AberdeenGroup

Open Source Databases: All Dressed Up, Only So Many Places to Go

An Executive White Paper

March 2004

Aberdeen Group, Inc.
260 Franklin Street
Boston, Massachusetts 02110-3112 USA
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

Open Source Databases: All Dressed Up, Only So Many Places to Go

Executive Summary

This Aberdeen *Executive White Paper* details the pros and cons of today's open source databases. These databases not only are widely downloaded by more than 10 million developers but also now have a track record of several years in support of key applications for some users. Recently, the major remaining suppliers of open source databases after the inevitable market shakeout — MySQL, PostgreSQL, and Sleepycat Software's Berkeley DB — have added significant new features that allow them to be compared in the same breath with so-called enterprise databases.

The time is therefore right for a reassessment of open source databases as an option for any developer or IT manager. How scalable, robust, flexible, manageable, and programmer productive are they? Where can they best be used? Do they really deliver advantages in cost of ownership (as opposed to license costs), and if so, when? What are best practices for implementation and maintenance?

During this assessment Aberdeen:

1. Surveyed today's open source database market and usage (This survey included in-depth qualitative interviews with approximately 15 users of open source and competing databases.)
2. Assessed the state of the art in open source database technology and noted how users view MySQL, PostgreSQL, and Berkeley DB
3. Identified types of database usage that are particularly appropriate for open source databases
4. Identified key target applications for open source databases
5. Compared open source databases with the competition
6. Discussed emerging "best practices" in using open source databases
7. Assessed the trends and projected usage of open source databases

Aberdeen's findings indicate that open source databases are indeed delivering medium-sized enterprise-scale performance, scalability, robustness, and manageability — open source databases are indeed "all dressed up." However, Aberdeen's interviews also show that today's proprietary installed bases and ISVs show little interest in replacing existing databases with open source ones — leaving open source databases with only so many places to expand in, such as being used in new application development, serving as an alternative database for ISVs, and providing additional services to their installed base. In these areas, over the next two to three years, open source databases are definitely worth investigating and prototyping.

Section I: Setting the Stage

Motivations for Choosing Open Source Databases

The reasons that users and ISVs (independent software vendors) choose open source products over proprietary products include:

- Control over maintenance and support
- Availability of source code
- Cost, such as for licenses and subsequent upgrades (Some cite development cost and TCO as well, but others find that open source database development cost and TCO are greater than for a proprietary alternative.)
- Flexibility to add new features as needed (Note that users rarely change open source databases.)
- Reliability

These reasons hold for DBMS users and for ISVs that build applications around a DBMS. They cite strong support (from the open source community) and control over maintenance as primary reasons for using an open source DBMS.

Users of open source DBMS products have often moved away from proprietary products because they have tired of the poor support that they receive from proprietary vendors. With open source, you can often download patches and install them in hours. In fact, some companies that have transitioned to Linux from Unix platforms buy minimal support from the Linux distributors and rely instead on support from the open source community. This approach comes at a cost — usually small — to the user or ISV because it requires that at least one employee work very closely with the open source community that supports the user's open source products.

The Open Source Database Leaders

Enterprise users and ISV developers surveyed by Aberdeen typically cite three open source databases as especially attractive:

- MySQL
- PostgreSQL
- Berkeley DB

The MySQL Open Source Database

MySQL is a full relational database with SQL support. Users view MySQL as simple, small footprint, and performant.

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

MySQL offers the following key features beyond the relational "basics":

- Tablespace backups, i.e., the ability to back up a set of tables rather than the entire database
- Deadlock detection
- ACID (atomicity, consistency, isolation, and durability) compliance (That is, MySQL guarantees that some system failure will not corrupt the database and make it transactionally inconsistent.)
- Row-level locking (typically required by large packaged applications)
- Open Database Connectivity (ODBC) support

MySQL does not offer the following key features typically provided in enterprise databases:

- Incremental and online/parallel backup/restore
- Encryption (security)
- Bit-mapped indexing (for large data warehouses)
- Single GUI administrative interface
- Views
- Stored procedures and triggers
- Object (complex data) table or data-type support

MySQL AB indicates that improvements in areas such as stored procedures are contemplated, but the next version of MySQL may arrive sometime in 2006, and the company is not committing to any particular improvements.

