

Open Source Database landschap heel gevarieerd

# Klaar voor het serieuze werk

Jos van Dongen

**Roep 'Open Source Database' en in 9 van de 10 van de gevallen wordt spontaan 'MySQL' geantwoord. Dat de marketing van het Zweedse bedrijf dik in orde is staat buiten kijf, maar het Open Source Database (OSDB) landschap is veel gevarieerder dan de meeste mensen op het eerste gezicht denken.**

Waar voorheen voornamelijk een OSDB als simpele back-end database voor interactieve websites kon worden aangetroffen, is er inmiddels een rijk scala aan verschillende databases beschikbaar. Het centrale open source verzamelpunt, de project- en download-site [www.sourceforge.com](http://www.sourceforge.com), vermeldt binnen de categorie Database Engines/Servers maar liefst 852 projecten! Omdat het wat te ver voert om alles wat enigszins met een database te maken heeft binnen de OS gemeenschap hier te bespreken, zijn er enkele criteria aan de dag gelegd om het kaf van het koren te scheiden. Zo moet er de laatste 12 maanden wel ontwikkelactiviteit hebben plaatsgevonden, moeten er tools en een of andere vorm van ondersteuning beschikbaar zijn en liefst ook een JDBC of ODBC driver. Zelfs met deze criteria in het achterhoofd hebben we niet de pretentie een volledig overzicht te bieden. Waar het hier om gaat is inzichtelijk maken hoe de OSDB 'markt' er uit ziet en wat voor consequenties dat heeft voor de 'closed source' wereld. Een overzicht van de hier besproken databases is te zien in de tabel in afbeelding 1.

### Soorten en maten

Wanneer we naar de bekendste OSDB's kijken, MySQL en PostgreSQL, dan lijkt het erop dat het alleen zou gaan om relationele database management systemen. Niets is echter minder waar. Vanuit het objectgeoriënteerde Java-paradigma zal het zal geen verbazing wekken dat er diverse OS Object Oriënted (OO) database-projecten gevonden kunnen worden. Er bestaan inmiddels ook pure XML databases en een aantal andere OSDB's bevat extensies om met XML-structuren overweg te kunnen.

Jos van Dongen spreekt tijdens Database Systems op 20 maart 2007 over Open Source Databases.

Het in een eerder artikel besproken Mondrian is een typische ROLAP database server en Palo is een krachtige MOLAP database. Kortom, voor elk wat wils.

Wanneer we gedetailleerder kijken naar de achterliggende filosofie en de ontstaansgeschiedenis van de meeste OSDB projecten, wordt ook duidelijk waarom. Een OS project wordt namelijk nooit gestart vanuit het idee om bijvoorbeeld een nieuwe database te ontwikkelen, maar op basis van een concrete case. Vrijwel altijd gaat het om enkele ontwikkelaars die voor een specifieke toepassing een opslagmechanisme zoeken, en als ze niet (helemaal) vinden wat ze zoeken, gaan ze zelf aan de slag. In veel gevallen ontstaat dan toch een product met op zijn minst relationele trekjes en een SQL interface. De meeste aandacht zal daarom in dit artikel toch uitgaan naar de relationele databases. Enkele jaren geleden zag deze wereld er voor de niet-ingewijden nog tamelijk overzichtelijk uit: commerciële producten waren duur, maar rijk aan functionaliteit en boden professionele ondersteuning. Open source producten waren gratis, beperkt in hun mogelijkheden en voor support was men afhankelijk van een legertje goedwillende hobbyisten. Voor zover deze vooroordelen al ooit geklopt hebben is er inmiddels wel het een en ander veranderd. Zoals recentelijk in een onderzoek door Forrester Research is aangetoond, staan OSDB's voor het normale OLTP-werk hun mannetje, en dat tegen aanzienlijk lagere kosten dan hun commerciële tegenhangers. De onderzochte producten waren MySQL, PostgreSQL, Ingres, Derby, Firebird en BerkeleyDB. Ook in andere onderzoeken en overzichten komen deze namen telkens weer terug, waarbij Derby en BerkeleyDB soms wel, en soms niet voorkomen. Helemaal toevallig is dat niet: bij de 'grote vier' gaat het om complete RDBMS's die direct concurreren met 'grote drie' Oracle, DB2 en SQL Server. Voor veel applicaties zijn deze producten nogal zwaar, vandaar dat ontwikkelaars vaker een beroep zullen doen op *embedded databases* zoals Derby, SQLite,

HSQldb en BerkeleyDB. Als we naar de Java-ontwikkelgemeenschap kijken zien we gelukkig ook enkele pure object-oriented databases. DB4O en Neodatis bieden een middel voor object-persistentie en voorkomen hiermee de objectrelationele mapping die nodig is bij het gebruik van een RDBMS vanuit een object-georiënteerde taal als Java.

