6,330
30	**5 Less: Allowance for bad debts**	697,575	0	0	0	0	0
31	**6 Inventories**	4,493,505	0	0	0	*6,395	0
32	**7 U.S. govt. obligations, total**	1,426,285	0	0	0	0	0
33	**8 Tax-exempt securities**	24,777	0	0	0	0	0
34	**9 Other current assets**	42,892,704	0	0	*5,699	*1,768	*2,729
35	**10 Loans to shareholders**	239,587	0	0	*1,780	*2,542	*6,663
36	**11 Mortgage and real estate loans**	1,419,012	0	0	0	0	0
37	**12 Other investments**	87,033,454	0	0	377	0	*59,222
38	**13 Depreciable assets**	392,914,921	0	*17,944	*29,519	*45,831	*235,553
39	**14 Less: Accumulated depreciation**	156,647,479	0	*15,081	*27,384	*17,509	*80,205
40	**15 Depletable assets**	1,281,715	0	0	0	0	0
41	**16 Less: Accumulated depletion**	666,393	0	0	0	0	0
42	**17 Land**	2,048,645	0	0	0	0	*6,814
43	**18 Intangible assets (Amortizable)**	15,265,890	0	0	0	*1,114	*17,103
44	**19 Less: Accumulated amortization**	1,717,630	0	0	0	*279	*4,871
45	**20 Other assets**	77,861,617	0	0	64	*955	9
46	**21 Total liabilities**	519,234,712	0	*5,393	*66,721	*76,834	*264,151
47	**22 Accounts payable**	29,119,949	0	0	927	*4,918	*20,222
48	**23 Mort. notes, and bonds under 1 yr**	34,523,586	0	7,407	0	0	*190
49							

Add Sheet ft1dv10se22ma221mi221100yr2000 9 V1.1.9j

Done

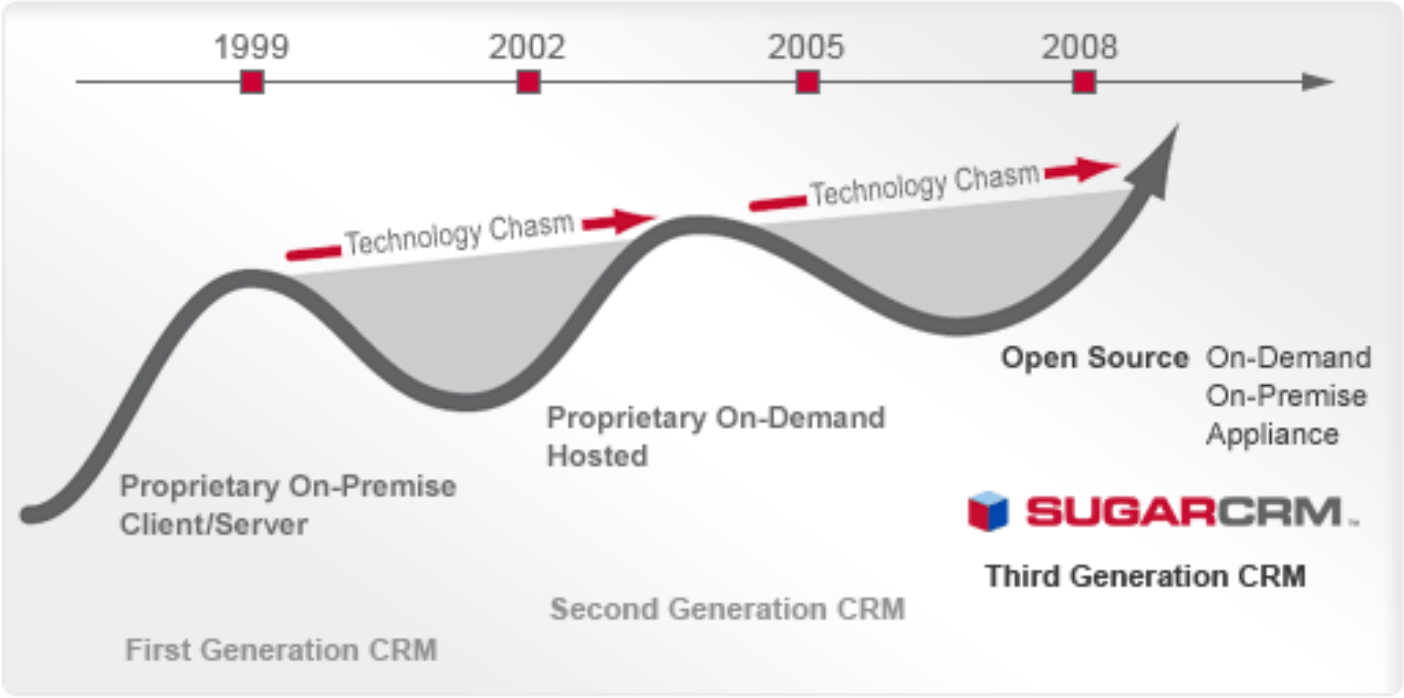
Enterprise applications

Enterprise applications include:

- Enterprise Resource Planning (ERP).
- Customer Relationship management (CRM).
- Supply Chain Management (SCM).
- Human Resources Management (HRM).
- Financial and Planning applications.
- Sales Force Automation (SFA).
- Marketing Automation, e.g., campaign management.
- Systems for manufacturing, logistics, and other functions.

Enterprise applications

SugarCRM's vision –



But...

Alta Plana

Enterprise applications

Open source –

- Doesn't significantly complete with Oracle, SAP, Salesforce.com, or even Microsoft Dynamics.
- Targets small-medium enterprises (SMEs).
- SugarCRM – FOSS and commercial versions.
  - Uses PHP and MySQL.
- ConcourseSuite (was CentricCRM) – OS community edition – Java.
- Compiere ERM and CRM – OS community edition with core capabilities – Java EE.

Questions?

Discussion?

Next: Business Intelligence and Data Warehousing

Agenda

The Business Case for Open Source

Licensing and support

Infrastructure, applications, databases, and tools

BI, data warehousing, and data integration

Evaluation and implementation best practices, strategy, and resources

BI, DW, ETL & Analytics

This segment is about 4 related technologies –

- Data warehousing.
- Business Intelligence.
- ETL/data integration.
- Analytics.

Data Warehousing

What's a data warehouse?

- A reference database structured for analysis.
  - Non-transactional.
  - Contents are cleansed, harmonized, and comprehensive.
  - Partitioning, bitmap indices, star joins, materialized views, & cluster/grid/SMP support help.
- ... with plenty of room for controversy:
  - Kimball versus Inmon/Imhof versus Teradata.
    - Normalized versus "dimensional" models.
  - DW vs. data mart vs. operational data store (ODS).
  - Real-time and "unstructured" data needs.

The Data Warehousing Scene

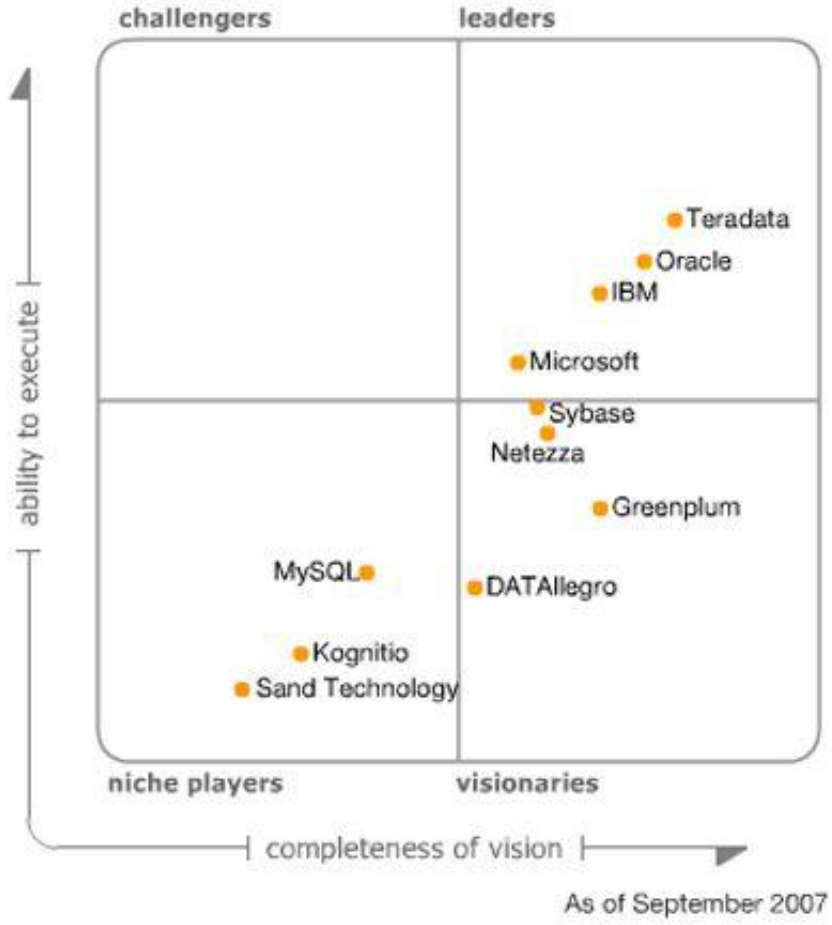
Regarding DW vendors,

- Teradata was the first *notable* DW pure-play...
  - blazing a trail for the DW appliance vendors, e.g., DATAllegro (Ingres), Netezza, and Sun-Greenplum.
- Every major DBMS vendor supports DW; specialized vendors include Kalido, Kognitio.
- Analytical tools will generally work with any DBMS that supports *standard* APIs/access methods.

What does this mean?

- DW techniques are portable to any DBMS platform with the necessary capabilities and tool support.

The Data Warehousing Scene



Gartner excludes PostgreSQL: "Support for these data warehouse DBMS products must be available from the vendor — community-supported open-source software (OSS) products are not included."

http://mediaproducts.gartner.com/reprints/microsoft/article19/article19.html

Database Systems for DW

There are three leading OS-DBMS players in the BI & DW world:

- Ingres.
- MySQL.
- PostgreSQL.

Ingres is possibly the most enterprise worthy.

- "We switched from Postgres to Ingres a few years ago after determining that Ingres was much more suited for data warehousing." -- Stuart Frost, CEO of DATAllegro.

Database Systems for DW

MySQL, popular but limited DW capabilities.

- Multiengine architecture. We're interested in –
  - MyISAM.
  - Merge.
  - InfoBright.
- Big strides with MySQL 5, out in late 2005.
  - Native functions, user defined functions, stored procedures, views.
- 5.1, due out any time, adds true partitioning.
- Nice query, admin & migration utilities. Toad for MySQL is free.

Gartner on MySQL

Strengths

MySQL has continued to mature — new functionality, growth of professional services, a growing sales force, an alliance with IBM, and the addition of many new third-party software vendors. With its new MySQL Enterprise offering (an installable system from a set of discs like most other DBMSs), it has seen rapid market acceptance. Many clients are beginning to use MySQL as a data warehouse engine for small data warehouses, up to about 200GB to 500GB in size. However, many data warehouse implementations begin small and grow over time. MySQL will see the same growth as its scalability is proven over time.

MySQL has several references with mutiterabyte data warehouses in production using a technique MySQL calls "sharding." This technique splits the database into smaller pieces of less than a terabyte. Although this requires more resources to manage the database and associated storage, it does represent another step in the direction of large data warehouse capabilities.

MySQL still maintains a low price point — a free license with support subscriptions ranging from $599 per year per server to $40,000 per year (for the unlimited server license of MySQL Enterprise).

The recent announcement by IBM and MySQL to port the MySQL DBMS to the System i opens MySQL to many new clients and can be expected to be used here for OLTP and data warehousing on the System i (see "MySQL Will Open IBM System i to New Applications and Customers"). Similarly, BrightHouse, Infobright's column-oriented engine, uses MySQL to create an analytic data warehouse solution. These possibilities are due to the architecture of the MySQL DBMS, allowing MySQL to work with multiple storage engines.

As number of downloads is not relevant to market growth (you cannot distinguish between experimental and educational downloads versus downloads for production), the increasing number of clients purchasing support services and MySQL Enterprise has led to revenue doubling year over year.

Gartner on MySQL

Cautions

MySQL continues to lack references for data warehousing that break the 1TB barrier in a single instance of the DBMS (see "sharding," mentioned earlier). To become a strong player in the overall DBMS market, and specifically the data warehouse DBMS market, it will need to spend 2008 concentrating on developing these accounts as referencable data warehousing customers with a range from 1TB to 5TB. In addition, it will need to begin to demonstrate scaling above the 10TB range in a mixed workload to dispel the perception of a lack of scalability of MySQL.

The company is facing increased competition from some of the new entrants using OSS DBMS technology, such as EnterpriseDB (just beginning to support Data Warehousing with EnterpriseDB GridSQL), ParAccel and Vertica — all of which are using PostgeSQL as a base.

Currently, MySQL still lacks many of the special features necessary to be a serious contender for large data warehouses. For example, current production version 5.0 does not have partitioning, which is due for version 5.1. Although MySQL has some basic functionality for workload management (such as storing query statistics), it will need to add more control and automatic management functionality to handle large data warehouses and the mixed workload.

The low entry cost of using MySQL does not always equate to low total cost of ownership (TCO), as the cost to manage a large data warehouse without the broad availability of management tools (as with the larger, more mature data warehouse DBMSs) leads to the use of resources to perform these management tasks manually.

Database Systems for DW

- PostgreSQL is a robust enterprise platform.
  - Greenplum database, which is parallelized, is designed for data warehousing.
  - Truviso adds streaming data capabilities.
  - Do NOT use Bizgres, the open source version of a spring 2006 Greenplum release. It's dead.
- EnterpriseDB's Postgres Plus is another PostgreSQL packaging.
  - Open source GridSQL adds optimization & parallelization.