Users interviewed indicate that MySQL is adequately robust in small and medium-sized applications.

The PostgreSQL Open Source Database

PostgreSQL is a full relational database with SQL support. Many users view PostgreSQL as more complex and powerful than MySQL, but much less powerful than an enterprise database. Interestingly, some knowledgeable MySQL users view MySQL as more scalable than PostgreSQL.

PostgreSQL offers the following key features beyond the relational "basics":

- Tablespace backups, i.e., the ability to back up a set of tables rather than the entire database — with limits
- ODBC support
- Online backup (with limits)
- ACID compliance for transactions
- Deadlock detection

- Unicode support (non-U.S. character sets)
- Views (read-only, i.e., with limits)
- Stored procedures and triggers (with limits on programming-language support, including no support for Java)
- Object table and data-type support

PostgreSQL does not offer the following key features typically provided in enterprise databases:

- Incremental and parallel backup/restore
- Encryption (security)
- Deadlock detection
- Row-level locking (typically required by large packaged applications) — although PostgreSQL does offer alternative sub-table locking schemes
- Bit-mapped indexing (for large data warehouses)
- A single GUI administrative interface
- View update/insert/delete

PostgreSQL users indicate that improvements in some areas are contemplated, but the next version of PostgreSQL may come as late as two years from now, and the community that drives PostgreSQL development is not committing to any particular improvements.

Sleepycat's Berkeley DB

Berkeley DB offers many of the same key features as a traditional database, but in a very different package. Berkeley DB supports ACID transactions and recovery; multiple processes and multithreading for high concurrency; and replication for fault tolerance and high availability. Berkeley DB is a library that runs directly in the application, not client-server. It also is accessed via simple programmatic interfaces. It does not support SQL, ODBC, or JDBC.

Berkeley DB provides the following key features:

- ACID compliance
- Database dump and load utilities
- Support for main-memory and/or on-disk databases
- Page locking
- Online backup
- Load-balancing support
- Recovery from single-node failure
- Replication for high-availability applications

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

Sleepycat also provides Berkeley DB XML, an XML layer built atop Berkeley DB for the storage and retrieval of native XML data.

Perhaps the most notable user of Berkeley DB is Sun Microsystems, which uses Berkeley DB as the embedded data store in its Sun Java Enterprise and Desktop Systems. Sun has already deployed Berkeley DB widely as the data store in its Directory.

Key Characteristics of Enterprise Databases

The right-hand column of Table 1 shows the key features that users expect from major enterprise databases, such as IBM, Oracle, and Microsoft. The left-hand column shows the main benefits that users expect from these features.

Table 1: Key Benefits and Technologies for Enterprise Databases

Criterion	Related Technologies
Scalability	OLTP scalability technologies: · SMP support · Clustering support · Multithreading Decision-support scalability technologies: · Query optimization · Replication · Cost optimization · Bit-mapped indexing "Mixed" scalability technologies: · Stored procedures · Distributed database synchronization (2PC) · Running the application "inside the database" · Load balancing
Robustness/administrative costs	· Online, parallel load and backup/recovery, monitoring, and metadata management · "Zero-administration" tools and automated reorganization · Cross-database tools

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

Criterion	Related Technologies
Flexibility	· Standards support - SQL- Java/EJBs - XML · Complex-data-type support - Rich-media support - File support - Object and object-relational technology - Content management technology · Integration with infrastructure software - Application servers - Infrastructure APIs and component libraries
Programmer-productivity support	· High-level development environment support ·Java programming support ·Frameworks

Source: Aberdeen Group, March 2004

Of these technologies, interviewees over 10 years and five previous editions of Aberdeen's *Database Buying Guide* often cited four technologies that help IT buyers distinguish between "low-end" and "enterprise" or "enterprise-ready" databases:

1. Stored procedures (and triggers)
2. Replication (and two-phase commit)
3. SMP (symmetric multiprocessing) support
4. Cost-based query optimization

All of these technologies testify to an enterprise database's ability to scale.

