

SAP DB – The Open Source Database from SAP AG

SAP AG has transferred its SAP DB database system to an Open Source system. The SAP DB database system has been offered to users of mySAP.com solutions as a free-of-charge alternative to database systems from other vendors since 2000. In accordance with the Roadmap published in October 2000, the source code for the database was made available on schedule at the end of April 2001. Since then, SAP DB can be used by any organization or any person free-of-charge and without restrictions **under a GNU General Public License (GPL)**.

The Open Source initiative for SAP DB does not represent the end of SAP AG developing its own basis technologies. On the contrary: a team of almost 100 developers is working on the further development of SAP DB alone. This is primarily motivated by SAP's efforts to avoid, as far as possible, being dependent on the technologies of other vendors. SAP DB is offered for all mySAP.com solutions as a license cost free equivalent alternative to the systems of other database vendors. In addition, SAP DB is embedded in a number of special application solutions, including mySAP SCM APO (liveCache) and SAP Content Server.

A significant motivation for the Open Source initiative is the price structure for database licenses. This affects not only SAP, but every provider of database-based solutions that must procure database licenses in order to sell them on as part of their own solution, therefore imposing costs on the end licensee. SAP AG hopes to enliven the market through the Open Source initiative and to encourage users to critically examine the high-price policies of the market-leading database providers.

This document provides you with information about the Open Source initiative for SAP DB, the development of the database market from the mid-1990s to the present day, the technological development status of relational databases, and the product positioning of SAP DB.

The Open Source Initiative for SAP DB

Roadmap Implemented

With the publication in April 2001 of the source code of the SAP DB database management system (DBMS), SAP AG implemented its intention, announced in October 2000, to transfer its own database system to an Open Source system. The first milestone in the Open Source roadmap was to make available, in autumn 2000, the database kernel and the query and database management tools in the form of executable programs for Windows NT/Windows 2000, Linux (Intel) and common UNIX systems. These programs are available on the SAP DB homepage www.sapdb.org and can be downloaded by any person or organization free of charge. The publication of the source code shortly afterwards represented the next milestone. Since then, any person or organization can view and change the source code of SAP DB and the SAP DB tools.

Made Available under GNU General Public License

The SAP DB database kernel is made available under a GNU General Public License (GPL), and the clients and programming interfaces under a GNU Lesser General Public License (LGPL). The GPL license allows developers to modify the programs as they wish, and to distribute the modified programs, as long as this is done by distributing the source code. The most popular Open Source system, the operating system Linux, is also distributed under the GNU GPL.

Unlike other well-known Open Source projects, control of the future development of the official SAP DB programs remains with one party; the SAP DB development team at SAP AG.

Users of the SAP DB database system outside the SAP user community receive free online support by e-mail. If desired, these users can take out a Premium Support Contract that guarantees full access to the SAP AG service infrastructure with its well-known high quality.

Efforts for Independence Motivate Development of Own Database

The development of the database technology for SAP's own application systems is motivated by SAP AG's efforts to avoid, as far as possible, being dependent on the technologies and products of other vendors. SAP's own database platform therefore primarily provides risk minimization. It is, however, also wise from a commercial perspective. The costs to SAP AG that result from buying database licenses are many times the cost of an in-house DBMS development team.

Against the Oligopolization of the Database Market

The cost pressure is increased due to the increasing domination of the database market by a small number of large companies and the high-price policies that result from this. Rising costs for database licenses are driving total license costs for SAP application systems noticeably higher. As the costs of the database licenses are linked to the cost of each SAP application as a percentage of the complete price, and are not discounted, a reduction in price on an SAP application system directly reduces the profit margin. Therefore, high prices for database licenses have a negative influence on the business results of SAP AG.

The distribution of SAP DB free-of-charge is to set a counterpoint to the progressive oligopolization of the database market and the high-price policies of the leading providers. With this Open Source step, SAP confirms its intention to permanently provide its own database system free-of-charge.

Embedded Engine for Special mySAP Applications

SAP not only provides its own database system SAP DB as an equivalent alternative to the products of the established database vendors, but also uses it as an embedded engine for special application solutions. An example of this is liveCache, with which you can process large quantities of complex objects very quickly in the main memory. liveCache is a database instance type, that is, a special variant of SAP DB, and an integral part of mySAP Supply Chain Management (Advanced Planner and Optimizer).

Another example is the SAP DB Document Server, which forms the technological basis of SAP Content Server and the SAP applications based on it. The SAP DB Document Server is a database instance type that can be used for the management of all types of documents.

