



— An IDG White Paper —

“IS YOUR DATABASE EATING OUT THE HEART OF YOUR IT-INFRASTRUCTURE?”

Pervasive Postgres can be a viable Open Source
alternative to commercial databases

Contents

1. Introduction	3
2. Open Source software: a collaborative software development method	4
Open Source licensing	5
3. An overview of Postgres	7
4. Is Pervasive Postgres an alternative to mainstream commercial databases?	8
Performance	9
Storage	9
Support and maintenance	10
Scalability	10
Platform support	10
Conclusion	11



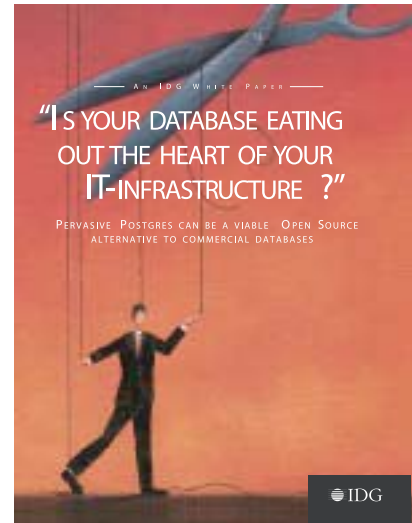
1. Introduction


Linux and many other open source software options have garnered acceptance by millions of users worldwide. These technologies offer an increasingly viable alternative to many commercial software products – operating systems, development tools, even databases. Depending on the specific needs of each client, the state-of-the-art commercial databases can still provide excellent value. Yet for mainstream applications, many users do not need the expense or feature overkill of the big brands. In many cases, a commodity database will suffice.

Does the acceptance of open source software fundamentally change the software industry as we know it? Can open source software provide a solid foundation for business applications? Which database is the appropriate fit for which applications? Is open source a viable business model? Can well-known software vendors sustain revenue based on support services? These are all important questions.

The obvious answer is... the customer is always right! For some applications basic open source software tools provide sufficient value. For more business critical situations, the same tools provide a solid foundation and form a complete solution when coupled with software assurance, platform support and professional services. Several software companies, like Pervasive, have chosen such a path. So is there an alternative to proprietary enterprise software, one that avoids the anarchy of open source do-it-yourself components? Yes. This White Paper reveals a third option.

***You only need your insurance when you're ill, don't you? Yet that's not how it works.”
(Petra Heinrich, Novell/SuSE)***





2. *Open Source software: a collaborative software development method*

Open Source software originated as a community driven phenomenon. Originally, users acquired software in a hardware-software bundle that included full source code. In the late 1970s, the emerging software industry changed this practice, and access to source code, including the ability to customize software for local needs, gradually disappeared. The open source software movement then emerged as the intellectual counterpart to this trend. The Free Software Foundation, among others, has been carrying the flag of Free Software since the late 1970's. The open source movement strives to allow and promote software development in mutual collaboration through sharing of public code, while also protecting developers' own intellectual ownership rights.

A first high-profile open source project was Richard Stallman's GNU – short for 'GNU's not Unix', a barebones flavor of the Unix operating system. The Linux operating system quickly became another high profile project after Linus Torvalds developed a Unix-like system kernel and published it in 1991. At this point, the open source community collided with the Internet revolution, bringing to life a very powerful movement capable of extending Linux from a small-scale student's project into a de facto industry standard.

Yet there have been many more open source software projects, ranging from highly successful to nearly invisible. According to a recent survey from Evans Data Corporation, over 1.1 million developers are working on various open source projects in North America alone. Successful examples are the Apache web server, the Sendmail e-mail server, and the Eclipse integrated development environment. Several databases have achieved open source success, including PostgreSQL, and more recently MySQL.

Although not essentially part of the Free Software concept, most open source software projects have been offered free of charge. Enforcing high licensing fees would not be easy, since anyone with basic skills of compiling source code could obtain a copy for free. In response, several large software vendors are currently working to construct a legally sound framework for offering warranty, certified interoperability, and support services.

Commercial software vendors are increasingly adopting the open source software model, often as a means of providing community-based customer support. Adopting this model necessitates a lower cost structure for the vendor than traditional direct-touch support models. Is this new development simply a ploy to revive interest for older software product lines approaching end-of-life? Was this the rationale for the open source afterlife that the venerable Ingres was awarded in 2004? Is the open source community the new paradigm for the software industry? Or is it the place where old code goes to die?



Looking at the open source database market, it is likely to be both a growth and a maintain strategy. According to Noel Yuhanna, senior analyst at Forrester Research, the open source database market will be growing from 120 million USD in revenue to 1 billion USD in only three years. This clearly indicates an emerging business model. Concurrently, we see other vendors releasing non-critical product lines as open source, the so-called “abandonware” movement.