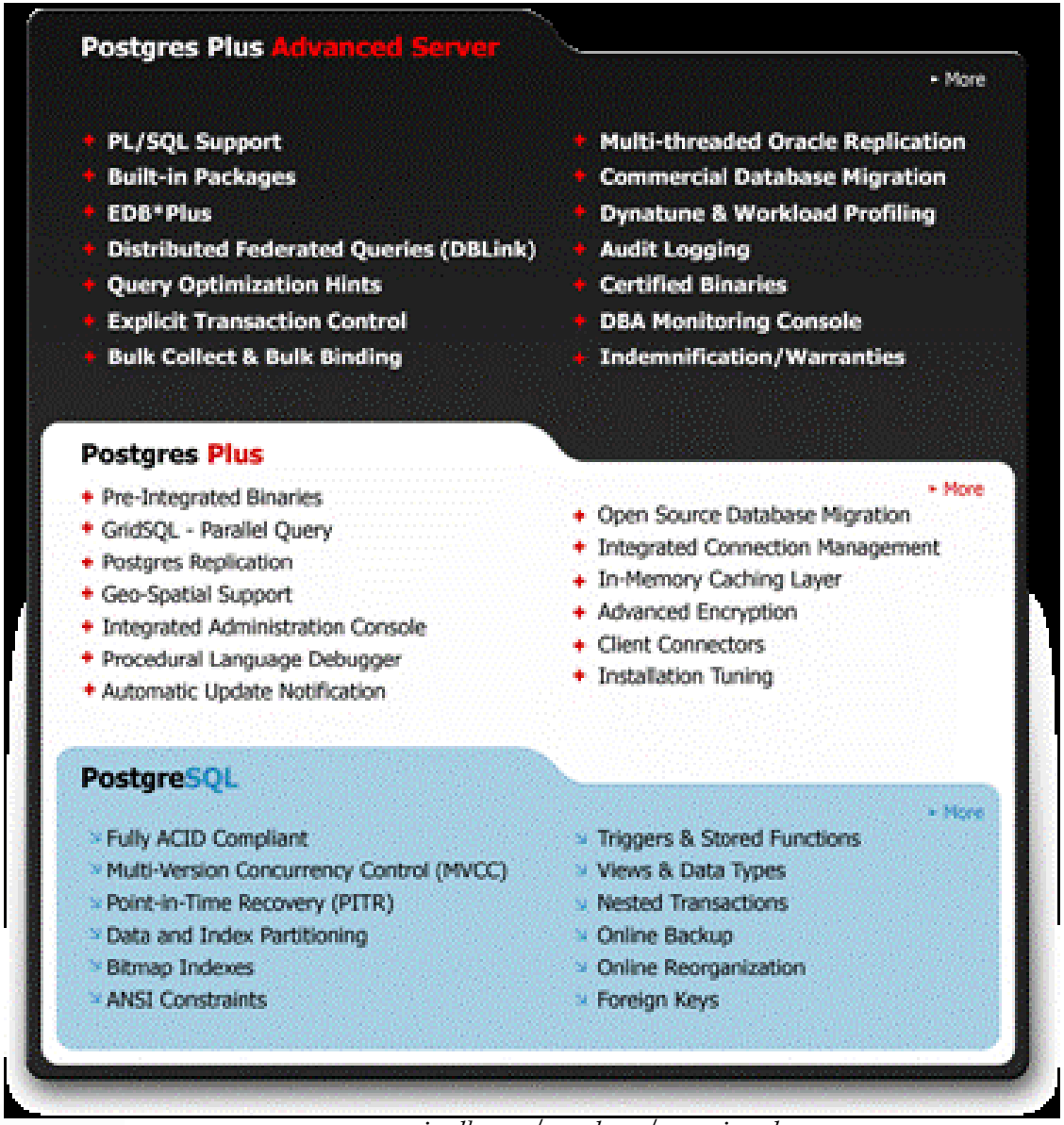
Database Systems for DW

Gartner on Greenplum –

"Demonstrated scalability in production to hundreds of terabytes and internally to over a petabyte (1,000TB). It has also demonstrated the ability to run and manage the mixed workload."

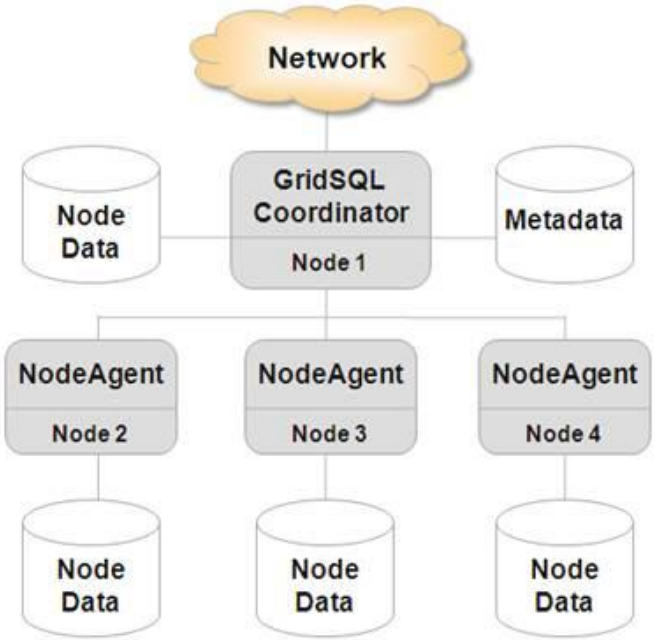
"Use of an OSS DBMS as the core work engine also helps to reduce costs while it concentrates on the management software surrounding the data warehouse and the optimization features necessary for a complex, mixed workload environment."

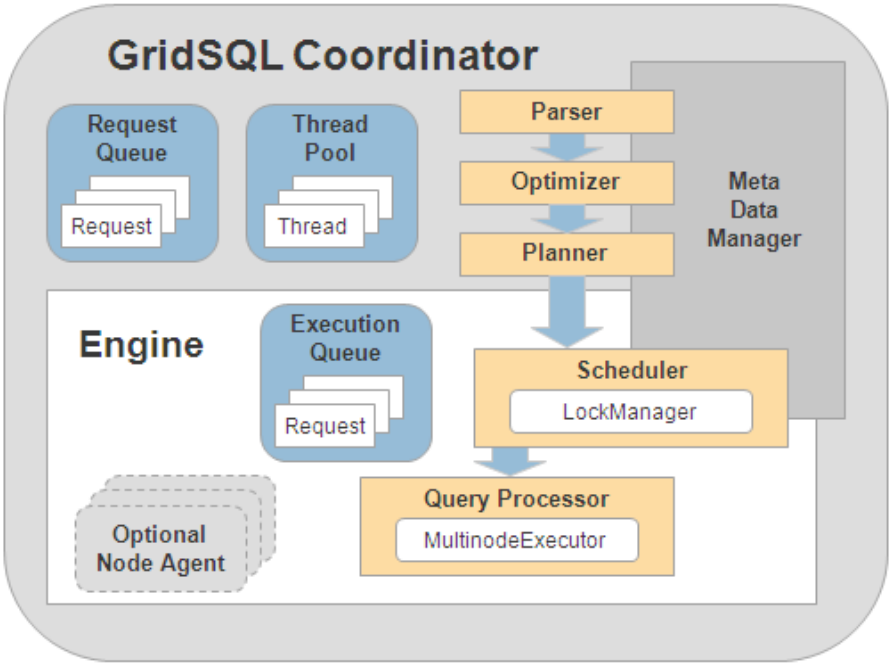
http://mediaproducts.gartner.com/reprints/microsoft/article19/article19.html



www.enterprisedb.com/products/overview.do

GridSQL adds parallelization for PostgreSQL.





www.enterprisedb.com/gridsql/architecture.do

Database Systems for DW

Yahoo! created its own PostgreSQL extension –

- Designed a column store for PostgreSQL with compression and parallization.
- 2 petabytes, expected to grow to 10s of pb by next year.
- Designed for analysis of the behavior of a half-billion monthly visitors.
- Processes 24 billion events a day.

http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9087918

DW Recommendations

Data warehousing is undergoing big changes.

- MPP, shared-nothing architectures.
- Column stores.
- Adaptive, search-reliant data warehouses.
- Data-stream management.
- Appliances.
- Embedded analytics.

Nonetheless...

DW Recommendations

There is an open source/OS based option that will meet most DW needs.

- For appliance needs, consider Ingres Icebreaker or Greenplum-Sun, but evaluate against non-OS based options including DATAllegro and Netezza.
- MySQL shops should try MySQL first, but consider it a prototype and evaluate performance and scalability. Consider an enhancement such as Infobright and/or Kickfire.
- Developer shops should try PostgreSQL. Do consider EnterpriseDB's Postgres Plus.

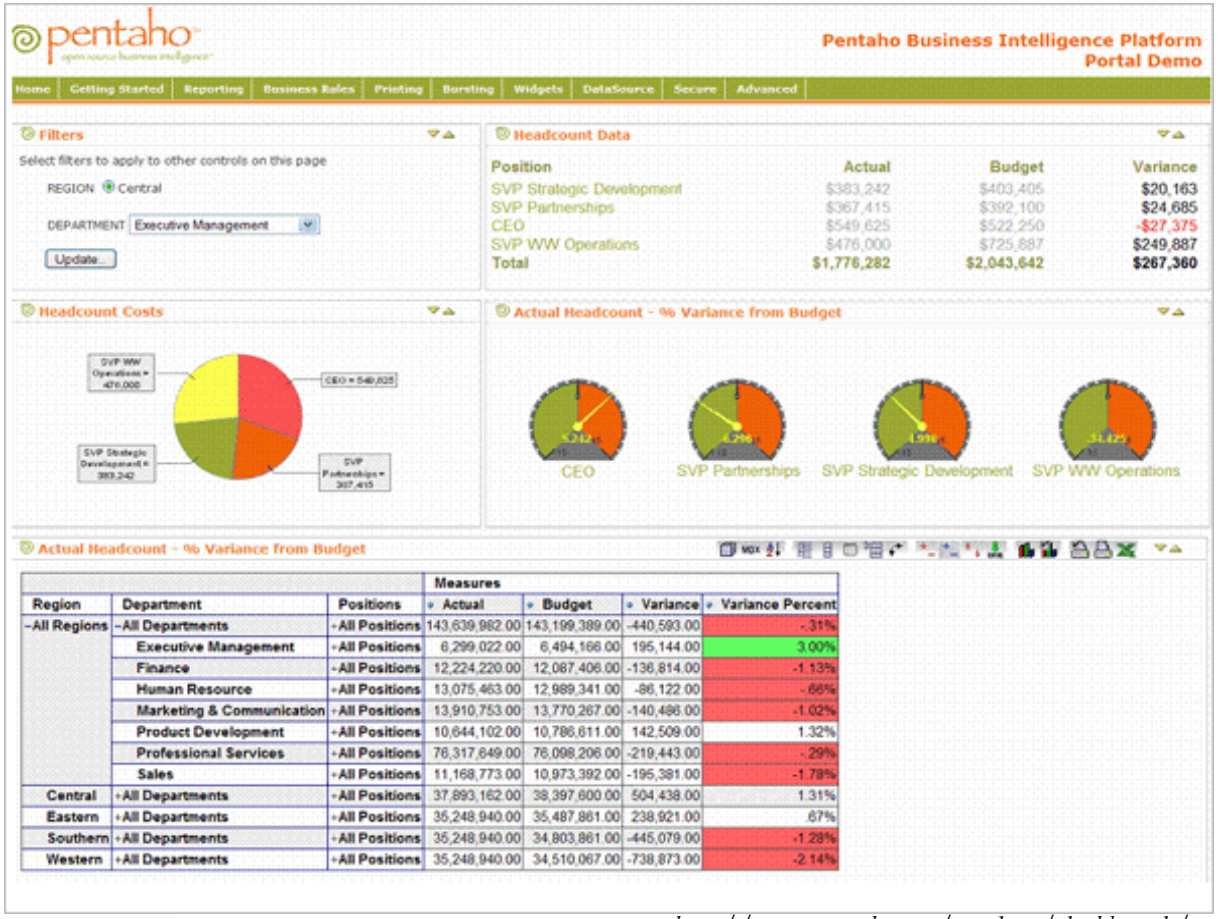
DW Recommendations

Continued...

- Enterprises looking for a supported DBMS that will run out-of-the-box should try Ingres and (EnterpriseDB) Postgres Plus Advanced Server.
- Enterprises doing high-volume data warehousing should consider Greenplum and Ingres.
- Palo is a datamart alternative.
- Keep an eye on LucidDB.
- Do consider MonetDB.
- Do not use Bizgres. It is dead. For moderately sized DWs, try Postgres Plus with GridSQL.

Business Intelligence

Typical BI looks like this:



http://www.pentaho.com/products/dashboards/

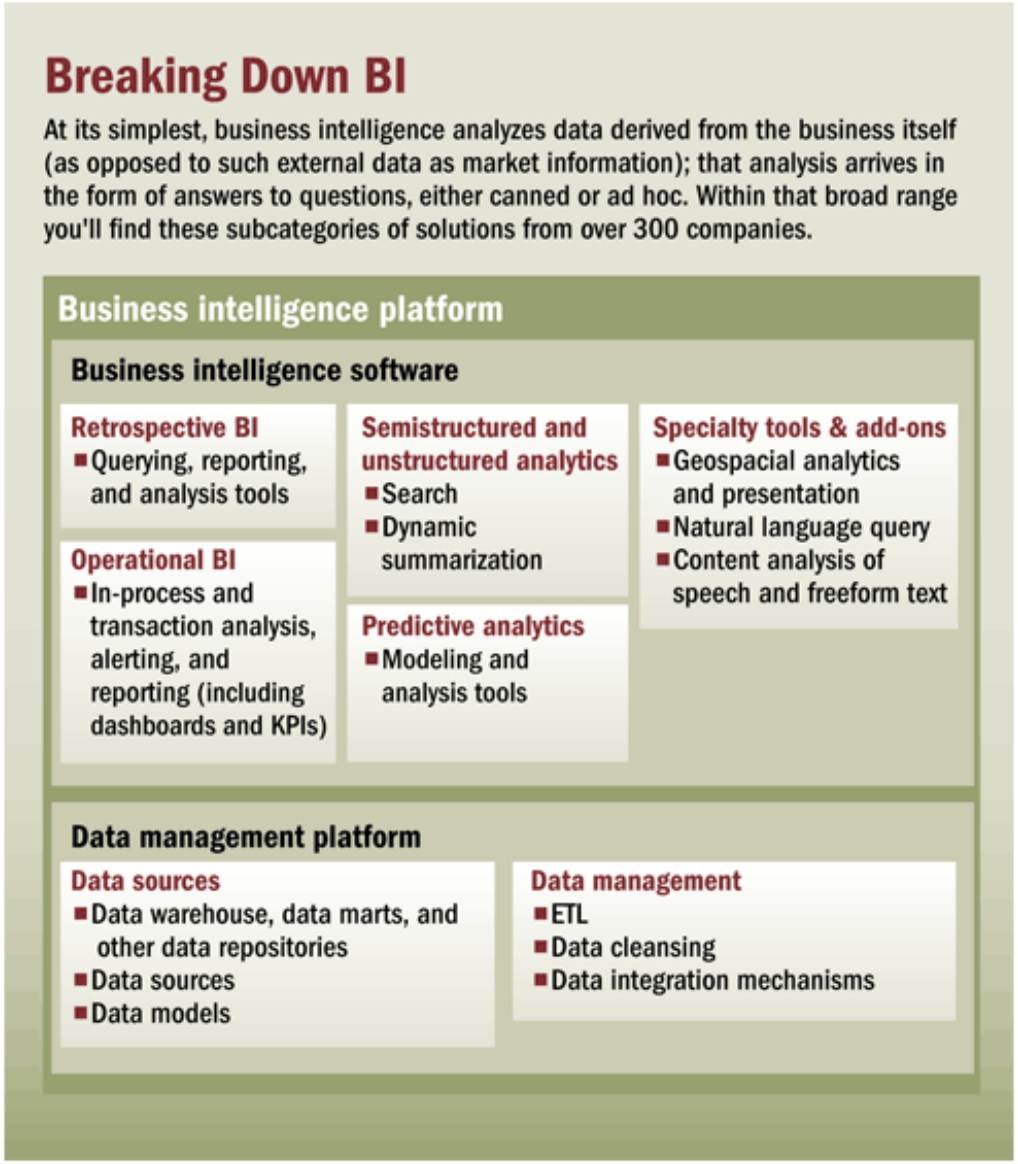
Business Intelligence

BI encompasses:

- Process: event → data → analysis → decision.
- Software.
- Information: a highly contextual business driver.