In both examples, SAP DB is used exclusively. New SAP applications that will run exclusively on SAP DB are currently being developed. SAP AG itself increasingly uses the SAP DB database system for internal database installations.

Open Source Instead of Marketing

Despite the increasing breadth of functions of the SAP DB database system, SAP AG does not intend to market the database system commercially, and to enter the DBMS market as a new competitor. Consequently, SAP DB is offered license-free in connection with SAP solutions. The users of mySAP.com applications must simply pay a support and maintenance fee that is determined as a fixed percentage of an accrued license price.

Instead of creating its own distribution channels for its database system, SAP AG decided to make it available free-of-charge through the Open Source initiative. Through this, SAP AG expects to primarily achieve very simple, Internet-based product provision and distribution,

including Internet-based support, and to significantly increase the size of the SAP DB user community. This should contribute to finding any possible errors or performance bottlenecks more quickly, in particular in SAP DB areas that are not used by mySAP.com applications. As SAP DB is positioned as a competitive DBMS that is suitable for all types of database-supported applications, the specification of SAP's own database development are not restricted to just the specific demands of the mySAP.com solutions.

Due to the high-level of technical complexity of a DBMS, it is less likely that a large number of programmers will make changes to the source code. Nevertheless, a wide user community can help to test the SAP DB database system in application scenarios more comprehensively than SAP's own test structures can.

The Linux community, which naturally has a high affinity to Open Source software, undoubtedly offers the best chance for the creation of a user community outside the SAP customer base. For this reason, SAP DB is part of the professional version of SuSE Linux (as of 7.2). SAP DB is also to be distributed with other versions of Linux in the future.

Competition with Other Open Source Databases

There are other Open Source database systems competing with SAP DB in the Linux environment. However, the SAP DB database system clearly distinguishes itself from the other database systems, such as PostgreSQL or MySQL, with regard to performance. SAP DB was designed from the start for use in OLTP applications with a large number of users and transactions. SAP DB is the only Open Source database that can adequately support such a complex and demanding SQL application as the SAP system and ensure round-the-clock operation when doing so. Due to the proven suitability for SAP applications, SAP DB is not only more powerful and more scaleable, but also significantly more robust.

Unlike the providers of the Open Source databases listed above, SAP AG does not depend on recovering its development costs for SAP DB through service income outside the SAP market or through additional products for the DBMS.

The Development of the Database Market since the Mid-1990s

A market consolidation has taken place since the mid-1990s, initiated in 1994 with the takeover of ASK/Ingres by Computer Associates, and reaching its highpoint so far in the middle of 2001 with the takeover of Informix by IBM. As Sybase has obviously pulled out of the database business, three big providers dominate the market: Oracle, IBM, and Microsoft. According to an investigation by the Gartner Group, all database vendors together realized 8.8 billion US Dollars worldwide in 2000 and therefore ten percent more than the previous year. Oracle made around a third of this turnover, IBM made just under a third, and Microsoft had a share of around 15 percent. Sybase and Informix had around three percent each. This means that the big three providers now share more than 80 percent of the market.

Market Domination by a Few Large Companies Prevents Price Reduction

This oligopoly is one reason that, despite many predictions to the contrary, relational databases have not become a commodity product and there is no sign of an imminent fall in prices. Microsoft has succeeded in delivering a competitive product with SQL Server and has taken the market lead for Microsoft Windows platforms with it. However, as this share is restricted exclusively to Microsoft Windows, it has hardly any influence on the price policies of the market leader Oracle (see Computerwoche No. 16/2001: "Oracle lässt sich seine Datenbanken vergolden" ["Oracle's databases turn to gold"]).

DBMS Do Not Become Part of Operating Systems

The prediction that DBMS would become part of operating systems, often made at the beginning of the 1990s, has proved to be false. On the contrary, the two database products that were developed and marketed by system vendors, Rdb from DEC and Nonstop SQL from Tandem, have become obsolete through being taken over by other vendors. Although the software is still supported, the products no longer appear in the market. Two of the three remaining large database providers are also operating system vendors – however, there is no indication that they have plans to integrate these database systems into their respective operating systems.