## Een stukje geschiedenis

Zoals veel OS ontwikkelingen is ook het ontstaan van de meeste OSDB's terug te voeren op de universitaire wereld. Een belangrijke naam die telkens weer terugkomt is Michael Stonebraker (zie ook DB/M 8, 2006 voor een interview met hem), inmiddels emeritus professor in Berkeley aan de gelijknamige universiteit en een grootheid in de database-wereld. Stonebraker stond aan de wieg van Ingres, één van de eerste relationele DBMS'en en via diverse omzwervingen door de commerciële wereld sinds twee jaar aan een nieuw leven begonnen in het open source domein. Vervolgens is hij medio jaren tachtig gestart met een nieuw project met als doel een 'postrelationele' database ontwikkelen. Hieruit is uiteindelijk PostgreSQL voortgekomen, inmiddels

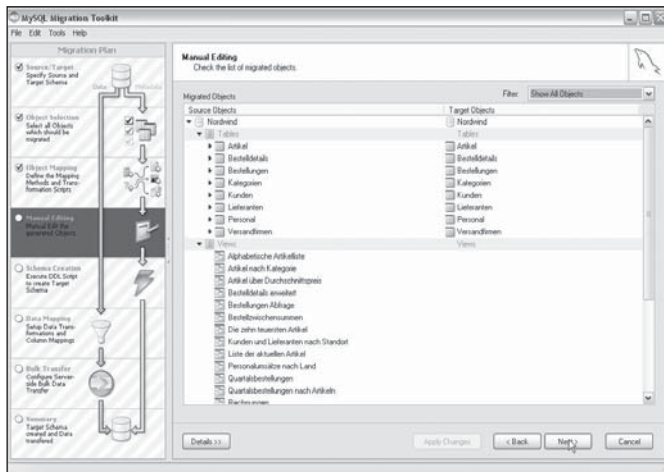
voorzien van een SQL interface en een uitgebreide feature set. Interessant om te vermelden is verder dat het oorspronkelijke Postgres in 1994 als Illustra een commercieel leven is gestart en later opgekocht is door Informix om onderdeel uit te gaan maken van de Informix Universal Server. Stonebraker was ook de PhD begeleider van Margo Seltzer, die later samen met haar echtgenoot het bedrijf SleepyCat startte en BerkeleyDB ontwikkelde.

## Ook grote research-bureaus hebben inmiddels de open source wereld ontdekt

BerkeleyDB is inmiddels één van de meest gebruikte embedded database systems en in februari 2006 overgenomen door Oracle. Dit soort vertakkingen, afsplitsingen, vercommercialisering en later weer teruggave aan het OS domein, is illustratief voor de ontwikkelingen in de OS wereld en zorgt voor een weliswaar

| Naam         | Type  | URL                                       | Bijzonderheden  |
|--------------|-------|---|---|
| MySQL        | RDBMS | www.mysql.com                             | Marktleider in OS DB's  |
| MaxDB        | RDBMS | www.mysql.com                             | Gestart als ADABAS D, door SAP als SAPDB doorontwikkeld en daarna geschonken aan de OS gemeenschap. Wordt beheerd door MySQL ab                   |
| Ingres       | RDBMS | www.ingres.com                            | Eén van de eerste RDBMS'en. Na zelfstandig bestaan overgenomen door CA, in 2004 aan de OS gemeenschap geschonken. Sindsdien succesvoller dan ooit |
| PostgreSQL   | RDBMS | www.postgresql.com                        | Object-relationele database   |
| EnterpriseDB | RDBMS | www.enterprisedb.com                      | Oracle compatible versie van PostgreSQL   |
| BizGres      | RDBMS | www.bizgres.org                           | BI/DWH versie van PostgreSQL; bitmapped indexing  |
| Greenplum DB | RDBMS | www.greenplum.com                         | Grote broer van Bizgres, bedoeld voor multi-terabyte DWH's  |
| ExtenDB      | RDBMS | www.extendb.com                           | PostgreSQL gebaseerde parallele database server voor DWH's  |
| DB4O         | OO    | www.db4o.com                              | Volledige Object Oriënted OS database; native object storage  |
| Firebird     | RDBMS | www.firebirdsql.org                       | Open Source doorontwikkeling van Borland Interbase  |
| Palo         | MOLAP | www.palo.net                              | In-memory OLAP database   |
| Mondrian     | ROLAP | www.pentaho.com                           | Onderdeel van Pentaho project   |
| SQLite       | RDBMS | www.sqlite.org                            | Zeer snelle in C/C++ based embedded database  |
| HSQldb       | RDBMS | www.hsqldb.org                            | Embedded in memory Java database; meegeleverd met OpenOffice 2.0; native XML storage  |
| Derby        | RDBMS | db.apache.org/derby                       | Embedded Java database van de Apache foundation   |
| BerkelyDB    | RDBMS | www.oracle.com/data-base/berkeley-db.html | Embedded DBMS, inmiddels van Oracle; XML add-on beschikbaar   |
| MonetDB      |       | www.monetdb.com                           | Vernieuwende Nederlandse implementatie van DBMS met XML extensies   |
| eXist        | XML   | www.exist-db.org                          | Native XML database   |
| NeoDatis     | OO    | www.neodatis.org                          | Native Java OO db & OO based kennis management systeem  |