All of today's open source databases are seen today as lacking especially in scalability, and to a lesser extent in robustness, flexibility, and programmer support. Therefore, they are not classified as "enterprise." Many are clearly deficient in at least the first three aforementioned technologies — they do not offer (or offer limited) stored procedures, do not offer two-phase commit, and do not offer exceptional multiprocessing support. However, interviewees' perception is that in many cases, today's open source databases are enterprise class in scalability and robustness. Moreover, like databases offered by lower profile proprietary suppliers, such as Pervasive and Progress, open source databases offer exceptional manageability, requiring little maintenance compared with major enterprise databases.

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

Section II: Open Source Database Markets and Use

A. Market Analysis for the Overall Open Source Database Market

The database market was large ($10.5 billion in revenue) in 2003. It will pick up steam in 2004 and beyond, although it will not return to the heady days of steady 35% growth that was seen in the 1990s. The open source database market, measured in revenue, is a very small fraction of the overall database market, with less than $100 million in revenue in 2003.

Functions

Typically, the database market is sliced first by transaction type:

- OLTP — Online transaction processing is usually a stream of updates, inserts, and deletes. OLTP is common in customer-interaction software, such as airline reservation systems, order-processing systems, and stock market information bases.
- Decision support/data warehousing — Typically, queries involve no updates, with mass inserts once each day (load). Data warehousing is primarily used by data miners (experts in statistical analysis) to gain insights from OLTP and other "mixed" information bases, without overloading those systems with the "query from hell" that prevents the systems from carrying out their main tasks.
- Mixed/application serving — Part update/insert/delete, part reads (access, but do not change one data item), part query. This market segment includes embedded databases that are dedicated to containing the data from one application, as well as Web databases that deliver rich-media data to a Web site.

Another common operation in legacy databases is batching — carrying out a sequential stream of updates, inserts, and deletes offline. In the past, batching has allowed OLTP systems to handle less important updates in a delayed fashion without burdening the business-critical customer-interaction function. However, batching is now a minor or vanishing part of most real-world databases.

Table 2 shows trends in the three key functions of databases over time. In the last five years, the data warehousing submarket has crested and matured, while the mixed market has come to dominate the overall database market because of the rapid increase in mixed Web database usage. Over the next five years, Aberdeen expects recent trends to continue, with the OLTP and decision-support submarkets decreasing in market share, while "mixed" applications continue to increase in market share despite the recent ebb in Web spending.

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

Table 2: Trends in Database Workloads

Function	1997	2002	2007
OLTP	70%	30%	25%
Decision support	15%	30%	25%
Mixed	20%	40%	50%

Source: Aberdeen Group, October 2003

The open source database market predominantly involves "mixed" transaction streams, especially Web ones. Open source database users value the ability to handle wide variations in load, and they process both large objects/text and relational data. They are less likely to value ultra-high performance, "extreme" scalability in numbers of users or amount of storage, and a high degree of administrative "ease-of-use."

Platforms

A second useful way to slice the database market is by operating system. Table 3 indicates trends over time in the operating systems used, and therefore on favored platforms. Because Aberdeen based Table 3 on revenue rather than units, it may underestimate the degree of commoditization of databases and the penetration rate of NT-based and Linux databases.

Table 3: Trends in Database Platforms (Share of Overall Database Revenue)

Operating System/Platform	1997	2002	2007
Unix/Linux	60%	50%	50%
NT	10%	25%	35%
Mainframe/proprietary	35%	25%	15%

Source: Aberdeen Group, October 2003

Linux is strongly associated with open source databases. Although Linux is displacing Unix in many shops, high-end Unix implementations are not changing to open source databases, although significant numbers of low-end SCO Unix users are changing over. The primary open source database market is new Linux applications.

Aberdeen expects open source databases on Linux to encourage the continuing commoditization of databases and the decrease in average database prices. However, the difficulty of migrating from legacy databases will prevent complete com-

moditization of the database market and offer a good continuing revenue stream to existing legacy database suppliers, such as Computer Associates and IBM.

Suppliers

Figure 1 shows our estimates of the market shares of the major database suppliers.