Status of DBMS Technological Development

Relational database management systems (RDBMS) are in the late phase of the “Technology Adoption Lifecycle” (as in Geoffrey A. Moore: “Crossing the Chasm” and “Inside the Tornado”). At the beginning of the 1990s, Michael Stonebraker, then still a Professor at Berkeley, was publicly asking the question, of whether RDBMS were a “done deal”. In fact, there had only been a small amount of progress made in the last ten years – with the exception of attempts to extend the relation model with object-oriented characteristics – with regard to the functional performance characteristics. The dominant vendors’ main interest lay far more in improving the performance, availability, scalability and handling of their database systems. The large providers therefore employed whole panels of specialists at times, in order to be able to continually present new Benchmark records. Nowadays, the interest in benchmarks is significantly reduced. In the meantime, it has become accepted that all database systems relevant to the market have achieved a performance level and a standard of robustness that make them appropriate database systems for business application solutions of (almost) any size. Benchmark tests with databases are nowadays mainly performed by hardware vendors and generally serve to demonstrate the performance capabilities and scalability of new server systems.

The Vision of an Object-Relational Database

The functional development of the RDBMS was driven by various visions during the 1990s. In general it had already been decided at the beginning of the decade not to pursue the concept of distributed databases, of which there were originally high expectations, due to the too close linking of database instances connected with it and the lack of application requirements. In the middle of the decade, there was a fear that object-oriented database management systems (OODBMS) might make a triumphal march. Even though the representatives of OODBMS only had a very modest market share, they did put their finger on a supposed weakness of the relational database model that, in concept, is designed exclusively for processing table-type structured, that is, relational information. This type of logical data organization is as unsuitable for processing complex structured or multimedia data as it is well suited to typical business applications. To avoid leaving the field to the representatives of OODBMS without a fight, the concept of object-relational databases was developed. Object-relational database systems were to remove the restrictions of the classic relational model by extending it through object-oriented concepts.

Data Blades Extend the Database Kernel

Informix was the leader in this development. Through the take over in 1996 of the startup company Illustra for 400 million dollars, they bought in the “Data Blade Technology” together

with the services of its originator, the database developer from Postgres, Michael Stonebraker. The Data Blades concept envisioned the extension of the database kernel with additional data types and functions, so that complex data structures could also be stored in the database system, and their data type could correspondingly be used. The idea was quickly picked up by the other big database providers. In this way, IBM extended their DB2 database systems with Extenders. Extenders are preconfigured packages for image and text analysis, the storage of audio and video sequences, and time series analysis. IBM then renamed the DB2 database system to DB2 Universal Database. Finally, Oracle introduced Data Cartridges with Oracle version 8. Object concepts and SQL were interwoven, therefore jeopardizing the robustness of the DBMS, as a programming error in a linked Data Blade, Extender or Data Cartridge endangers the integrity of the database.

Not Accepted by the Market

A few years later, it should be noted that the object-relational database concept did not prevail, and there are practically no applications of Data Blades, Extenders or Data Cartridges. An exception to this is a Data Blade that, together with associated functions, implements a data type for geographical coordinates. The term "object-relational" disappeared at the end of the 1990s, as it was no longer suitable for making a separation between competitors in marketing terms.

OODBMS in a Niche

Object-oriented database management systems could not – despite the predictions of many analysts – establish themselves as a legitimate successor to the RDBMS. They can nowadays only be found in a few areas - such as in the area of CAD/CAE applications or as persistent object stores for Web Application Servers. Relational DBMS will therefore be the predominant database technology for most business applications for at least another decade.

Web Application Servers Displace the Vision of the Universal Database

There is another reason for the fact that RDBMS are mainly reduced to their original core functionality nowadays: modern application systems are generally implemented with a three-tier architecture. This includes clients (usually Web clients), Application Servers and a database server. Although the first version of the SAP system involved no Web clients, it was one of the first systems that were based on this three-tier architecture. The introduction of the application server layer had the consequence that the database system was restricted to its most important tasks. Many tasks that had to be taken over by the database server in a two-tier client-server architecture can be better dealt with by the application server. If, in addition, there is a requirement for application portability over different database systems, you cannot use non-standardized and therefore proprietary functions.

It is the modern Web Application Servers that have contributed to the displacement of the vision of the universal database. The persistent storage of business data will in future be hidden behind Web Services or Data Sources with Internet connectivity, which are integrated and presented uniformly in an application by a Web Application Server. Therefore, the future belongs to the slender SQL database systems, rather than to those overloaded with functions. Functions that are only required in special cases are ballast for normal operation, use unnecessary resources and make the DBMS more complex and therefore more vulnerable.