**Afbeelding 1:** Overzicht besproken producten.



Afbeelding 2: MySQL migratie-toolkit.

levendig, maar niet altijd even overzichtelijk landschap van verschillende producten en projecten.

MySQL bijvoorbeeld is deels gebaseerd op mSQL (voor 'Mini-SQL'), oorspronkelijk ontwikkeld door David Hughes aan de Bond Universiteit in Australië als tool om SQL statements 'on the fly' te vertalen in PostQUEL, zodat gebruik gemaakt kon worden van een Postgres database als back-end. In een later stadium is een eigen back-end ontwikkeld waarmee het huidige mSQL ontstond. De geïnteresseerde lezer kan de complete ontstaansgeschiedenis van mSQL en MySQL terugvinden op [www.unix.org.ua/oreilly/linux/sql/ch01\\_04.htm](http://www.unix.org.ua/oreilly/linux/sql/ch01_04.htm). Interessant feit is dat MySQL 1.0 pas in 1995 op 'de markt' is verschenen en dus ten opzichte van de meeste commerciële tegenhangers pas net komt kijken. Ingres tenslotte leidde een kwijnend bestaan onder de vlag van CA, maar is sinds de schenking van het product aan de OS gemeenschap in 2004 aan een tweede, zeer succesvolle, jeugd begonnen.

## Closed versus open source

Een aantal OS databases begon ooit het bestaan als commercieel product (Firebird, Ingres) maar is inmiddels veel succesvoller onder de OS vlag. Het is dus interessant te kijken waar precies de verschillen zitten wanneer een keuze gemaakt dient te worden tussen closed source en open source database-producten. De enorme interesse die er momenteel bestaat voor OSDB's, is namelijk niet alleen te verklaren vanuit het kostenaspect. Oracle, Microsoft en IBM leveren immers gratis Express versies van hun Enterprise-producten. Deze kennen weliswaar beperkingen in het te gebruiken aantal processoren, hoeveelheid geheugen en de grootte van de database, maar voldoen prima voor kleinere implementaties. Ook als er naar de feature set gekeken wordt valt op dat in de OS producten nog steeds veel geavanceerde onderdelen ontbreken. Er moeten dus andere aspecten zijn die het succes bepalen.

Een groot deel van deze aspecten valt te groeperen onder de noemer 'vrijheid'. Vrijheid in tool-keuze, in keuze van support, in aantallen gebruikers, de vrijheid om desnoods code aan te passen en de vrijheid van processor-, memory- en data-restricties. Het

belangrijkste aspect lijkt echter de ondersteuning. Naast het leveren van geoptimaliseerde binary's (de gecompileerde installers voor diverse platformen) is de enige manier om geld te verdienen met een OS product het bieden van betaalde service, training en support. Dit betekent dan ook dat men er alles aan zal doen om een perfect product te leveren en de klanten tevreden te houden. Het gaat dan niet meer (alleen) om een kostenvoordeel, maar vooral om een ondersteuningsvoordeel.

