# Opensource Databases: A Comparative Analysis

-Amit Kumar

# Overview

Open source databases come with the advantage of being free, available for understanding, modification, and improvement as per needs.

There are three ways of using an open source database - do-it-yourself, third-party, and outsourcing. The choice of an open source database depends on factors, such as the intended usage, duration of use, and the cost involved as the main factor. Often the best comparison is obtained by first-hand evaluation. However, the views of the user community regarding the most suitable open source database are worth observing. The points of comparison are players with license, popularity, features, and selling points.

This paper is an effort to provide a background to help you choose an open source database that suits your needs.

# Players

The top three vendors in the database market are: Oracle, IBM (DB2 and Informix), & Microsoft (SQL Server). In the open source market, there are six important players (on the basis of acceptance in the software community at large):

1. MySQL - GNU (General Public License) as well as commercial
2. PostgreSQL - BSD (Berkeley Software Distribution) license
3. Firebird - Initial Developer's Public License (IDPL)
4. Ingres - GNU General Public Licensev2
5. MaxDB (earlier known as SAP DB, it is being developed in alliance with MySQL especially for SAP environments) - GPL
6. HSQLDB - BSD license

# Popularity

The popularity of an open source database seems like a trivial concern to the technical community, but the fact stands tall that a product's success is based on other reasons in addition to the technical features. Some of these are:

a) Support available from community as well as tool vendors

b) Momentum of the product being released in and open source arena

These are the two most important business concerns for choosing an open source product. The former ensures integrity with various other technologies and the latter ensures that open source is not going to die or fade out easily.

The six open source databases that we selected, fare differently on the popularity aspect.

• MySQL - most popular open source database boasting more than 8 million active installations. Starting since 95 (internally), the MySQL community is the largest and has frequent releases; version 4.1 came out in Apr-03, current version 5.0 in Aug-04, and the much-talked and awaited future release version 5.1 is out as beta since Nov-05.

• PostgreSQL - has hundreds of companies listed as Users. The community is as old as Aug-96. It is much-adored among technical jig wigs for its huge support of features. **EnterpriseDB** is a product built over it. Last few releases have been version 7.4 in Nov-03, current version 8.0 in Jan-05, and the new release 8.1 in Nov-05.

• Firebird - about 0.1 million deployments as of Sep-05. This is a very old community since Sep-84 and has a huge fan following for the number of good and bad times it has been through. The versions out are version 1.0 in Mar-02, current v1.5 in Feb-04 and 2.0 alpha is out there.

• Ingres - about 5000 existing customers. After changing many hands, the product has been made available to open source since '04, though it has been in existence along with PostgreSQL. The current release - r3 is out since Aug-04 and r4 is much-awaited for its support to materialized views.

• MaxDB - about 6000 customer installations. This has been specific to SAP applications earlier known as SAPDB. Started since Oct-00. It was available in version 7.5 in Nov-03 and version 7.6 in Nov-05

• HSQLDB - 0.3 million downloads. The project is finding hard to get contributions and is slower than others. It is the youngest in the group, formed since 01. It was out with version 1.7 in Feb-05 and version 1.8 July-05 (It released 6 versions since Apr-01).

Platform Support also speaks about the popularity and acceptance. Regarding the OS, we have major players as Linux, UNIX, Apple, and Windows. Besides, we also have Novell with Netware and BSD.

- MySQL & PostgreSQL fare at the top of the group as they support all the listed platforms
- Firebird & Ingres lack support to Novell OS.
- MaxDB favors the market with Linux, UNIX, and Windows platforms only.
- HSQLDB is a java-based database, and hence, is more of JVM-dependent than platform-independent.

# Open Source Database Features

In general, most database features are supported by all databases with minor deviations. To name a few, we can categorize them as follows:

- SQL Compliance
- ACID compliance, Locking and concurrency support, Foreign Key constraint & Transaction support
- Functional features:
  - Views, schemas, sub-select, stored procedure, triggers, xml support
  - Tablespace feature
  - UNICODE support
  - API support
- Non-functional factors:
  - Performace & Availability
  - Stability & Flexibility
  - Web and Data-warehousing strengths
  - Security, Authorization, and SSL support
  - Replication, Load-balancing, and Clustering support
  - Ease of installation, configuration, management, and administration
  - Hot backups
  - Admin and migration tools/scripts
  - Easy to learn & use
  - Good documentation
  - Distribution License & Support

It would be a continuous task to figure out what features distinguish one database from others. To begin with, here are a few pointers

# Comparison of Selling Points

| | MySQL | PostgreSQL | EnterpriseDB | FireBird | Ingres | MaxDB | HSQLDB |
|---|---|---|---|---|---|---|---|
| Temporary Table | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Materialized Views | ✓[1] | ✓[2] | | | ✓[3] | | |
| Indexing | ✓[4] | ✓ | | | ✓[5] | | |
| Domain | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Cursor | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Partitioning | ✓[6] | ✓ | | | ✓ | | |

MySQL**Points in favor** -• This is the most widely accepted open source database in use. This simple reason is quite compelling to select among various options.

- It claims to be comparatively easier to learn and use.
- The popularity provides for a lot of options in Administrative tools.
- The architecture provides for plugging storage engines. Many options exist in the market to cater to various needs e.g. InnoDB.
- Commit grouping, gathering multiple transactions from multiple connections together to increase the number of commits per second.
- It is available on Novell Netware OS also besides the generic lot of Win, UNIX, and Mac family. **Points against it** -• Even the current version lacks many SQL RDBMS features. For instance, its handling of dates allows storing a date with a day beyond the last day of a month with less than 31 days, and arithmetic operations are vulnerable to either integer overflow or floating point truncation. Though these can be taken care of by running special SQL modes. There are many such features that have been criticized and are slowly being taken care of in newer releases. Version 5.0, for example, supports views, stored procedures and cursors, and version 5.1 will support triggers.
- Other criticisms include its divergence from the SQL standard on the subject of treatment of NULL values and default values.
- It does not support Roles.