We're addressing software here, but we want to keep in mind:

- Integration with operational systems.
- Embedding analytics in line-of-business applications.
- Collaboration.



www.infoworld.com/article/07/04/02/14FEbizintel_1.html?source=NLC-SR&cgd=2007-04-03

Business Intelligence

BI software consists of:

- Reporting; dashboards; ad-hoc query.
- Analysis, especially OLAP.
- Advanced analytics, e.g., statistics and data mining.
- Office/applications integration including EAI.

BI relies on:

- Information movement & integration, e.g., ETL.
- Data warehousing; metadata management.
- Visualization.
- Search.

The Business Intelligence Scene

There are many BI vendors.

- Pure-/almost-pure-plays: Business Objects, Cognos, Hyperion, Information Builders, Microstrategy.
- The (would-be) dominators: IBM, Microsoft, Oracle
  - ... and their toadies such as Panorama.
- Visualization, performance management, reporting, dashboard specialists: Actuate, arcplan, Spotfire, Tableau, TIBCO Spotfire.
- Analytics heavyweights: SAS, SPSS.

Recent, major consolidation.

- Hyperion→Oracle, Cognos→IBM, BOBJ→SAP.

The Business Intelligence Scene

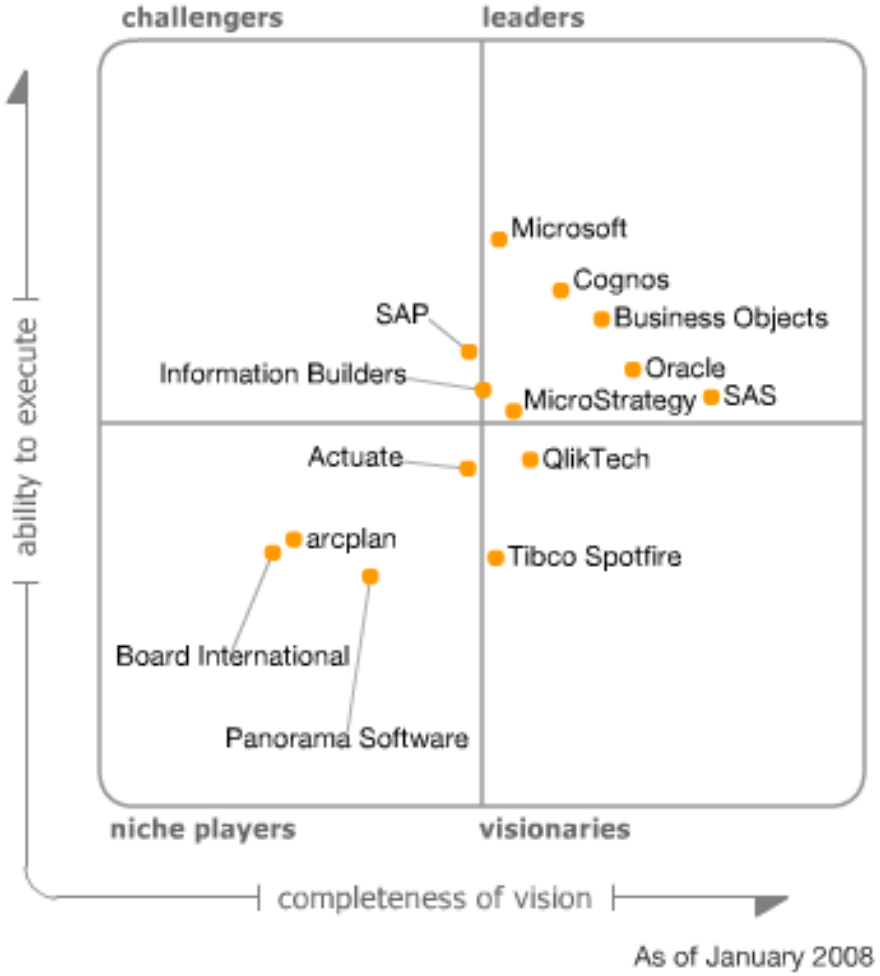
BI has an Excel problem, an artifact of the PC devolution, but it's still not pervasive.

What does this crowded-segmented field mean?

- Vendor lock-in.
- When it comes to "out-of-the-box" end-user BI, **open source is nowhere to be seen**.

But let's look at mainstream perceptions...

The BI World According to Gartner



http://mediaproducts.gartner.com/reprints/microsoft/vol7/article3/article3.html

What Do the Analysts Think?

Nigel Pendse is author of the OLAP Report –

Actually, I've been quite surprised at how little impact open source BI solutions seem to be having. I was expecting much more.

I guess there are two parallel universes: customers in OSW (open-source world) have decided for idealistic, economic or technical reasons that they must have an open-source solution, and don't even consider any proprietary options, while most other people ignore open-source solutions....

Current OS OLAP solutions are quite weak (at least a decade behind the current proprietary products), whereas the reporting solutions may be better...

The proprietary BI software vendors seem to be genuinely unconcerned by open-source BI. They never mention it to me, and they seem quite surprised if I ask them about it. A few have looked at briefly products like Pentaho, and seem totally unimpressed/unconcerned. I guess they don't sell into OSW anyway, and therefore aren't losing any business to OS BI that they are aware of.

Category Error

My guru friends have made a "category error."

- Open source does not succeed (best) by replicating commercial, proprietary, closed source software and processes.
- The most successful open source projects are not imitative, they are innovative.
- Think about Internet, server, and desktop computing in this light.

OSBI has NOT aimed to replace closed-source, commercial solutions... yet.

BI Components

For reporting –

- JasperReports.
- Eclipse Business Intelligence and Reporting Toolkit (BIRT) from Actuate.
- JFreeReports.

Online Analytical Processing (OLAP) –

- Mondrian Relational OLAP Server.
- Palo Multidimensional OLAP Server.

Components

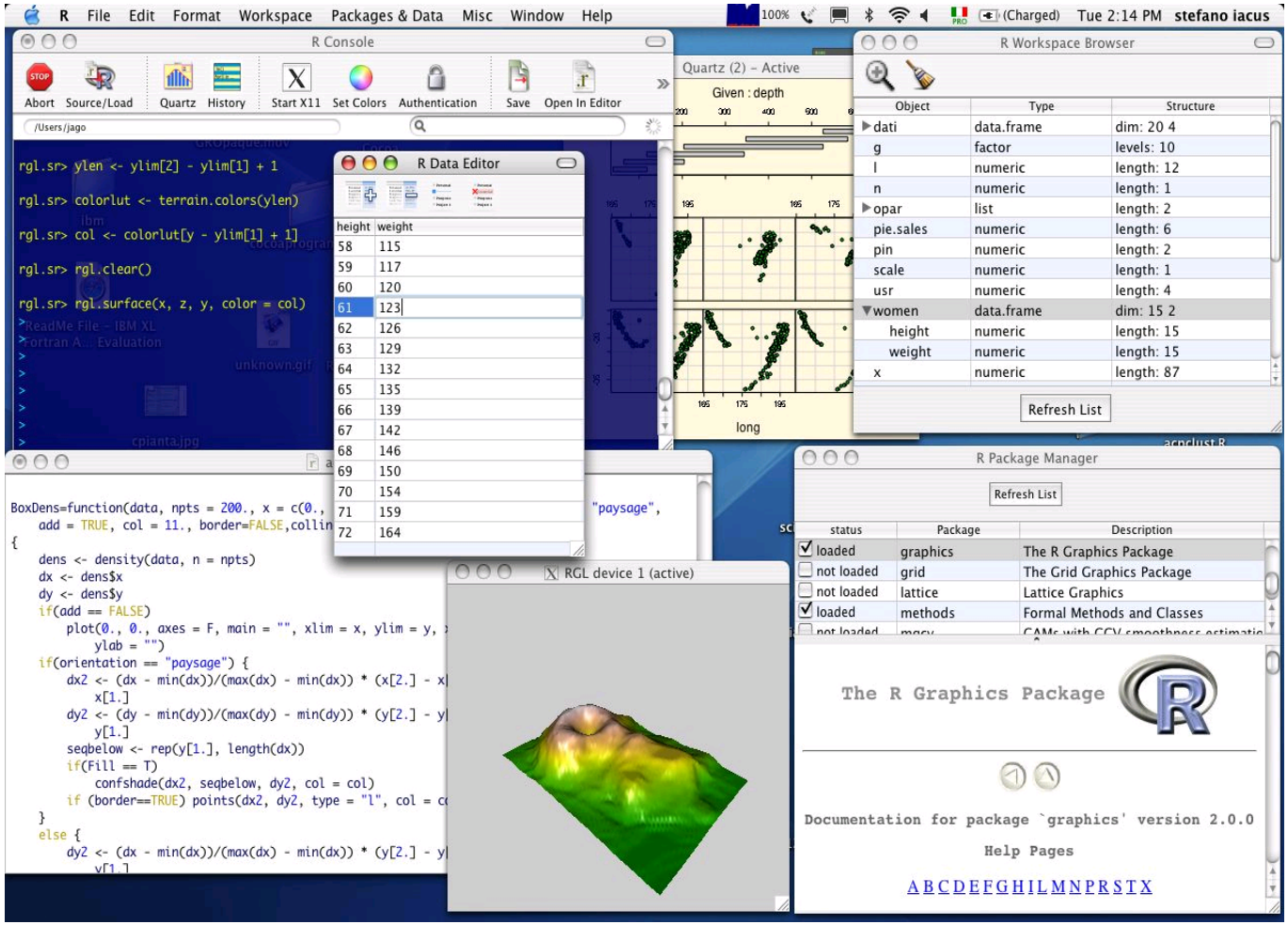
- Application Programming Interfaces (APIs) –
  - JPivot JSP (Java Server Page) tag library.
  - OLAP4J Java API (*http://sourceforge.net/projects/olap4j*).
- Data integration/ETL –
  - Enhydra Octopus, Kettle ETL, Kinetic Networks Extract Transform and Load (KETL), Talend.
    - See *www.manageability.org/blog/stuff/open-source-etl/view* .
- Visualization and text –
  - AT&T Graphviz.
  - Gate, Lucene, UIMA for search/text analytics.

BI Components

For data mining –

- R is an open source implementation of AT&T's S statistical programming language.
  - R-Python links let you extend Postgres!
- Weka is a machine learning and data mining tool.

Sample: R



ETL & Data Integration

Apatar (*http://www.apatar.com/*) – dual GPL/commercial license.

- Mashup Data Integration
- Operational Integration
- Application Integration

Pentaho (*http://kettle.pentaho.org/*) – LGPL v2 license.

Talend (*http://talend.com/*) – GPL v2, support & SaaS licenses.

XAware (*http://xaware.com/*) – GPL v2/commercial.

- Real-time data access; SOA implementation.

Packages

Include components, applications & frameworks.

Pentaho –

- JPivot, Mondrian, JFree reports, Kettle ETL, Weka data mining, Excel services with portal tools and workflow management.

JasperSoft BI Suite –

- JasperReports, JasperStudio (designer), JasperServer, Talend ETL, and Mondrian OLAP.

OpenI and SpagoBI provide other frameworks for Mondrian and Jpivot and more.

JasperSoft

JasperSoft claims –

- 80-100 thousand deployments.
- 8,000 customers including 100s of subscriptions.
  - Others are docs, services, incident-based support.

Embedded/OEM licensing is half of JasperSoft's business.

Ingres Icebreaker BI appliance includes Talend ETL, using Linux on commodity hardware.

http://www.ingres.com/products/icebreaker-bi-appliance.php

Other partnerships include ParAccel.