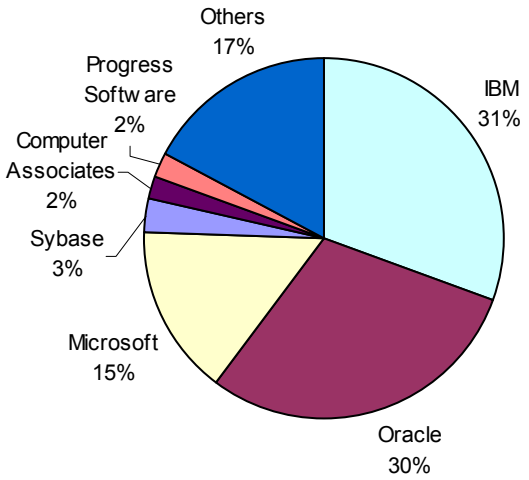
Although open source database suppliers continue to be a small part of the market as measured in revenue, they are a much larger part as measured in installations (more than 10 million).

Size or Level of the User Base

A fourth useful way to slice the database market is by the size or "level" of the user base. Traditionally, users are classified as follows:

- High — Large enterprise data centers, in which a large enterprise has more than $50 million in revenue
- Medium — A fairly large-sized enterprise with $25 million to $75 million in revenue or a division, line-of-business, or "mass-deployment" set of workgroups within a large enterprise
- Low — An SME (small to medium-sized enterprise) with less than $25 million in revenue or a workgroup within a large enterprise

Figure 1: Market Shares of the Major Database Suppliers



Source: Aberdeen Group, March 2004

Table 4 shows the shares of overall database market revenue that each submarket claims today. These shares tend to track the purchasing power of particular sizes of enterprises and should therefore show relatively little change over the next two years.

Table 4: Database User Size and Market Shares (of Overall Database Revenue)

Size of Database User	High	Medium	Low
Market share	54%	28%	18%

Source: Aberdeen Group, March 2004

Open source databases fall almost exclusively in the low end of the market. The SMB space has been doing somewhat better than other areas of the market, and, therefore, open source databases should gain in market share along with the rest of the low database market.

Niches

A final way to slice the database market is by global, or "enterprise," databases versus more specialized databases. These database specialties, or niche markets, in descending order of revenue size, include:

- "Embedded" databases that support particular applications or run "hidden" on small-scale devices (perhaps $400 million)
- Content management databases and other "specialty" databases (perhaps $400 million)
- Mobile/desktop/workgroup databases characterized by a small "footprint" (perhaps $150 million)
- "Second-tier" databases that act as "cache" or high-performance databases on the second tier of a Web architecture, with an enterprise database on the third tier (perhaps $100 million)
- Open source databases (less than $100 million)

Open source databases therefore are currently (and in the next year) the smallest separate database submarket.

Summary: The Open Source Database Market

The open source database market is small as measured in revenue — Aberdeen estimates it at less than $100 million worldwide — but millions of units are already deployed. These databases have not yet made major inroads into established suppliers' installed bases at the high or low ends. However, open source databases now have a strong niche in the low-end space.

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

Open source databases are running particularly on the Linux platforms that have superseded SCO's low-end Unix. They have achieved significant Windows penetration. These databases are especially attractive to highly technical developers. They are strong in such low-end, Unix-friendly vertical markets as telecommunications and retail. Finally, Linux in embedded applications, such as consumer electronics and cash registers, is displacing proprietary operating systems to some extent. Open source databases have an increasing role in embedded applications.

The history of the database market shows that many of today's open source database users will inevitably look for more features than today's suppliers offer as their applications scale. Interviews with open source database suppliers indicate that these suppliers will be slow to provide the features needed most (stored procedures, cost-based optimization of queries, and two-phase commit). However, interviews with users suggest that they see no immediate need to change databases.

Likewise, the installed bases of enterprise databases are, as shown by their history, strongly committed to them. Aberdeen does not expect significant numbers of these users to migrate their in-house applications to open source databases.