1. Emulates Views using Stored Procedures and Triggers
2. Emulates Views using Stored Procedures and Triggers
3. Support from r3
4. Limited R tree and hash indexing support
5. Expression indexing support from r4
6. Partitioning supported from MySQL 5.1

We can conclude that even though MySQL lacks compliance with the SQL standards, it provides for comparable performance. It serves the purpose for users who are willing to accept the program's limitations (which decrease with every major revision) in exchange for speed, simplicity, and rapid development.

It is often included as a default component in low-end commercial web hosting plans. For application developers (mostly using PHP and Perl), MySQL is the only DBMS choice unless they want to operate their own web hosts.

Ingres

• The database is technically at par with big databases as Oracle. Besides having an open source advantage, business and other peripheral factors can only justify the use of Ingres.

MaxDB

**Points in favor** -

• The database is specifically made for SAP and performs best with it. It covers most of the shortcomings in terms of features over its sister MySQL.

**Points against it** -

• It is quite complicated to install & has poor documentation.

HSQLDB

**Points in favor** -

• It offers a fast, small (less than 100k in one version) database engine which offers both in-memory and disk-based tables. Embedded and server modes are available.

• It is best known for its small size, ability to execute completely in memory, and speed. It can also run on free Java runtimes such as, Kaffe.

• It is very fast for simple queries, e.g. HSQLDB takes less than 4 seconds to INSERT 50,000 rows, whilst MySQL takes over 20 mins!

**Points against it** -

• It is not a good choice if the DB size is large. It not only sucks system resources, but degrades DB startup performance as well.

• If you benchmark complex queries (joins with sub-selects, IN clauses, etc), you'll find that its a weak spot is its query optimizer.

• HSQLDB requires Java, which could be a sufficient constraint not to use it.PostgreSQL

**Points in favor** -

• PostgreSQL supports a richer SQL dialect and SQL sub-queries.

• PostgerSQL use a multiple row data storage strategy called MVCC to make it extremely responsive in high volume environments. The leading proprietary database vendor uses this technology for the same reasons.

**Points against it** -

• It is Considerably slower than MySQL

• It does not support the entire ANSI SQL 92' standard, much less the ANSI SQL 99' standard.

• PostgreSQL isn't hierarchical. Databases hold only tables, not other databases.

Firebird

**Points in favor** -

• Mature Windows support. It has been supported on Windows for a very long time and it is well-tested.

• Mature ADO.NET provider. Npqsql (PostgreSQL ADO.NET provider) is still in beta version. Firebird ADO.NET provider supports the embedded Firebird, services API (backup, restore, statistics, batch SQL execution, etc).

• Embedded version (with 2 MB runtime and easy switching to a standalone server) seems to be one of the biggest advantages of Firebird.

• It is reliable, stable, fast, and low-cost, and with a small footprint. And not so hard to learn!

**Points against it** -

- Stored procedures are difficult to implement in Firebird.
- Most RDBMSs stores column names lower case. Firebird stores them in upper case.
- Firebird OAT (oldest active transaction) & OIT (oldest interesting transaction) needs constant monitoring. If the OAT gets stuck, the database starts to get huge as transaction data is stored in the database file itself. Firebird does not have the concept of a separate transaction log.
- PostgreSQL supports temp tables and tons of built-in functions, which is missing from Firebird. Firebird also has 3 SQL dialects - SQL PSQL, DSQL, etc and you can't use one dialect from the other dialect.

ConclusionIt is a question of what you desire from a database that matters.For small and web-based retrieval applications, the in-built pre-configured MySQL in many application servers is served just right.The moment we need more from our databases, be it features or performance, or the lack of skilled resources, we start looking the other way.

We all know how user/developer-community in general is reluctant to dispense with a setup database if it serves the basic purpose. The reluctance can be explained from the fact that everything would come at a cost be it installation, configuration, timely upgrades (maintenance), administration, or even training the resources for the new database server.

Hence, for anything playful or level-I, it's MySQL. Anything serious and level-II, it is usually replaced by PostgreSQL or Ingres or Firebird. If the application is SAP-based - the default choice is MaxDB. More often than not we tend to fall on the proven and 'built-for' products. If the user is Java-savvy and has a love for fast but basic database, not to mention the fascination of such a low memory print, the usual option is HSQLDB. Some tools claim it to be 20 times faster, which is a lot!

As I mentioned, that there are positives as well as negatives for PostgreSQL, Ingres, and Firebird. All the three have similar features. It's really difficult to justify the best among the three. It's a particular feature in the wish list or inclination with (out) reasons (like the support community which is usually positive, but an opinion clash could always be there) - which impacts in the choice. All the three have their limitations and have also proved their worth as a part of successful enterprise solutions.

Amit Kumar is a graduate from the Bengal Engineering College, Shibpur and specializes in software product engineering with Java/J2EE.

Amit Kumar works for VAssure Inc, a premier software engineering vendor with marquee customers like Gemalto, Avolent, Mantas, and Nextone to name a few. He has deep understanding in areas related to open source technology and has a wealth of experience interacting with customers, and understanding their vision so that the same can be implemented at the Global Deliver Centers at VAssure