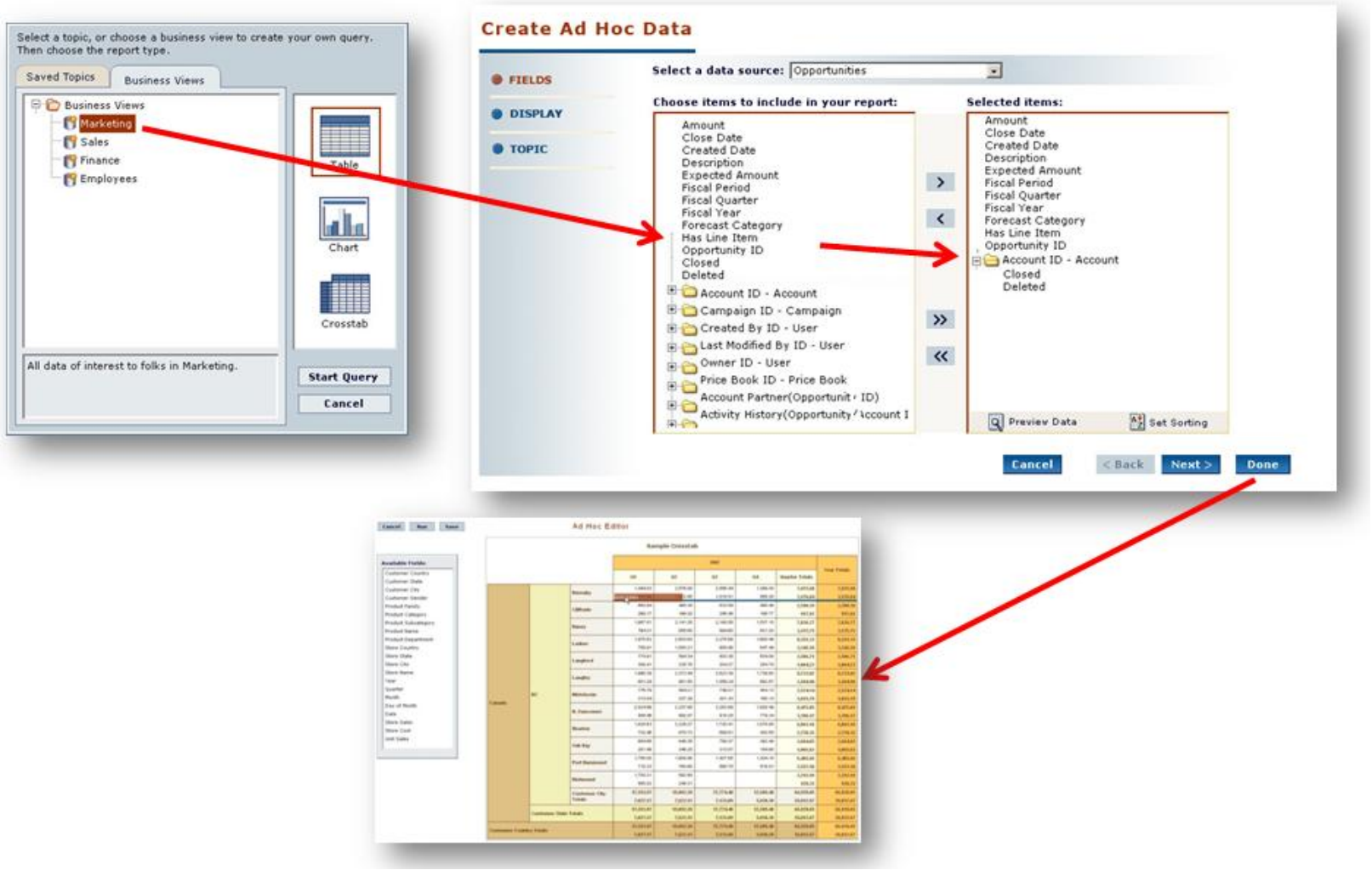
JasperSoft

With V3, JasperSoft is targeting enterprise deployments.

- Professional (non-OS) versions will feature –
  - A business abstraction layer called a domain.
  - AJAX dashboard elements.
- OS version will feature an enhanced repository manager.

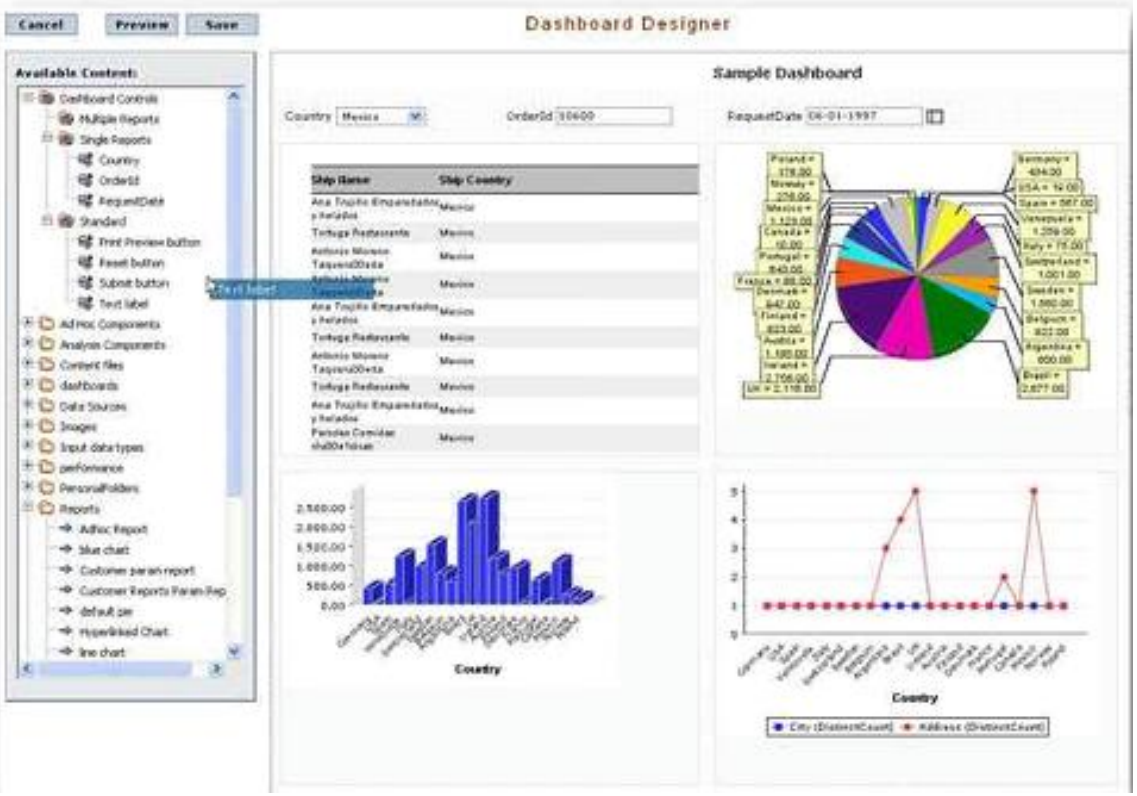
ODBO Connect for Excel is not open source.

JasperSoft v3

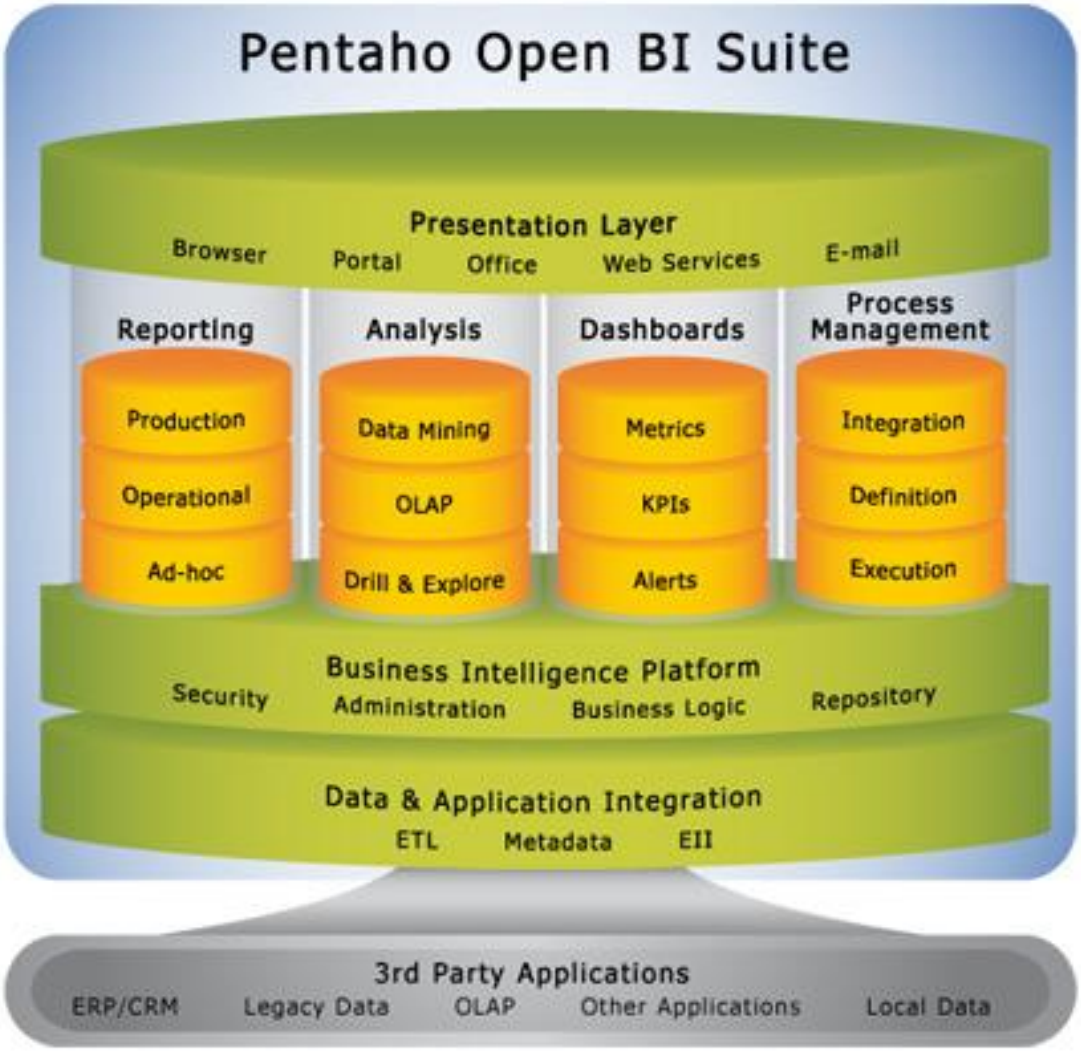


JasperSoft v3

- **Drag & drop reports onto Dashboard**
- **Dynamic resize & layout**
- **Global controls**
  - Input parameters
  - Other controls
- **"Twinkling" dashboards**
  - Dynamic frame refresh in seconds



Pentaho

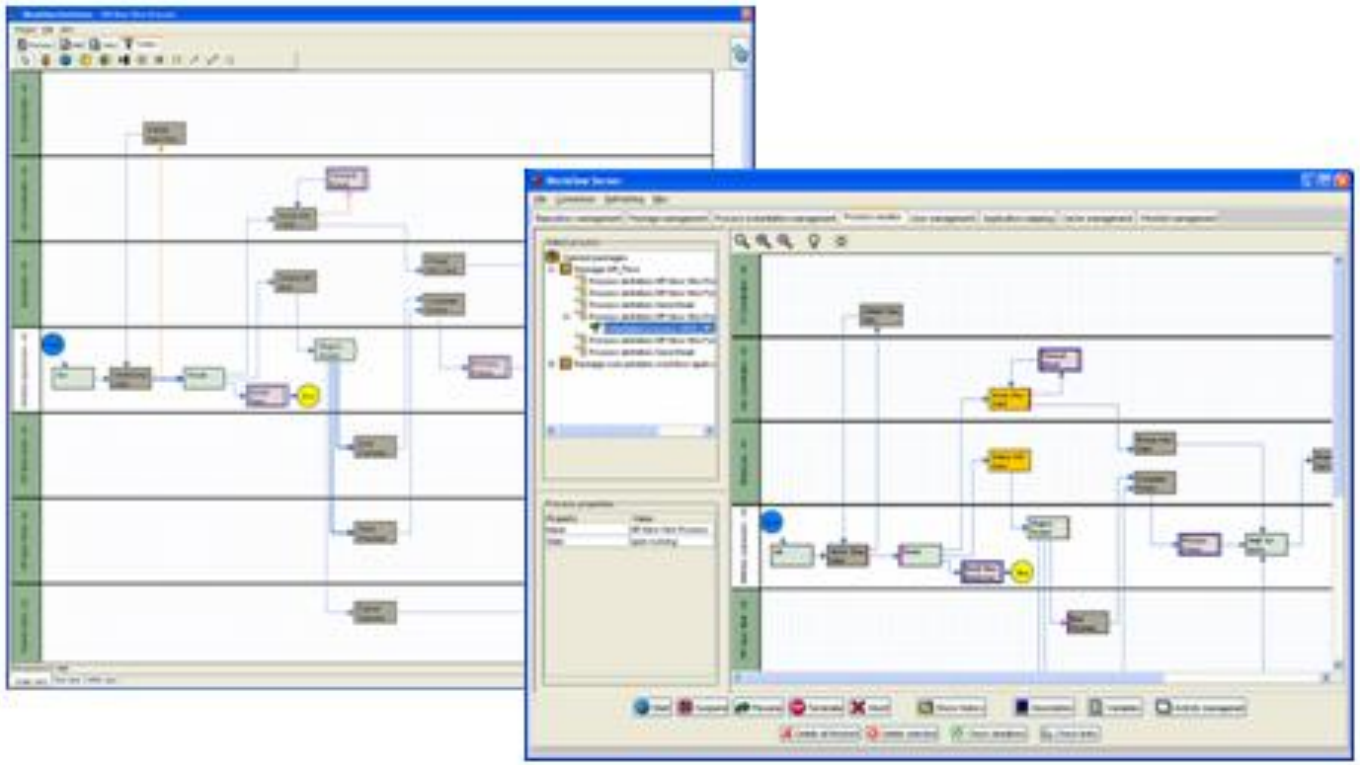


Pentaho

- Pentaho Analysis Services = Mondrian ROLAP.
  - Pentaho Spreadsheet Services comes from Simba Technologies. It is not open source. It uses XML/A to provide a Mondrian interface in Excel.
- Pentaho Dashboards.
- Pentaho Data Integration = Kettle ETL.
- Pentaho Reporting
  - Report Creation Tools.
  - Report Distribution.
  - Embedded Reporting = JFreeReports.

Pentaho

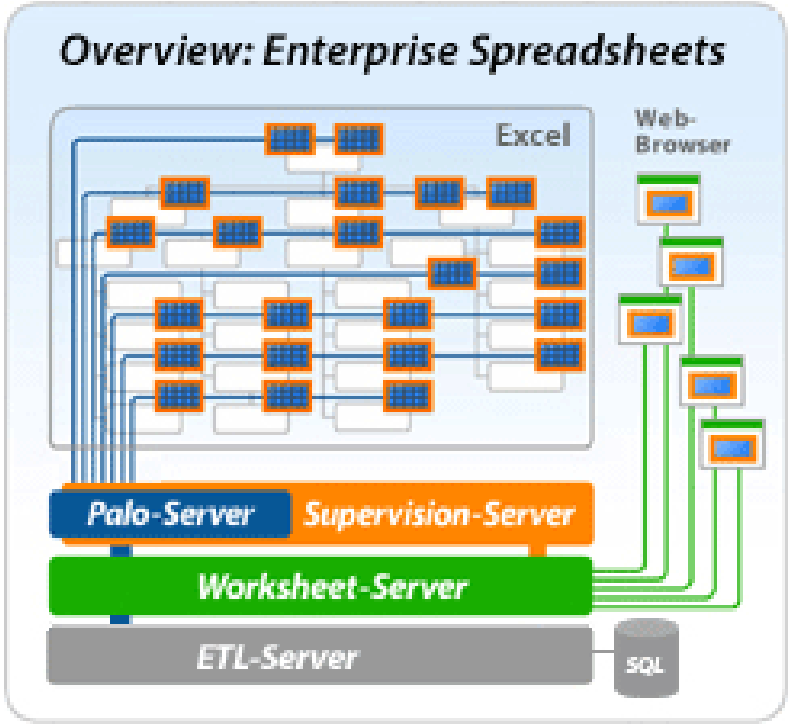
Pentaho Workflow Services is a distinguishing feature.