Users of large-scale packaged applications such as SAP and PeopleSoft value scalability and robustness, but are willing to consider a less functional solution, such as Microsoft SQL Server. However, it usually takes several years and careful partnerships with the packaged application suppliers to "crack the market."

Users of smaller scale vertical packaged applications are much more willing to consider nontraditional solutions. However, these users strongly demand not only greater scalability but also "near-lights-out" robustness; ease of administration, deployment, and upgrade; and strong programming support for rapid development and customization. Today's solutions would require significant enhancements to compete with Progress, Pervasive, Sybase, and now IBM DB2 Express.

The largest area of usefulness for an open source database is therefore in the area of new in-house applications, both low end and high end. This market is currently relatively large (in the hundreds of millions), but it is "in recovery," especially with offshore outsourcing of in-house development. In the low end, key user buying criteria are low license cost, rapid Java development support, and sometimes scalability. In the high end, key user buying criteria are scalability, robustness, flexibility (to fit in enterprise architectures), and Java development support.

B. Key Potential Applications for Open Source Databases

Table 5 shows examples of key potential uses and applications for particular vertical industries.

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

Table 5: Key Database Vertical Markets

Vertical Market	Example of Companies	Typical Database Usage
Financial	Investment, credit card, bank, insurance	All use databases to run the business, banks to handle ATM data, and investment to analyze stock information ("quants").
Telecommunications	Telephone companies and PTTs, wireless suppliers	All use databases to run the business, and phone companies use databases to deliver robust storage of voice and data switch information.
Retail	Department store chains, restaurants	All use databases for point-of-sale data (OLTP), decision support, and business operations.
Manufacturing	Car makers, such as General Motors and Ford; equipment makers, such as Caterpillar; semiconductor suppliers	All use databases for ERP applications (mixed); carmakers use databases for design (CAD/CAM).
Health care	Hospitals and hospital chains	Hospitals use databases to run the business and increasingly to store and protect key rich-media data, such as charts.
Oil/gas/aerospace	Energy suppliers, such as Exxon Mobil; aerospace manufacturers, such as Boeing	Oil/gas companies use databases to store and analyze geological data; aerospace companies use databases to store designs (CAD/CAM) and handle production (ERP).
Travel/entertainment	Hotels; luxury cruise lines; television, movie, and music companies	Hotel chains store movies-on-demand in common databases; increasingly, news organizations use content management to handle rich-media data; movie companies store video in large information bases.

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

Vertical Market	Example of Companies	Typical Database Usage
Transportation	Truck, train, and shipping companies	All use databases to store key scheduling data.
Government	NSA, Department of Defense	Databases store rich-media satellite data.
Education	Universities	Databases store student and administration data securely.

Source: Aberdeen Group, October 2003

The sharp downturn in the telecommunications vertical market has had a major impact on the database market in general. Likewise, cost cutting across vertical markets — such as financial and retail — have had a negative impact on most database suppliers. Recent increases in IT and health-care spending are beginning to boost database revenue in 2004.

The markets in which open source databases have had the most success are retail and telecommunications, followed closely by financial (but not insurance) and government. The health-care and manufacturing markets are relatively resistant to open source databases.

As noted above, open source databases have had relative difficulty in penetrating the markets for large-scale packaged applications and vertical applications, and an open source database will need major customization and careful attention to channel penetration in order to succeed — in other words, the odds are against this penetration occurring in the next year.

C. Competition for Open Source Databases

Dynamics of Open Source Market Competition

Interviews with MySQL, PostgreSQL, and Berkeley DB users show that the recent surge in interest in open source software has had its counterpart in the database market. Although some open source suppliers have fallen by the wayside as the Web frenzy subsided and they were forced to identify sources of revenue, it now appears clear that several open source databases — including MySQL, PostgreSQL, and Sleepycat Software's Berkeley DB — are here for the long haul. They have not yet made major inroads into established suppliers' installed bases at the high or low ends, but they clearly have a strong niche in the low-end space, running particularly on the Linux platforms that have superseded SCO's low-end Unix, and with significant Windows penetration. These databases are especially attractive to highly technical developers, and particularly in such Linux-friendly vertical markets as telecommunications.