Open Source licensing:

Although the principal aim of most open source advocates has been to develop software, the high rise of the open source movement would have been impossible outside the legal framework of copyright and licensing. During the last decade, advocates and users have tried out many formats, each with different implications for using, adapting and commercializing software. One thing they all have in common is that the code is made public – mostly published on the Internet.


The availability of public source code – and the ubiquity of the Internet – has initiated a whole new methodology, more similar to that of the academic world than the typical 'closed source' software industry. Advocates of 'closed source' software have even described some of the licensing schemes as 'viral' in nature. Since many open source software licenses stipulate that additions and improvements on published code also be made public under the same license, this statement is true to some extent.

A short survey:

- GNU Public License (GPL): Licensing scheme used and promoted by the Free Software Foundation, used by many open source projects including GNU, the Linux kernel development, and Sendmail (e-mail server). It uses the label 'copyleft' – denying any adaptation of the code or incorporation in other applications, unless they are made public under the same terms. The ethical objective of the GPL is to create an island of free software, and the “viral” license language exists to force applications that touch GPL covered code to move to that island. The “Lesser” GPL (LGPL) is a less stringent derivative of the GPL, essentially allowing LGPL-protected code (typically software libraries) to be incorporated in closed source applications without causing a 'viral' effect.



- Berkeley Standard Distribution (BSD): Simple software licensing scheme initiated by the University of California, Berkeley. It allows adaptation and redistribution of licensed software code, provided the initial owner is mentioned as the copyright holder and with a warranty disclaimer. The BSD license is generally perceived a business friendly basis for collaborative code development and deployment. Variants of the BSD license include the Apache license and the Mozilla license.
- Creative Commons: Inspired by the GPL, the recently developed license (2001) was created as a legal framework for 'creative works' in general. Summarized as "Some Rights Reserved" the CC-license allows copying and redistribution with correct attribution of the owner, but no third party commercialization.
- Gratis. Some vendors give away executable versions of their software for free, usually as a mechanism to generate demand for paid versions. Obviously this is not an open source approach, but does provide users with zero cost software.
- Public domain: This is not exactly a licensing scheme, but it is an important legal concept, mainly in the U.S. Public domain software is owned by the public, and nobody can claim further ownership or patents.



3. *An overview of Postgres*

The PostgreSQL relational database system is a result of the POSTGRES project established at the University of California at Berkeley by Michael Stonebraker, also architect of INGRES, one of the very first (°1972) relational database systems. INGRES was later developed into Sybase (and is therefore a distant parent of Microsoft SQL), Informix, and several other commercial database systems. POSTGRES was started in 1986 as the next generation of INGRES, using funds from DARPA, the National Science Foundation, the Army Research Office and ESL Inc. It took four years for Stonebraker and his students to construct the basis of what is now PostgreSQL.

POSTGRES originally used a language called PostQUEL for retrieving database information. Following the industry trend towards SQL, an interpreter was added in 1994. The project was licensed under the BSD license and renamed PostgreSQL shortly thereafter. The open source community happily picked up the project and contributed additional features. The liberal BSD license allowed for several commercial derived works including Illustra, which merged with Informix.

In January 2005, Pervasive Software announced a private label version of PostgreSQL called Pervasive Postgres, as well as production support for PostgreSQL users and a suite of related professional services. Pervasive's product and support offerings are based on PostgreSQL version 8, the first to provide fully native Windows support. As with other open source business models, Pervasive will offer the database for free and charge for a range of support packages, ranging from 99 to 4999 USD per year per server. Pervasive technicians will be available for five-day engagements to migrate customers from Oracle, DB2, Sybase, MySQL and other databases, and also for ongoing traditional professional services like tuning and training. Besides services, the company promises a full suite of integrated software and a roadmap for future improvements, beginning with automated installation and simple administration tools. All improvements to the core PostgreSQL server will be contributed back to the PostgreSQL development community.

Open-source databases are following a similar path to Linux in the enterprise. According to industry observers, the movement started with low-end uses and is gradually moving up into mission critical business applications. No single company controls an open-source project, and development is shared amongst enthusiasts and employees of various companies. In theory, and increasingly in practice, this development model provides low-cost, commoditised replacements for much enterprise software infrastructure, analysts say.

4. *Is Pervasive Postgres an alternative to mainstream commercial databases?*

Adding commercial support from a well-known database vendor still does not make up an enterprise database. Can Pervasive Postgres rival industry leaders such as IBM, Oracle and Microsoft? PostgreSQL's BSD licence allows royalty-free use in a commercial setting, and the ongoing roadmap includes some needed extensions. PostgreSQL already has enterprise features lacked by many other open source databases, such as views, triggers, stored procedures, and rich security. These features will make it easier to port existing applications from commercial databases like Oracle or DB2.

PostgreSQL stacks up well against commercial rivals, said Noel Yuhanna, senior analyst at Forrester Research. "It certainly has some robust enterprise features, which are comparable to some of the commercial databases." The syntax of PostgreSQL's stored procedure language PL/PGSQL is very similar to Oracle's PL/SQL, while also comparing to Informix' SQL 99-compliant language. "Our estimate is some four out of ten Oracle users would be able to migrate to PostgreSQL right away, and four more with some migration planning and effort," stated Pervasive's Lance Obermeyer.