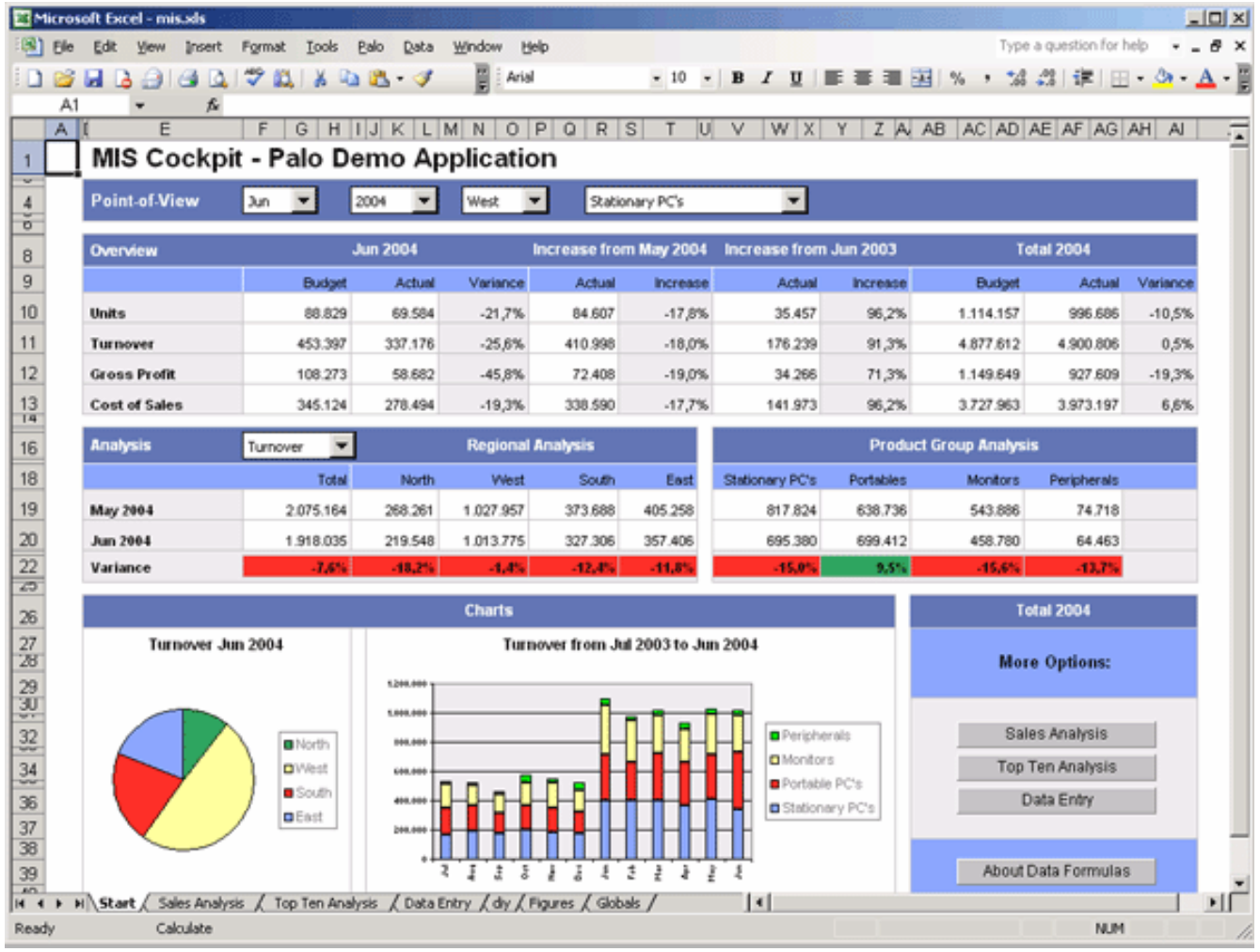
Palo

MOLAP server from Jedox, designed for spreadsheet services. (*http://www.jedox.com/*)

- Excel add-in.
- .NET, Java, PHP, C APIs.
- SAP, DBMS connectivity.
- GPL/commercial license.

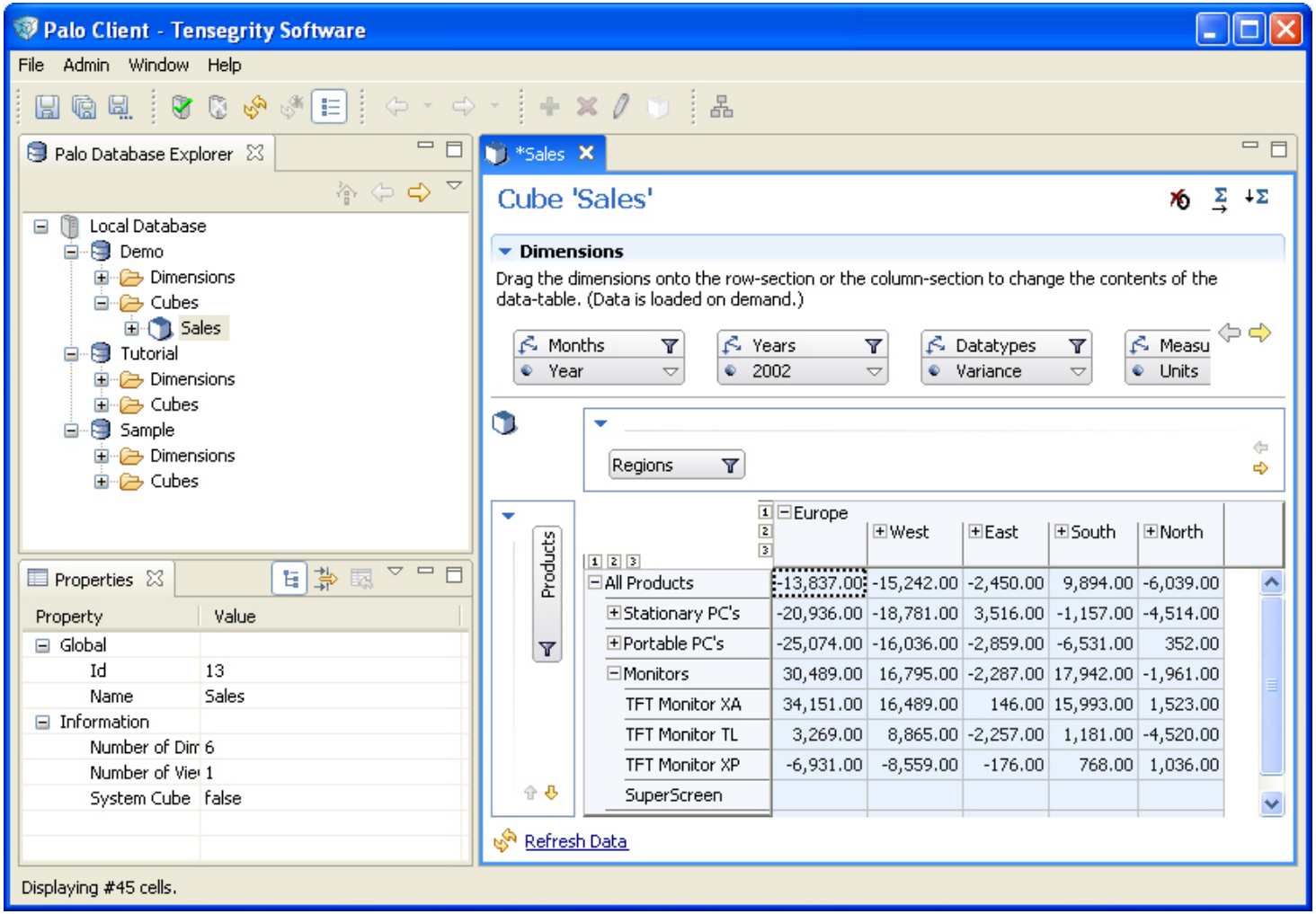


Palo – Excel



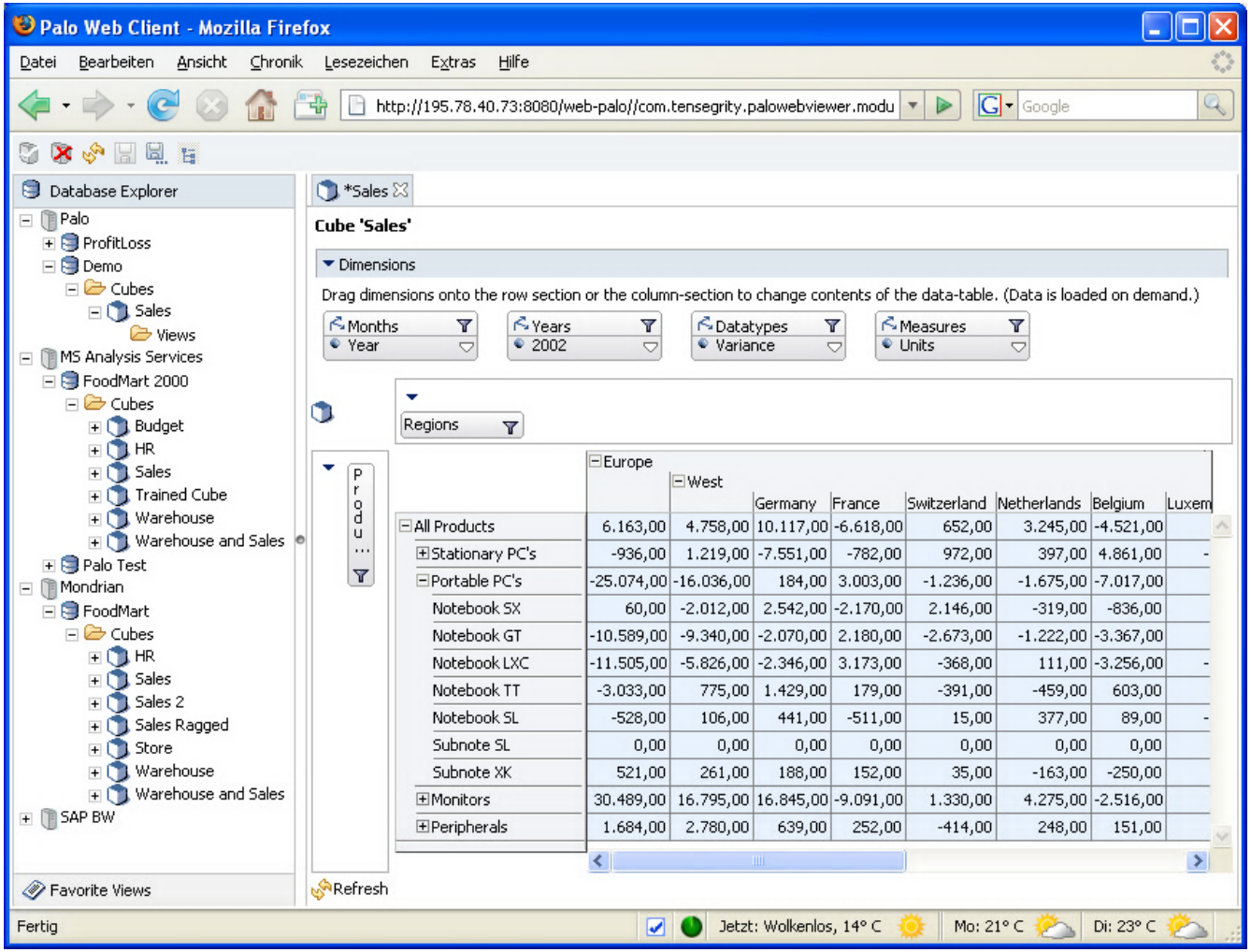
http://www.jedox.com/en/enterprise-spreadsheet-server/excel-olap-server/palo-server.html

Palo Client – Tensegrity – Eclipse



http://www.jpalo.com/en/products/palo_client.html

Palo Web Client – Tensegrity



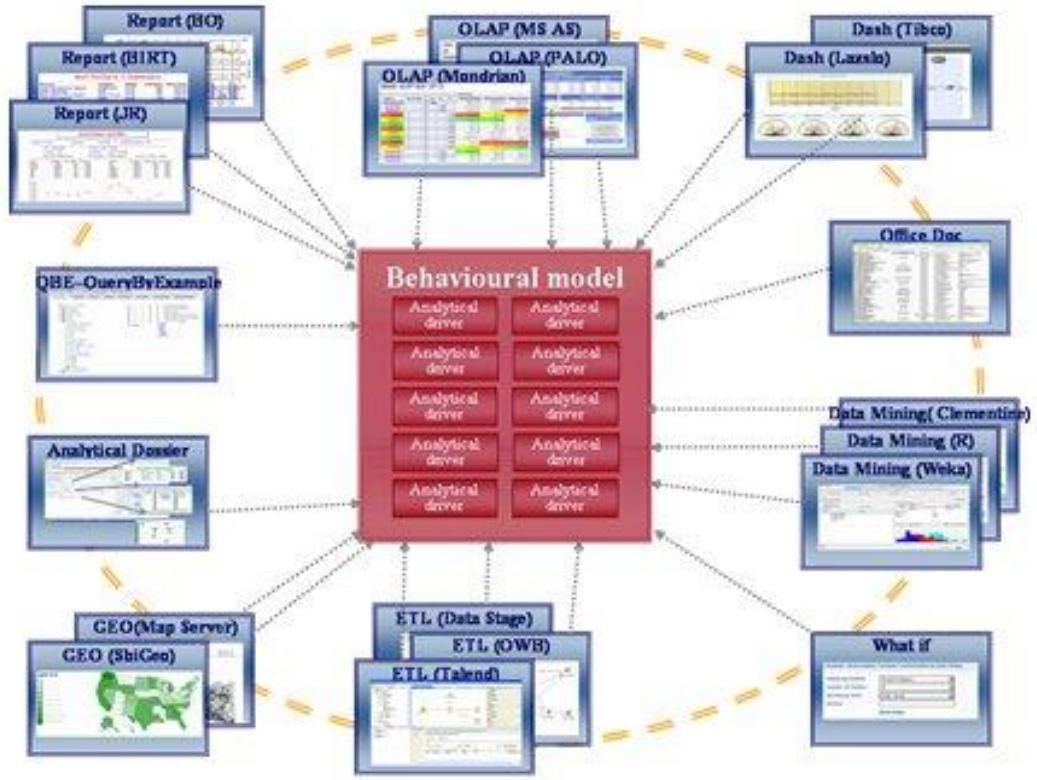
http://www.jpalo.com/en/products/palo_web_client.html

Eclipse BIRT

- Eclipse is open source.
  - The leading Interface Development Environment (IDE) and Rich Client Platform (RCP).
- Business Intelligence and Reporting Tools (BIRT) is a top-level Eclipse project.
  - Java/J2EE.
  - Actuate is the project lead.
  - Report designer.
  - Web-server deployable run-time component.
- Eclipse BIRT is of interest to Java developers.