The IT buyer does not expect these open source databases to be equal in breadth of features and robustness to offerings in any other database market; however, they have made significant strides over the last few years and can now offer surprising performance/scalability, manageability (that is, the technical specialist finds administration straightforward, but the average database administrator may not), and programmer support. In flexibility, their standards-based approach can often make up for their lack of specific features for integration with other parts of a software infrastructure such as enterprise application integration (EAI). As a result, IT buyers value these open source databases not only for their low license costs but also as the database for Web-friendly applications that must be created quickly in-house.

IT buyers assessing these open source databases look primarily at the typical criteria by which all databases should be judged — scalability; robustness and manageability; flexibility; and programmer productivity. Most open source database users are "self-selecting" — users who choose these databases do not have an immediate need for advanced features.

Over the next two years, the market will reach a "tipping point" at which a larger range of vertical application and line-of-business programmers will find open source databases' low cost and association with other open source software such as Linux a good reason to include open source databases in their plans. At that point, open source databases will begin to have a significant impact on the overall database market, on database pricing, and on the readiness of the market for an "enterprise-scale open source database."

D. Differences in Implementing and Maintaining Open Source Databases

Interviewees indicated that they value the wide availability of programming help for open source database programming in the open source community; the fast response by open source database suppliers to requests for bug fixes (whereas large database suppliers usually proffer several levels of less technical support, interviewees typically can talk directly to the engineer designing the open source database); and the ability to turn for help to the open source community when the vendor's support fails.

New users should therefore expect to set up communications with not only the supplier of the database but also a larger open source community. This community can aid in implementation and advice on problems as well as programming "on top of" the database. In other words, users should expect that the main difference in their experience from a proprietary database is dealing not with a large proprietary vendor service arm but with a more amorphous open source community and that "best practices" will take full advantage of that community.

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

Users should also note that open source licenses are different from proprietary ones. Users should understand the differences and then rejoice in the ease of maintenance of open source licenses, which do not require extensive administration.

E. Future Trends in Open Source Databases

Aberdeen expects open source databases' installed bases to continue to grow at a slower pace over the next two years, with more of their new market coming from Windows than previously. A small number of high-end installations will emerge, and both databases will begin to penetrate the manufacturing, oil/gas/aerospace, and insurance verticals, with health care taking longer owing to health-care professionals' strong aversion to perceived risk and slow buying cycles.

These databases' penetration into vertical applications will continue a steady increase, especially in those verticals that offer their customers a choice of embedded databases. However, this increase will be limited by the fact that in many cases these databases offer little cost, support, and development advantages over existing proprietary suppliers — in fact, the lower level programming tools associated with open source databases may give them a programming disadvantage in data-intensive application development. Enterprise applications will likely see little penetration by these databases in the next year, although the SAP DB acquisition may give MySQL a small penetration into that market in 2005.

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

Section III: Findings from Database User Interviews

A. Findings from Open Source Database User Interviews

The interviewees were a combination of users and ISVs. They all had considerable DBMS experience, with open source and also often with proprietary databases.

What is your environment and applications?

Most interviewees typically oversaw relatively small implementations. Most used Linux hardware and a three-tier Web architecture, with some using Unix and Windows 2000.

Some applications included querying photo images for the "eBay of amateur photography," querying current weather data for the FAA, providing a search capability for classified ads, and offering MySQL as an optional embedded database for a packaged application. Additional applications included a temp replacement application that matched jobs to people, a financial application, an account management system, PostgreSQL as the backend for an accounting/invoicing application, a POS application to process debit/credit cards, software to support content management and digital media distribution (Berkeley DB), and use as an embedded store for directories, calendaring, and similar applications (Sun/Berkeley DB). The most sophisticated application was a full-blown ERP suite with a dozen modules ranging from accounting to inventory to hard-core manufacturing and a graphical report writer.

Most applications involved a substantial transaction stream of reads, with relatively few — but a significant number of — complex queries and updates. One application involved a 60–40 mix of reads and updates. One interviewee reported up to 250 concurrent transactions.

What conclusions can you draw from your experiences?