The Positioning of SAP DB

What appeared to be a weakness at the end of the 1990s – restraint in the case of topics such as the object-relational concept and deliberately forgoing the battle for yet more functions – prove today to be strengths of SAP DB. The SAP DB database system delivers high performance, can be automated to a large degree, is versatile with respect to the types of information that can be processed, meets the current Standards, and provides excellent value in its total operating costs.

Modern Architecture for High Performance and Availability

Due to its modern architecture, SAP DB provides a high level of performance, scalability and robustness. In this way, the database can fulfill the performance demands of application environments with thousands of concurrent active users and very large data volumes. At the center, the multithread/multi-server architecture ensures a high degree of scalability with sparing handling of server resources. SAP DB fits flexibly into modern server architectures such as multi-processor systems or cluster configurations, and uses the advantages – for example, where high availability is concerned – without costly configuration. Due to the customizable architecture, SAP DB is suitable as a central database system both for three-tier and two-tier client-server environments.

Effective locking mechanisms, efficient caching of data, intelligent optimization of SQL applications, extensive parallel processing of read and write processes, and strategies to minimize the required write operations are among the architecture characteristics that significantly affect response times and throughput.

SAP DB is designed for interruption-free round-the-clock operation. Required maintenance tasks such as configuration customization, the extension of data or log areas, data backup, the creation of table indexes, and so on, can be performed during production operation without affecting the active users.

High Level of Automation for Unattended Operation

As well as performance and robustness of the DBMS, the main focus of the further development of SAP DB is on simpler operation of the database system. The vision of a database system that automatically manages itself to a large degree, and only requires minimal monitoring by the database administrator, is the guideline of the development.

Both the setting up of the database system and the running operation are largely automated. During configuration, SAP DB automatically sets the core parameters in accordance with the existing system environment. During the definition of database objects such as tables and

indexes, the database administrator works exclusively at the level of logical schemas: SAP DB automatically makes the assignment to the physical data structures in mass storage. The database system also handles the growth of tables and indexes completely dynamically. Prior allocation of disk space and regular reorganization runs, that with other databases are tasks that the database administrator must perform, are not required. Due to the dynamic disk space management, interfering interruptions of production operation are not necessary. The database administrator can concentrate on tasks such as user administration or assigning authorizations, and does not need to constantly monitor system resources.

SAP DB is delivered with a package of convenient tools that support the database administrator in performing the few necessary tasks. The most important SAP DB tools are available with a Windows and a Browser interface. In mySAP.com environments, you can additionally incorporate the monitoring of the database system into the monitoring performed with SAP CCMS.

Access to Many Information Types

While SAP DB was first optimized as a database system for the SAP system, the application areas now reach far beyond online transaction processing. The data store is no longer restricted to table-type structured information, but now includes both structured and unstructured documents. In this way, SAP DB meets the requirements of the ever more important Internet applications, which mostly use a combination of unstructured data such as HTML pages, images, audio and videos, and increasingly also exchange information with the aid of partly structured documents in the XML format.

Examples of this type of application are the previously mentioned SAP Content Server and mySAP Product Lifecycle Management. An extension to SAP DB conforming to the Web Distributed Authoring and Versioning (WebDAV) standard for the Internet-based storage of documents is in development.

Conforms to Standards

SAP DB fulfills the official ISO SQL Standard (ISO-SQL 92, Entry Level), meaning that SQL programs that were created for other database systems can easily be ported to SAP DB. SAP DB offers developers all of the interfaces that could be expected from an open DBMS. For program development, there is a precompiler for C/C++ and interfaces for the Perl, Python, and PHP script languages. Database developers that have previously worked with other systems require practically no (re)training to be able to immediately work with SAP DB.

The standard-compliant ODBC driver allows the use of Windows-based Office or reporting tools. With the help of the JDBC driver, you can easily connect the SAP DB database system to Java-based application servers.

Broad Platform Availability

SAP DB is available with the same range of functions for a large number of operating system platforms:

- Compaq Tru64 Unix (Alpha)
- IBM AIX (PowerPC)
- SUN Solaris (SPARC)
- HP-UX (HP-PA)
- Linux (Intel)
- Siemens Reliant Unix (MIPS)
- Windows NT/2000 (Intel)

Total Cost of Ownership Represents Incomparably Good Value

The total cost of ownership (TCO) for SAP DB is such good value partly because there are no license costs. The total running costs are, however, not determined just by license and maintenance costs, but also strongly influenced by employing and training of operators.

Therefore, characteristics such as simple handling and operation without constant monitoring contribute to extremely low total operating costs.