SpagoBI

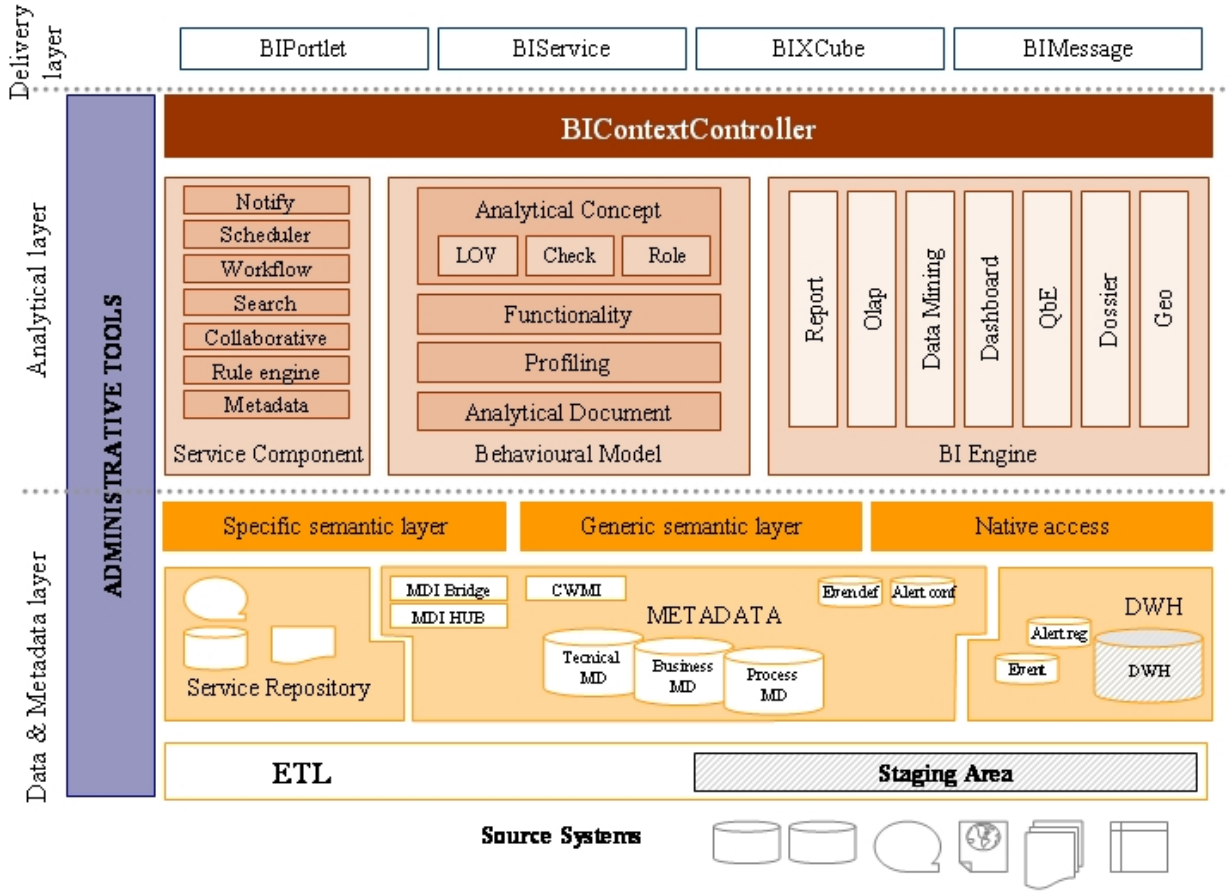
LGPL license; completely open source.



http://spagoworld.com/ecm/faces/public/guest/home/solutions/spagobi

SpagoBI

The architecture supports interoperability.



Choosing an OSBI Product

So far as I'm concerned, Pentaho and JasperSoft are functionally equivalent although Pentaho's suite is broader.

- Both are developer, end-user, and enterprise friendly.
- Both have strong support networks.

Palo offers similar capabilities with less focus on reporting and less DB flexibility.

For straight reporting, try JasperReports.

For reporting developers, Eclipse BIRT, Pentaho, or JasperReports.

Choosing an OSBI Product

For spreadsheet services, Palo. (Pentaho & JasperSoft require a professional license.)

For data mining, R or Weka (via Pentaho or not).

For ETL, Talend or Kettle (via Pentaho or not); XAware is also credible.

The SpagoBI framework provides for tool interoperability, e.g., JasperReports on a Palo server.

Market Reaction

How have vendors of proprietary, closed source, commercial software reacted to OSBI?

- By porting to Linux, providing limited MySQL support, and embracing Eclipse.
  - I interpret these steps as mostly positioning for now.
- By moving up the applications stack into –
  - Business Performance Management.
  - Planning & Budgeting, Compliance.
  - Industry verticals.
- By attempting platform lock-in –
  - Microsoft Sharepoint.

Market Reaction

But the established vendors shifted tactics before OSBI emerged. What pushed them?

- Competition.
- Commoditization: Microsoft SQL Server OLAP, Analysis Services.
- Opportunity (i.e., $$) generated by the enterprise-applications space: SAP, Siebel, Oracle.

Market Analysis

Is OS BI-DW a threat to established vendors?

- Not while OS projects/vendors are providing tools but few solutions.
- Not until it establishes an end-user presence.
- Not until there are more, credible user stories showing robustness, scalability, reliability.
- Not while alliances break out of the open-source/small-shop world.

Market Analysis

Expect a three-wave approach to open-source BI adoption:

Wave 1: 2004 to 2007 – early adopters.

Wave 2: 2008 to 2012 – driven by midmarket enterprises.

Wave 3: 2012 and beyond – just another aspect of sourcing.

– Who's Who in Open-Source Business Intelligence," Andreas Bitterer, Gartner, 18 April 2008 (*http://www.stratebi.es/todobi/may08/whos_who_in_opensource_busin_156326.pdf*).

Market Analysis

The answer to the "category error"?

- OS BI-DW is doing quite nicely providing developer tools for end-user and embedded applications.
- Their route to enterprise acceptance is:
  - by leveraging the OS stack.
  - by appealing to in-house developers.
  - by supporting development shops.

Will OSBI provide the tools (and cost model) to enable the much-talked-about democratization of BI?

Questions?

Discussion?

Next: Best Practices, implementation strategy, and resources.

Agenda

The Business Case for Open Source

Licensing and support

Infrastructure, applications, databases, and tools

BI, data warehousing, and data integration

Evaluation and implementation best practices, strategy, and resources

Best Practices

- There are many thousand OS projects.
  - How do you find the tools you need?
    - The BI/DW world isn't so large; we'll cover this anyway.
  - How do you assess tool:
    - Capabilities?
    - Total Cost of Ownership (TCO)?
    - Viability?
  - How do you evaluate candidate tools?
- How do you implement?
- How can you participate and contribute back?

Finding #1

Starting point #1: development and distribution sites.

- *Sourceforge.net.*
- *ObjectWeb.org.*
- *Eclipse.org.*
- Microsoft's CodePlex, *codeplex.com/*.
- GNU, *www.gnu.org/software/*.

There are also tools/vendor specific "forges":

- *JasperForge.org*, JasperIntelligence BI.
- *SugarForge.org*, SugarCRM.

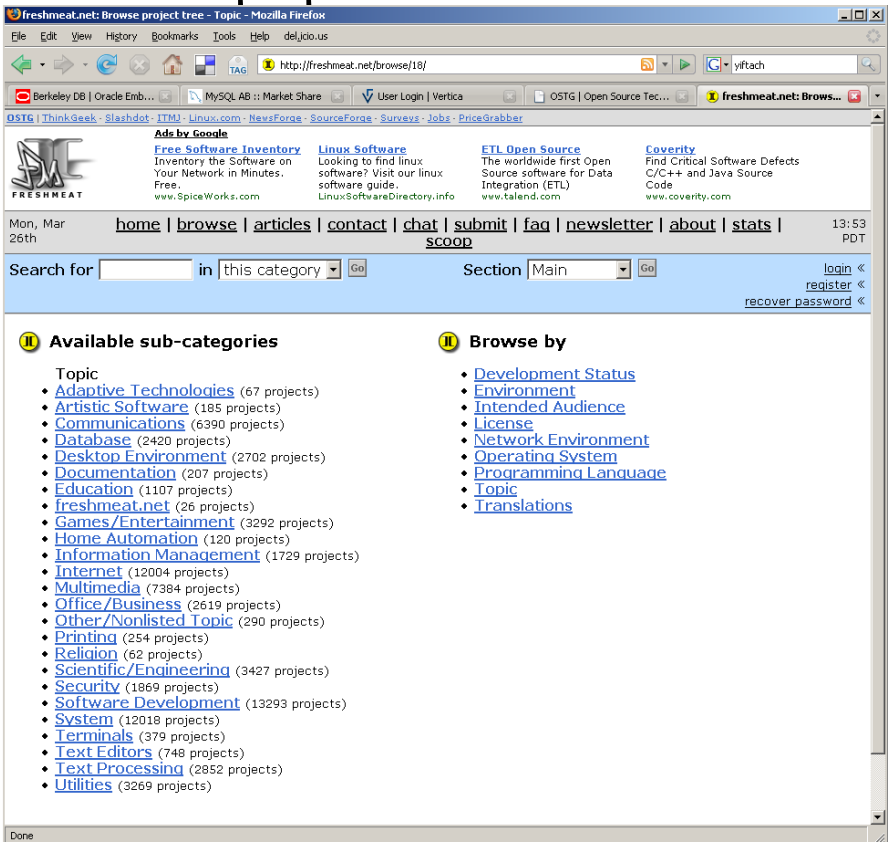
Finding #2

Starting point #2: news and information sites.

Newsforge.com, "the online newspaper for Linux and open source."

Freshmeat.net

try the browse page –



Freshmeat

"freshmeat maintains the Web's largest index of Unix and cross-platform software, themes and related 'eye-candy,' and Palm OS software. Thousands of applications, which are preferably released under an open source license, are meticulously cataloged in the freshmeat database, and links to new applications are added daily. Each entry provides a description of the software, links to download it and to obtain more information, and a history of the project's releases, so readers can keep up-to-date on the latest developments."

freshmeat.net/

Finding #3

Starting point #3: key open-source platform / stack / solution providers.

- Apache.
- GNU.
- Red Hat/JBoss.
- Novell.
- MySQL.

A good place to start if you already run software that is a part of a stack.

Finding #4

Starting point #4: Attend project / vendor / open-source / business conventions.

… or just review their Web sites.

Speakers and panels.

Exhibitors.

Finding #5

Also, look into industry alliances –

- Open Solutions Alliance.
- Interop Vendor Alliance.
  - Established by Microsoft; open source and not.
  - Includes Red Hat, Sun, SugarCRM, Novell.

interopvendoralliance.org/Directory.aspx

- This possibility may be most useful if you already use one of the alliance solutions.

And use the course Resources, which follow.

Assessing

Once you've identified candidate software:

- Look for published surveys and assessments.
- Check out ohloh – more in a moment – and look for blog entries and articles at sites such as SlashDot.
- Look for implementation case studies.
  - Prominent users and users whose IT environments and/or business needs are like yours.

There's a presumption that market presence ("mindshare") + and funding = worthiness.

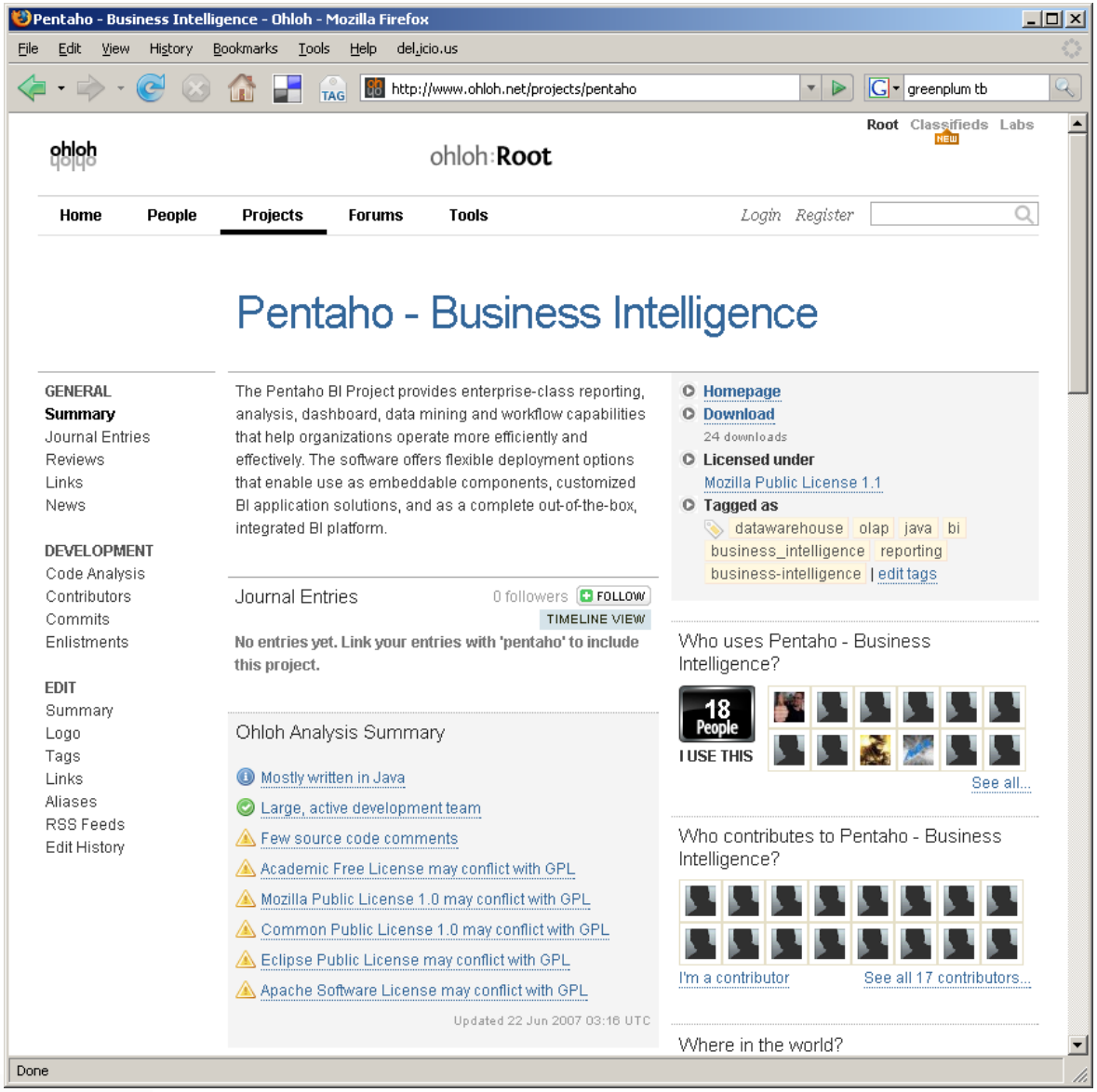
- Is this true?