Interviewees' comments ranged from "world's greatest DBMS" to "easy to install and configure." An ISV interviewee noted that the open source database was "very transparent" and "fast," requiring "very few support calls." One interviewee cited "an unbeatable combination of performance, robustness, flexibility, simplicity, and cost (Berkeley DB)."

How important are the following features — robustness, performance/scalability, support for programming/deployment/maintenance, and cost of administration/deployment/upgrade — and how well are you satisfied with your open source database with respect to the features?

All interviewees rated robustness as high in importance and said that they were satisfied with the robustness of their open source databases. Comments ranged

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

from "rock solid" to "good for the price." One user had 40 to 50 PostgreSQL systems running in the field and had had no database problems for about a year. Another noted that Berkeley DB had been in "heavy use" for two years, and the company had never lost data.

All interviewees rated performance/scalability as high in both importance and the open source database's ability to deliver these features. Comments ranged from "it scales with replication" to "faster than enterprise databases" to "able to handle 500-gigabyte data stores."

Most interviewees rated support for programming, deployment, and maintenance as medium in importance. Interviewees typically did not use many supplier support services (just the "book and Web site" in one case); therefore, they found this question irrelevant. An ISV interviewee noted that its customers do not deal with maintenance because an open source database does not require them to do so. Several interviewees noted the importance of direct, immediate contact with technically knowledgeable people, which one interviewee noted "have been a joy to work with."

All interviewees ranked cost of administration, deployment, and/or upgrade as three or above in importance, and they also rated the open source database's ability to deliver these features as high. Comments included "no major difficulties," "low effort," and "it works out of the box." An ISV interviewee stated that costs were "on a par with other databases," and another noted that Berkeley DB "doesn't require an RDBMS license or a DBA."

Overall, interviewees appeared to rate performance/scalability as the most important feature, with robustness second. It should be noted that (in response to another question) interviewees cited "zero license costs" as a key factor in buying and staying with their open source databases. An ISV interviewee noted that "lower costs" were a major factor in 50% of the vendor's new customers choosing MySQL rather than an alternative proprietary embedded database.

How does this compare with Oracle, DB2, or Microsoft?

All interviewees had looked at Oracle, two evaluated Microsoft, and one tried out DB2. All felt that proprietary databases' license costs were too high, were associated with too many patches, or had an organization too big for fast bug fixes.

Why did you choose open source? Do you have any reservations?

All interviewees immediately cited "price" (license cost) as the main reason. They chose open source databases over proprietary alternatives because they liked the control that open source provides with respect to maintenance and support, lower cost (in some cases, no full-time database administrator is required), availability of source code, and the flexibility to add new features (and make them available via open source) if needed. One noted that Berkeley DB could scale better than Ora-

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

cle for the company's application, and another said that "we like the reassurance of having the source to all the code."

Some ISVs indicated that they can make higher margins by not having to include a costly proprietary license as part of their package to SMBs, which are generally limited with respect to how much they can spend on software. In many cases, no full-time DBA is required.

All indicated that they had no reservations at all about open source databases in general and their own open source database in particular.

Do you feel you can grow your business on an open source database?

All interviewees said yes, without reservations. The time horizon appeared to be two to three years. One noted that his database was growing at the rate of 10% to 30% per month, and there was still "lots of room to throw hardware at the problem."

What would you like to see from your open source database that isn't there yet?

One interviewee cited full-text indexing, one cited triggers and a "geographic information system (GIS)," one cited rollback, and one cited a GUI "control center." Several interviewees indicated that they would like to see improved replication capabilities. A couple of interviewees requested clustering similar to Oracle9i RAC (Real Application Clusters). Two-phase commit is not available in PostgreSQL. Every PostgreSQL user requested this feature. One interviewee indicated that she would like to see an improved planner to recognize and handle badly formed queries and also have an incremental backup tool. One Berkeley DB user indicated a desire that it "come up the stack" and add more data-modeling capabilities.

All indicated that they were in no hurry for these and that they believed their supplier would provide them in time for their needs.

B. Findings from Proprietary Database User Interviews

The proprietary database interviewees included some using Microsoft SQL Server 2000 and some using Oracle9i.