Features:

PostgreSQL has many of the database features that Oracle, DB2, or MS-SQL have – including views, inheritance, sequences, cursors, and user-defined data types, in addition to the aforementioned views, triggers, and stored procedures. Hence PostgreSQL offers an easy migration path from commercial databases..

Data Integrity: ACID compliance, Row-level locking, Partial rollbacks, Savepoints, Point in Time recovery and Multi Version Concurrency Control (MVCC)

Capabilities: Stored procedures, Views, Triggers, Sequences, Cursors, User-defined data types

Indexes: Single column, Multi-column, Primary key, Full text, relational integrity

Replication: Single master/multiple slave

Interface Methods: ODBC, JDBC, C/C++

Performance:

When talking about open source databases, many in the open source community focus their attention heavily on MySQL. The product of a Swedish open source development company, MySQL is highly popular on the Internet for powering database-driven websites. MySQL is generally perceived as simple, but delivering excellent performance for low-level read and write operations. While it is capable of fast reading performance MySQL suffers serious shortcomings in data management features. "MySQL is the database equivalent of a cardboard box. Sure you can keep things organized in a simple way just using a cardboard box", said Jan Guldentops, researcher at BA Testlabs. Conversely, PostgreSQL's deep technology around indexing, optimization, and concurrency control provide it a speed advantage for demanding workloads, like complex queries or tens to hundreds of concurrent users. Also, the use of stored procedures can move processing from the application into the database, further benefiting performance.



MySQL is one of the second-generation open source companies, actually owning its entire code base. While its database software has been published under the GPL, the company simultaneously sells 'closed source' commercial licenses to anyone wishing to commercially deploy or distribute the technology. PostgreSQL uses a non-threaded model where every new connection to the database gets a new database process. Creating a new process for each user slows startup time, though freely available connection-pooling software can reuse existing connections without starting new processes. On the other hand, the process model offers guaranteed isolation between users and increased stability since a system fault cannot crash the entire system.

Storage:

PostgreSQL version 8 has tablespaces, a mechanism for administrating multi-gigabyte databases. This feature allows an administrator to control the placement of large tables and indexes onto individual disks or arrays, improving query performance and storage efficiency. It also provides methods for performing hot-database backups (backing up a database without shutting it down). PostgreSQL uses a system called write ahead logging to ensure database consistency. Like datatables, the placement of system logs is controllable, providing increased safety and throughput.

Support and maintenance:

PostgreSQL has enjoyed splendid community support, just like many other open source projects. Yet the commercial support offered by Pervasive, which includes defined response times and service levels is likely to add tremendous value for business customers that do not want to gamble on support.

Scalability:

Scalability and high availability are often synonymous with clusters. This is where the commercial database platforms are still well ahead of open source. Multi-server clusters, whether for performance or reliability, remain anchored to products like Oracle's 9i and IBM's DB2 UDB, and increasingly Microsoft's SQL Server. But there are a host of new application areas that do not require the complex, and expensive, features found in these products.

Platform support:

While both Oracle and DB2 run on multiple platforms, Microsoft's SQL Server is limited to Windows. PostgreSQL supports many different platforms, including Windows, Linux, FreeBSD, and MacOS X. The PostgreSQL on Linux combination will be the logical choice for anyone involved in the community over the last few years. The latest PostgreSQL release will also run natively on Windows – making it an even better alternative for many Oracle DBAs currently working on the Microsoft platform. PostgreSQL's PL/pgSQL language for writing procedures is highly similar to Oracle's PL/SQL.



Conclusion

Pervasive Postgres, more so than any other open source database offering is perfectly suited to challenge vendors like Oracle, IBM, Microsoft and Sybase in mainstream business applications. Quite simply, most users do not need many of the features these vendors offer. The path Pervasive has taken is to support the existing open source database while simultaneously pursuing an ambitious development roadmap, making it a real alternative to the mainstream databases over the next few years. Similar to PL/SQL, Pervasive Postgres is extremely interesting to Oracle DBA's in need of a less complicated and less expensive solution for second tier applications. Furthermore, Pervasive Software's professional support and services can provide a digital life insurance policy, easing the process of adoption.

“Open-source developers are in much the same position today that IBM was in 1981 when it changed the rules of the computer industry, but failed to understand the consequences of the change, allowing others to reap the benefits. Most existing proprietary software vendors are no better off, playing by the old rules while the new rules are reshaping the industry around them”

(Tim O'Reilly, in “How open is the future?”, Brussels, 2005)

“Postgres offers me most of the essential features found in Oracle, with many of them implemented in a more logical and more ANSI-compliant way. Still, documentation is scarce and there are few outside options.”

(Robert Beer, "long-term Oracle DBA and recent Postgres convert", Computerworld)