Assessing: Ohloh

Ohloh –

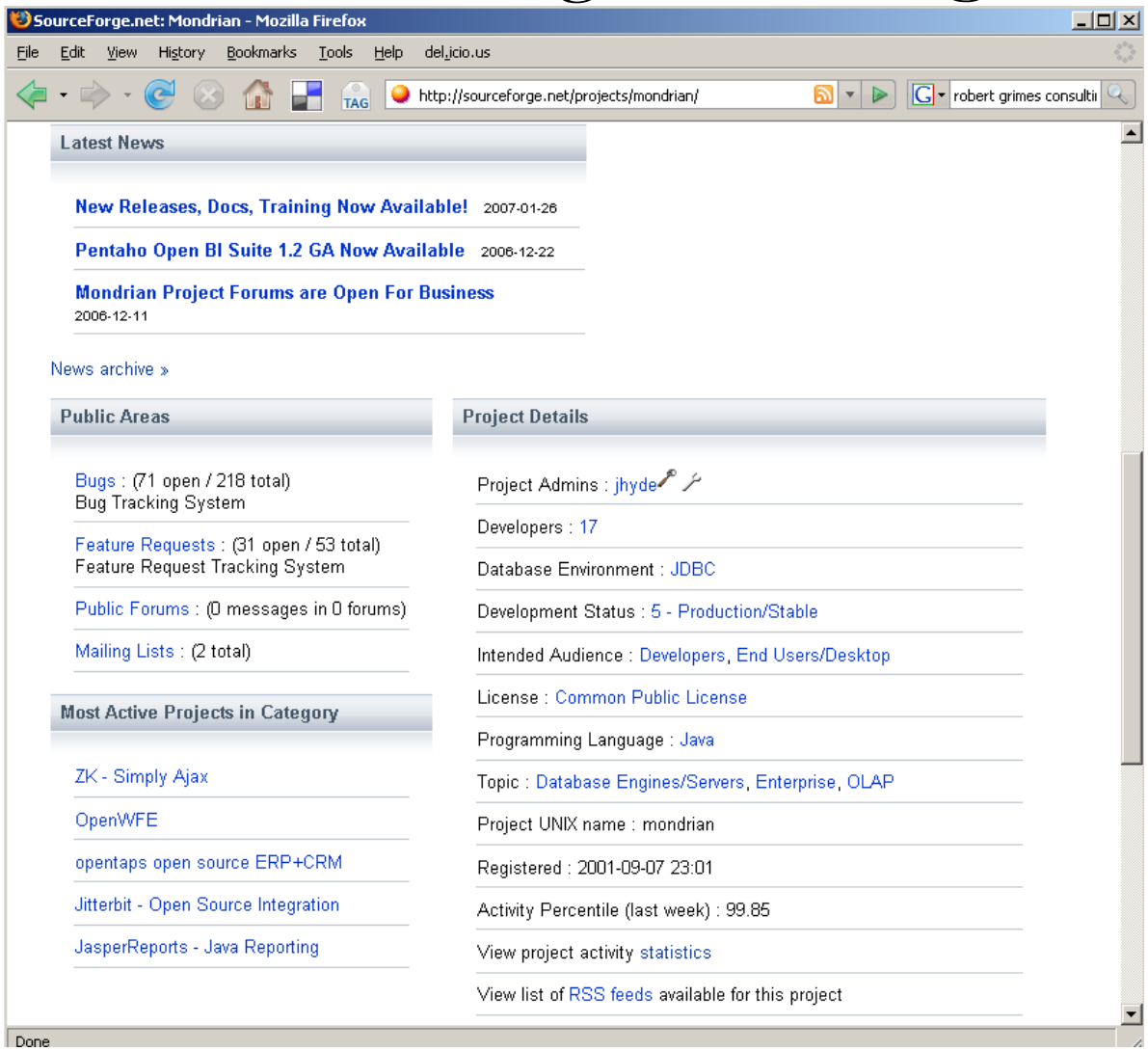
- "Mapping the open source world by collecting objective information on open source projects"
- "Ohloh collects software metrics from a variety of sources including the project's source code and the software development infrastructure used by the project's development team."
- "So far we've indexed over 3,000 projects and 220 million lines of source code." (2007)

www.ohloh.net/



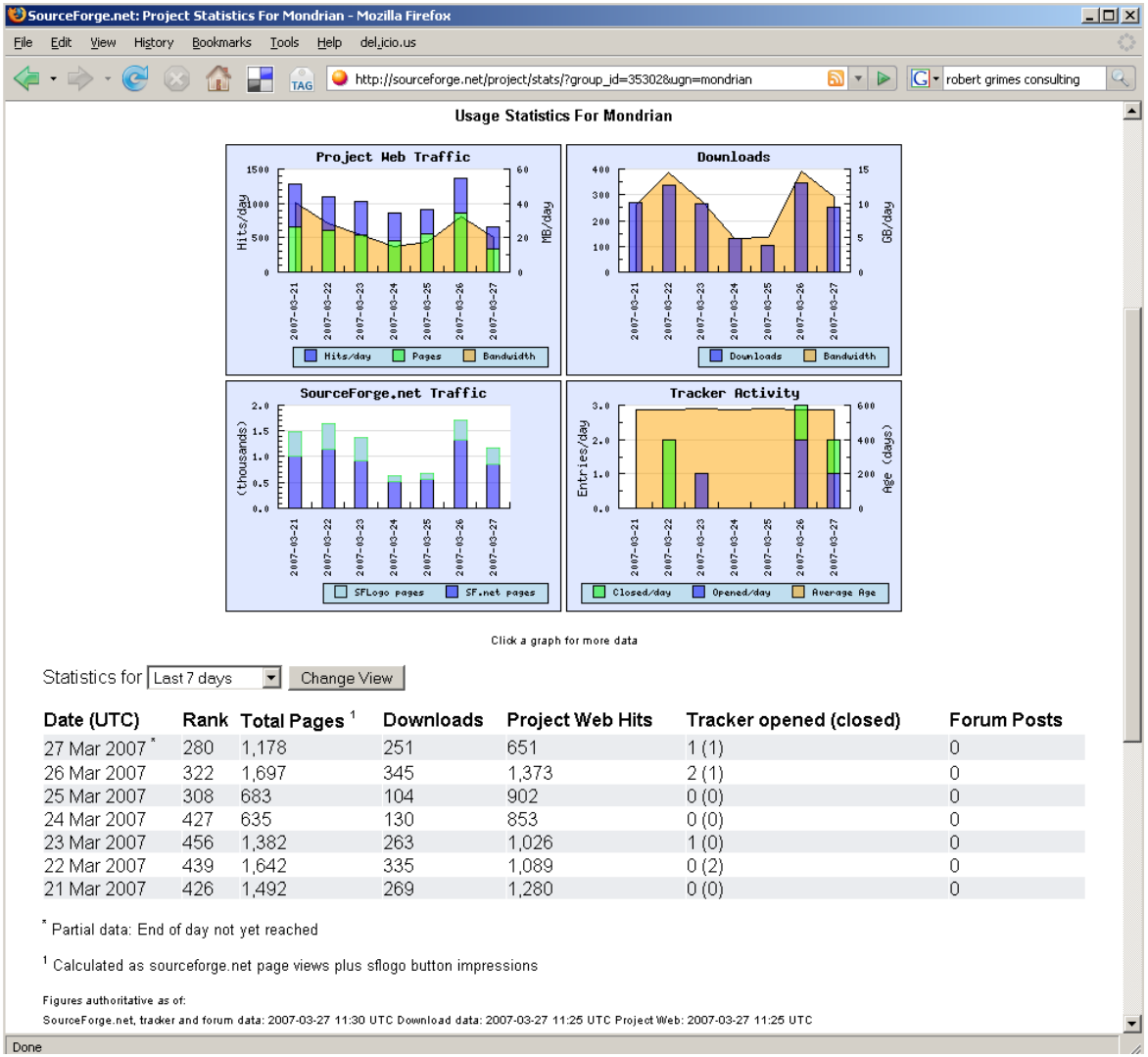
Assessing: *SourceForge.net*

Look for project information and for maturity and progress indicators.



Assessing: *SourceForge.net*

Look at project statistics.



Assessing: positioning

Look at download numbers…

… which may be distorted. What's being counted?

Review forum entries –

What's the nature of bug reports, feature requests?

What is the competitive positioning…

Relative to competing open- and closed source?

Related to your business domain, users, and IT environment?

Assessing: resources

What resources are required to support implementation and maintenance?

Assessing: licensing

What are the license terms?

- Are terms compatible with those of other software packages and hardware you're using?
  - Some hardware manufacturers' warranties are invalidated by open source (Linux) installation!
  - You must be extra support from Microsoft to run Windows in certain virtual environments!
- Will you create a derivative product for free or commercial distribution?

Assessing: licensing

What are the license terms?

- Does the free license scale to your needs?
  - Capability/functions.
  - Number of users allowed.
  - Number of installed nodes.
  - Number of objects (e.g., database tables).
  - Number of data records.

Assessing: project

How is the project run?

- Vision and roadmap.
- Past forks.
- Participation
  - Number of deciders, committers, and users.
  - Variety of backgrounds, skills, and interests.
- Sponsors and backers; partnerships and alliances.
- Management approach –
  - Use of mature management methods and tools.
- Documentation and supporting material.

Assessing: project

What support is available?

- Community.
- Vendor.
- Integrator.
- Consultant.

What in-house capabilities have you?

Assessing

Thesis:

Many or most business-related assessment questions can be avoided by going with a distribution of a supported stack.

Thesis:

Most technical evaluation points are the same or at least analogous for open source and closed solutions.

Assessing: OSMM

Open Source Maturity Model

- Developed by Capgemini in 2003.
- Factors grouped in categories.
- Scoring model that prioritizes factors.

www.seriouslyopen.org/nuke/html/modules/Downloads/osmm/GB_Expert_Letter_Open_Source_Maturity_Model_1.5.3.pdf

Assessing: OSMM

Usability – The intended user audience, the experience of that group.

Interfacing – Required connectivity, which standards are applicable. How does this fit into the organisation?

Performance – The expected load and processing capability. The performance demands that must be met.

Reliability – What level of availability should the product deliver?

Security – What security measures are required, what restrictions are imposed onto the product.

Proven technology – Does the product use technology that has proven itself in daily production?

Vendor independence – What level of commitment between supplier and user does the product demand?

Platform independence – Is the product available for particular ICT environments only, or does the product allow a wide range of platforms.

Support – What level of support is required.

Reporting – What reporting facilities are required.

Administration – Does the product allow the use of existing maintenance tools, the demands for operational management.

Advice – Does the client require validation / recommendation by independent parties, if so, what is required.

Training – Required training and facilities.

Staffing – Is product expertise bought, taught or hired.

Implementation – Which implementation scenario is preferred?

Assessing: Business Readiness Rating

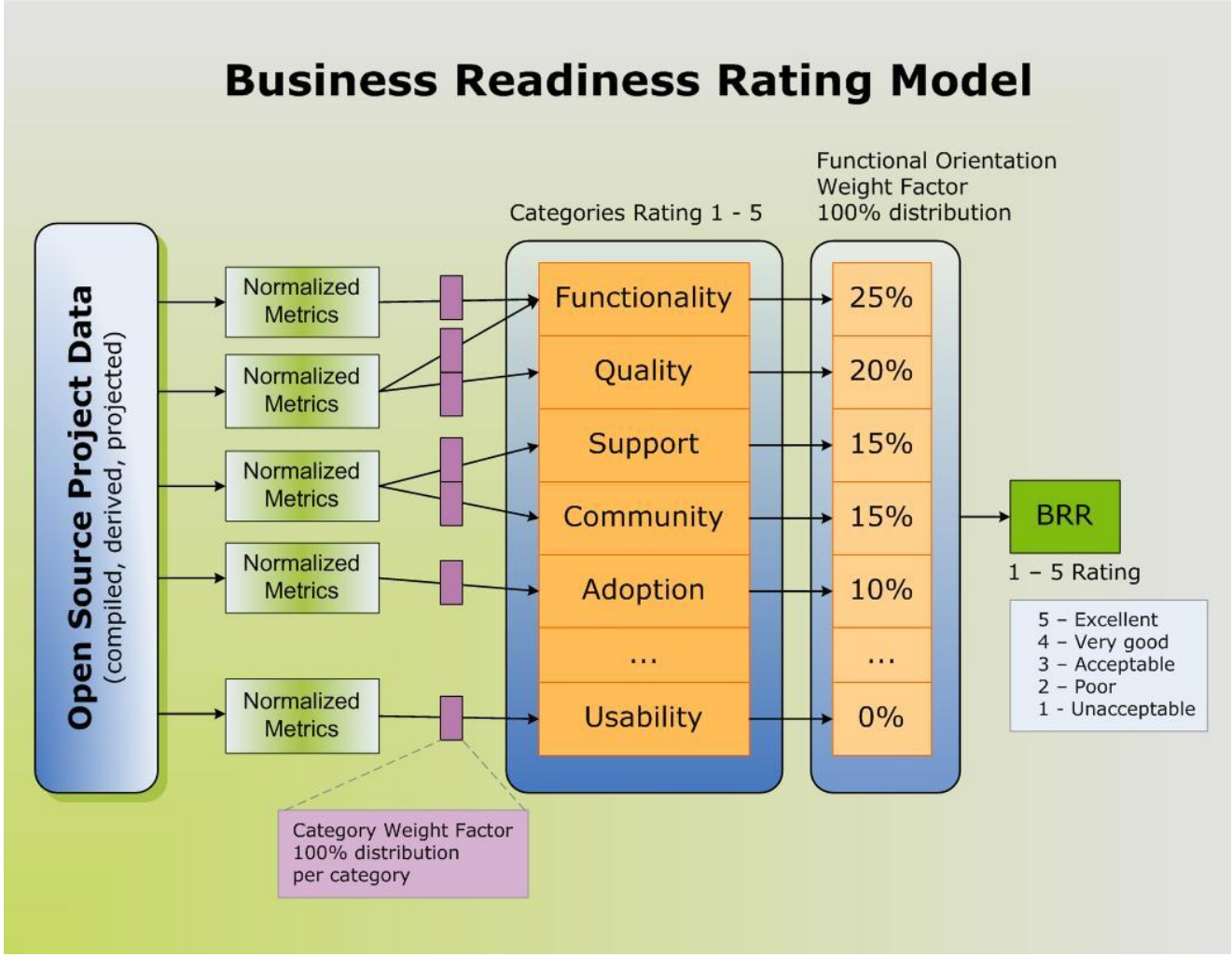
BRR was a 2005-6 attempt at "a new standard model for rating open source software."