All interviewees had relatively long (four years or more) experience with their present databases. Most were medium-sized implementations (40 GB to 200 GB of database storage).

What is your architecture?

All interviewees had numerous PCs, both as clients and (sometimes) as servers, with one to a few servers. Application servers were used rarely as load balancers and more frequently as platforms for developing second-tier code.

What are your applications? (Packaged/ISV? Mission critical?)

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

Most interviewees were running packaged applications over their databases, with at least 50% of these being mission critical for all interviewees. One interviewee was running J.D. Edwards; another, a custom order entry and billing system for the communications industry; and a third, a custom dental claims adjudication application.

What could be improved with your current database?

Microsoft users indicated that upgrades and patches, backup across tables, and the ability to create remote data storage without a physical attachment could be improved. No interviewees indicated that these improvements were vitally needed.

What has kept you from using an open source database?

Microsoft answers typically focused on perceived risks of switching to open source, costs of administration and development (because Microsoft developers and administrators are readily available and SQL Server 2000 is perceived as robust), and costs of switching.

Would you be interested in trying an open source database?

No interviewees indicated any interest.

C. Findings from Aberdeen TCO Studies

Findings from Aberdeen's TCO studies of SMB/workgroup users of proprietary databases are applicable to open source databases. Although Aberdeen's interviews did not target cost of ownership directly, it is possible to draw tentative conclusions about open source database cost of ownership relative to both SMB-targeted (e.g., Pervasive) and enterprise (e.g., Oracle) proprietary databases.

Aberdeen's interviewees indicated that license costs, administrative costs, and deployment costs of open source databases were significantly below those costs for major enterprise databases such as Oracle. However, administrative and deployment/upgrade costs appeared comparable with or greater than those of embedded database suppliers, such as Progress, Pervasive, and Sybase (iAnywhere Solutions). Development costs appeared significantly above those of both enterprise and embedded database suppliers.

Overall, open source five-year database cost of ownership in the SMB/workgroup space appears superior to enterprise databases, but comparable with or slightly higher than SMB/workgroup databases — although your mileage may vary.

Aberdeen Conclusions and Recommendations

Open source databases have improved dramatically since the early days, in scalability, robustness, and manageability — to the point where they deserve to be compared with enterprise databases, and even to seem superior in some SMB applications. They offer attractive rapid response to bugs, low to no license costs

260 Franklin Street
Boston, Massachusetts 02110-3112
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

and easy license management, and a large pool of open source developers on which to draw. At the same time, they remain less productive in creating applications that do not primarily involve Java and Web sites.

Open source databases are likely to prove most useful to IT organizations creating new applications that involve Java and/or the Web, as well as to vertical ISVs seeking a low-license-cost alternative to enterprise proprietary databases. Vertical markets that will find open source databases especially useful should be financial services and telecommunications, with retail and government close behind.

IT buyers and ISVs considering an open source database should begin to assess and prototype at least one in the near future. These users should plan for close relations with an open source community, take the time to understand the nature of open source licenses, and identify in-house resources with the technical skills in areas such as Linux or generic databases to handle development on and administration of an open source database.

Overall, the prospects for open source databases are moderately bright — limited by the market's strong attachments to existing databases, not by major technical deficiencies. Although they do not offer the bells and whistles of proprietary databases, they do offer a cost-effective alternative to these in many cases. Aberdeen's view is that the all-dressed-up open source databases will start filling their dance cards with departmental applications developed on a shoestring by enterprise IT staff and savvy vertical application developers.

Aberdeen Group, Inc.
260 Franklin Street
Boston, Massachusetts
02110-3112
USA

Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

Aberdeen Group is a computer and communications research and consulting organization closely monitoring enterprise-user needs, technological changes, and market developments.

Based on a comprehensive analytical framework, Aberdeen provides fresh insights into the future of computing and networking and the implications for users and the industry.

Aberdeen Group performs projects for a select group of domestic and international clients requiring strategic and tactical advice and hard answers on how to manage computer and communications technology. This document is the result of research performed by Aberdeen Group that was underwritten by Sleepycat Software. Aberdeen Group believes its findings are objective and represent the best analysis available at the time of publication.