- Invitation-only community sponsored by Carnegie Mellon West Center for Open Source Investigation, O'Reilly CodeZoo, SpikeSource,Intel.
- Currently moribund, but let's look at their material...

  www.openbrr.org/
- The project's 2005 white paper is worth a look:

  www.openbrr.org/wiki/images/d/da/BRR_whitepaper_2005RFC1.pdf

Assessing: Business Readiness Rating



www.openbrr.org/wiki/images/d/db/Business_Ready_Rating_Model.jpg

Assessing: Navica OSMM

Navica is a systems integrator that has published an Open Source Maturity Model.

- Structured process to objectively perform assessments along critical dimensions:
  - Software.
  - Support.
  - Documentation.
  - Training.
  - Integration.
  - Service availability.

www.navicasoft.com/pages/osmm.htm

Assessing: Navica OSMM

	Phase 1: Assess Element Maturity				Phase 2	Phase 3
	Define Requirements	Locate Resources	Assess Element Maturity	Assign Element Score	Assign Weighting Factor	Calculate Product Maturity Score
Product Software						
Support						
Documentation						
Training						
Product Integrations						
Professional Services						

Open Source Maturty Model

Product Requirements Template
OSMM Product Requirements Checklist

Product Requirements	
Method	*Notes*
Create Requirements Task Force	Consider: User Organization, Development/Engineering, Operations, Legal, Customer Service
Poll User Community for Requirements	
Group 1 Requirements	Purpose for Requirement
Group 2 Requirements	Purpose for Requirement
Review Applicable Standards	List formal standards and current status (e.g., draft, comment, etc.); also analyze future standards under development
	List informal or industry standards as appropriate
Standard 1 (Name)	Applicable functionality
	Current Status
	Future Status
Standard 2 (Name)	Applicable functionality
	Current Status
	Future Status
Review Commercial Vendor Product Materials	
Vendor 1 (Name)	Whitepapers
	Technical Specification Papers
	Functionality Checklist
Vendor 2 (Name)	Whitepapers
	Technical Specification Papers
	Functionality Checklist
Review Applicable Analyst Firm Materials	
Firm 1 (Name)	Functionality Checklist
	Vendor Recommendations
Firm 1 (Name)	Functionality Checklist
	Vendor Recommendations

Technical evaluation

- Not much different from evaluation of a commercial product.
  - Ascertain needs.
  - Prioritize requirements.
  - Create an assessment approach:
    - Function points to test.
    - Evaluation criteria and scoring.
    - Decision criteria.
  - Conduct evaluation.
  - Perform a Gap Analysis: What's missing?

Technical evaluation

Evaluate:

- Function points.
- Security.
- Integration.
- Interoperation.
- Non-functional criteria:
  - Scalability.
  - Usability.
  - Extensibility.
  - Performance and throughput.

Technical evaluation

Process:

- Design and conduct a functional evaluation.
- Create a proof-of-concept prototype.
  - Vertical slice: All functions for a subset of data or users.
  - Horizontal slice: Selected functions for all data.
    - Tests capacity and scalability.
- Create a nominal implementation plan.
  - Nominal = high-level.

Cost evaluation

You must account for:

- License fees.
  - (If a "professional" or "enterprise" or indemnified version.)
- Support costs.
- Infrastructure: hardware and computing environment.
  - Typically less for open source given platform scalability, deployment flexibility (e.g., add as needed).
- Required mediation, that is, gap bridging.
- Participation cost, that is, to "give back."

Migration strategy

Rationalize your computing architecture.

- Tiers:
  - User interface (UI).
  - Middleware, business logic.
  - Back-end, e.g., database, content management, and legacy services.
  - Hardware and network.
- Understand:
  - APIs (application programming interfaces).
  - Business processes.

Model your computing environment.

Migration strategy

Rebaseline your requirements.

- What functions are no longer used?
- What new functions are needed?
- What "non-functional" characteristics need improvement?

Assess your budget and staff.

Plan!

Migration strategy

"Big bang" system replacement.

- Faster.
- Cheaper if it works.
- Riskier.

"Divide and conquer" component replacement.

- You can start with a "face lift" replacing UIs.
- You can possibly replace the back-end systems or the middleware while supporting current UIs.
- You could possibly move (some of) the current software to an OS operating system.

Migration strategy

Infrastructure/encapsulation.

- Replace APIs, interoperation framework, for instance with SOA.

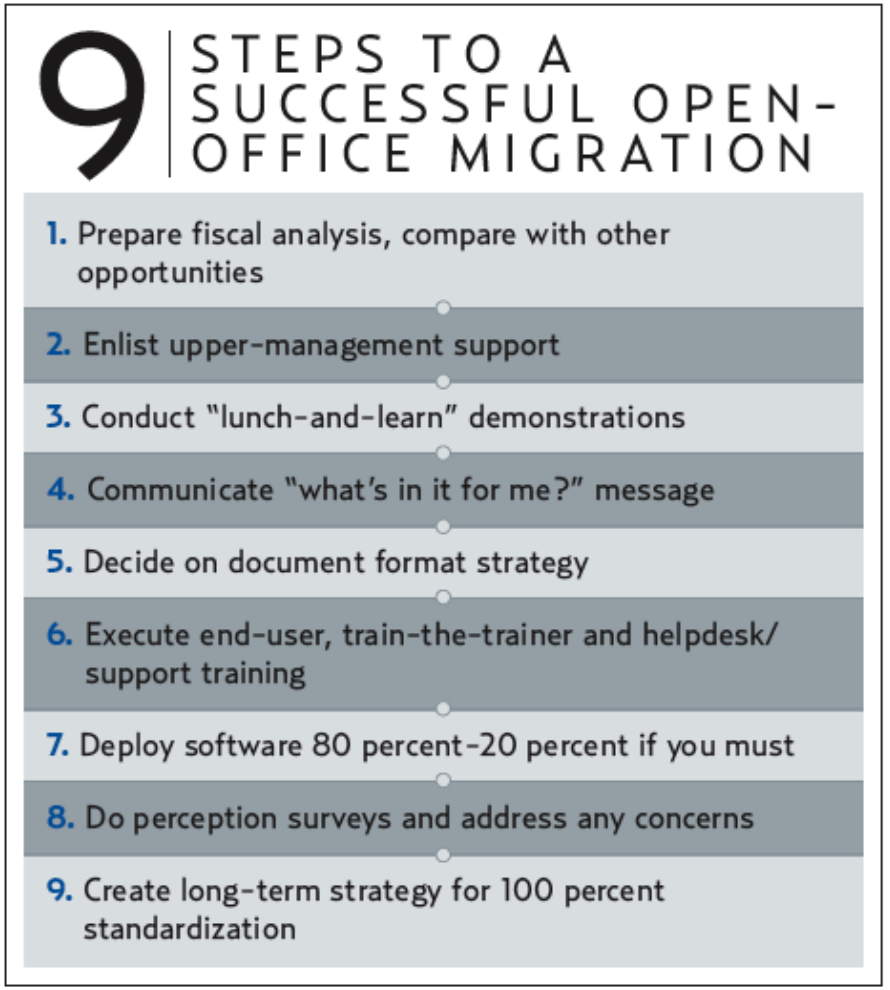
Platform.

- Move existing applications to a new operating system and possibly hardware.
- Not always feasible because of application dependencies.

Or simply "Contain and Retire."

Open Office migration

One take, from *Network Computing* magazine…



www.networkcomputing.com/gallery/2007/0219/0219f2b.jhtml

Open Office migration

Do a cost comparision…

COST SAVINGS OF A MICROSOFT DIVORCE AND AN OPENOFFICE MARRIAGE

	YEAR 0	YEAR 1	YEAR 2
Savings on Office licenses	$60,000	$60,000	$60,000
Training ("train the trainers")	–$10,000		
Productivity loss during training	–$20,000		
External support, hourly contract	–$8,000	–$5,000	–$5,000
Total	$22,000	$55,000	$55,000
NPV* of switching		$122,837	

ASSUMPTIONS/CONSTANTS	
Cost of OpenOffice and Evolution licenses	0
Number of office productivity users	1200
Cost of OEM office license	200
OEM license turnover/year	300
Cost of capital for organization	6%

*For a primer on discounted cash flow, an analysis that produces a project's NPV (net present value), see *nwc.com/showArticle.jhtml?articleID=171000416*.

Using the principles of discounted cash flow and applying opportunity cost to the cost of capital projects, this number represents the present value (net of inflows and outflows) of future cash expenditures and savings.

www.networkcomputing.com/gallery/2007/0219/ 0219f2e.jhtml

Policy

"Is it now possible to build a complete open source enterprise software stack? I put together the following table that suggests you could get pretty close."

-- Matthew Aslett, February 2006

www.businessreviewonline.com/os/archives/2006/02/a_complete_open.html

Anti-spam - SpamAssassin/SendMail/Thunderbird
Antivirus - ClamAV/Open Antivirus
Instant messaging - Jabber
Email client - Evolution/Thunderbird
Browser - Firefox
Office productivity - OpenOffice.org/Koffice
Mobile infrastructure - Funambol
Content management - Alfresco/Plone
Collaboration software - Openexchange/Sendmail/Zimbra
BI applications - Pentaho/JasperSoft/GreenPlum
CRM applications - SugarCRM/Compiere/Daffodil
ERP applications - Compiere/ERP5/OFBiz
Programming - PHP/Perl/Python
IDE/dev tools - Eclipse/Mono/NetBeans

Web server - Apache
ESB - Celtix/Mule
Middleware tools - JBoss/Apache/ObjectWeb
App server - JBoss/Geronimo
Database - MySQL/PostgreSQL/Ingres
File/print services - Samba
Virtualisation - Xen/OpenVZ
Operating system - Linux/BSD/OpenSolaris
Systems management - openQRM/Groundwork
Network management - OpenNMS/Groundwork
Backup/archive - Amanda
Application security - AppArmor/SE Linux
Storage management - Aperi
Security - OpenSSH/OpenVPN/OpenLDAP

Policy

Consider using a (commercial) solutions packager. Red Hat is one. Others include –

- SpikeSource.
- SourceLabs.
  - "SourceLabs simplifies and mitigates the risk for large companies adopting open source software."
  - SASH consists of 50+ open source projects and runs on all leading J2EE application servers.
  - SourceLabs tests, applies patches, packages software into a single distribution, and provides support, maintenance, and upgrades.

sourcelabs.com/

Policy

SpikeSource's package is called SpikeIgnited.

- Integration: Ensures that all open source components …work together seamlessly.
- Configuration: Easy installation and configuration of open source components.
- Testing and Security: Extensive testing, code reviews and packaging by SpikeSource …plus immediate notification of security issues and delivery of patches via SpikeNet.
- Maintenance: [Keep] up with open source infrastructure versions and updates.
- Support: 24x7 unlimited support from SpikeSource.
- Partners include JasperSoft.

Methodology

Sourcesense proposes the SHARE methodology.

- Educate: "Ensuring proper understanding of how Open Source works, what is the business model around it, what to expect from Open Source and how to mitigate risks."
- Extend: "Introduce Open Source in a structured and controlled way."
- Engage: "Sound strategic planning that takes into account the business, legal and community aspects of Open Source business."
- Partners include Pentaho.

Resources

Given that open source revolves around community and collaboration, there are many:

- Organizations.
- Portals.
- Conferences.
- Blogs.
- Projects.
- Publications.

I will provide a sampling…

Organizations

Free Software Foundation

www.fsf.org/

www.fsfeurope.org/index.it.html

The Linux Foundation

Merger of the Open Source Development Lab (OSDL) and the Free Standards Group (FSG)

www.linux-foundation.org/en/Main_Page

Open Source Initiative (OSI)

www.opensource.org/index.php

Open Solutions Alliance

osa.org/

Associazione per il Software Libero...

Organizations

L'Associazione per il Software Libero...

... è un'associazione senza scopo di lucro che ha come obiettivi principali la diffusione del software libero in Italia ed una corretta informazione sull'argomento.

Site at *www.softwarelibero.it/* contains a glossary, material on licenses, e-mail lists, etc.

GNUvox

"Le voci di FSFE e delle comunitá italiane del Software Libero: uno strumento per fare informazione, mettere in comune strumenti e cultura, sostenere iniziative condivise a sostegno della libertá di pensiero."

www.gnuvox.info/

Portals

Microsoft's Port 25.

port25.technet.com/

OSDir.com.

News, software tours.

osdir.com/

OSTG, the Open Source Technology Group.

A network of technology sites that includes:

Freshmeat.net.

Newsforge.com.

Slashdot.org.

Sourceforge.net.

www.ostg.com/

Conferences

O'Reilly's Open Source Convention

conferences.oreillynet.com/ oscon

LinuxWorld

www.linuxworld.com/

ApacheCon Europe

www.eu.apachecon.com/

Open Source Business conference

www.osbc.com

Open Source Think Tank

thinktank.olliancegroup.com/

Blogs

Matt Asay and Dave Rosenberg, Open Sources, in InfoWorld

weblog.infoworld.com/openresource

451 CAOS [Commercial Adoption of OS] Theory

blogs.the451group.com/opensource/

Alex Fletcher's Open Source Unleashed

alexfletcher.typepad.com/all_bets_off/

O'Reilly Open Source

radar.oreilly.com/open_source/

Dana Blankenhorn

blogs.zdnet.com/open-source/

Blogs

Roberto Gallopini, Commercial Open Source Software

robertogaloppini.net/

Gianugo Rabellino (SourceSense), Boldly Open

boldlyopen.com

Open Source Solutions

press.teleinteractive.net/oss

Planet PostgreSQL

www.planetpostgresql.org/

Projects

FLOSSMetrics

Sponsored by the European Commission.

Stands for Free/Libre Open Source Software Metrics.

Aim is to "construct, publish and analyse a large scale database with information and metrics about libre software development."

flossmetrics.org/

Publications

Open Source Forge

News, tutorials, articles, press releases.

www.osforge.com/

Enterprise Open Source Journal

Suspended publication, but back issues are useful.

www.eosj.com/

The 451 CAOS Community

www.the451group.com/caos/caos_community.php

Questions?

Discussion?

Thank you!