Als we bijvoorbeeld kijken naar de support-opties voor MySQL dan zien we in de meest uitgebreide variant een ongelimiteerd aantal incidenten, een groot aantal advies- en review-diensten en een 7\*24 beschikbare telefonische helpdesk met 1 uur responstijd, en zelfs 30 minuten bij calamiteiten. Per server per jaar wordt hiervoor € 4.995,- in rekening gebracht. Let wel, dit is inclusief de licentiekosten! U mag zelf uitrekenen wat vergelijkbare licentie- en support-opties van uw favoriete closed source database op een 8-processor machine gaat kosten. MySQL beschikt dan ook over een uitgebreide migration toolkit om de overstapdrempel zo laag mogelijk te maken, zie afbeelding 2.

EnterpriseDB doet dit weer op een andere manier. Deze op PostgreSQL gebaseerde OSDB claimt volledig Oracle compatible te zijn ('runs most applications written for Oracle unchanged') waardoor bestaande Oracle applicaties een-op-een over te zetten zijn op een EnterpriseDB database (en vice versa). Een ander niet te onderschatten voordeel van OS producten is de actieve community die, georganiseerd in forums, vaak sneller problemen signaleert en oplost dan de leverancier. Betalende klanten, die vaak nog op oudere versies draaien, plukken hier weer direct de vruchten van omdat het ontwikkelteam fouten op deze manier al uit een nieuwe versie heeft gehaald voordat er een upgrade plaatsvindt.

Ook de grote research-bureaus hebben inmiddels de open source wereld ontdekt. Beperken IDC en Gartner zich in hun rapporten nog voornamelijk tot open source software in algemene zin en Linux in het bijzonder, Forrester Research heeft zich al gewaagd aan het eerder aangehaalde specifieke OSDB onderzoek.

De meest interessante uitspraak van het rapport is dat 80 procent van de applicaties slechts 30 procent van de functionaliteit van de top DBMS'en gebruikt, en dat de meeste applicaties voldoende hebben aan de basis database management functionaliteit. Omdat vrijwel alle OSDB's deze functionaliteit bieden leveren ze dus in de meeste gevallen een passende oplossing.

## OSDB's voor Business Intelligence

Dat OSDB's op grote schaal voor OLTP toepassingen worden ingezet en met het aantal implementaties veruit marktleider zijn bij interactieve website-ontwikkeling is evident. In het BI domein ziet het plaatje er vooralsnog anders uit. Zeker als we kijken naar grote datawarehouse-implementaties moeten de OSDB's vooralsnog met een lampje gezocht worden, hoewel er al wel enkele indrukwekkende cases beschikbaar zijn. De reden hiervoor is dat aan datawarehouses andere eisen gesteld worden dan aan

OLTP databases, en dat ze ook andere eigenschappen vereisen. De belangrijkste zijn tabelpartitionering, bitmapped indexen en materialized views. Op dit moment is er nog geen enkel OS product dat al deze features herbergt. Er zijn echter wel ontwikkelingen gaande. PostgreSQL ('the world's most advanced open source database') beschikt over een rudimentaire vorm van partitionering, en Bizgres (een speciaal voor BI doeleinden afgeleide OSDB op basis van PostgreSQL) voegt daar bitmapped indexen aan toe. Met behulp van triggers kan tevens een soort van materialized views in elkaar worden geknutseld. Deze worden echter niet automatisch herkend bij het parsen van de query, zodat deze informatie nog steeds in de rapportage/analyselaag moet worden opgenomen. MySQL gaat in de komende versie 5.1 met partitioneren een flinke stap maken en gaat qua mogelijkheden aardig richting Oracle en DB2. Alleen jammer dat (tot nu toe) het 'ALTER TABLE TRUNCATE PARTITION' commando ontbreekt, dat zeer handig is bij het bijwerken van een datawarehouse. Bitmapped indexen zijn voor 2007 aangekondigd, maar over materialized views heerst nog volledige radiostilte.

## Aan datawarehouses worden andere eisen gesteld dan aan OLTP databases

Voor multi-terabyte datawarehouses claimt Greenplum met BizgresMPP een oplossing te hebben, ook op basis van PostgreSQL/Bizgres. Greenplum heeft het Bizgres-project geadopteerd en heeft samen met SUN een complete datawarehouse appliance ontwikkeld die voor minder dan 20.000 USD per bruikbare Terabyte zeer concurrerend geprijsd is. Ook ExtentDB ziet een markt in de grote datawarehouses en heeft met PostgreSQL als basis een interessante parallelle database server ontwikkeld. Deze is weliswaar tot vier nodes gratis maar is geen OS product is en hoort dus eigenlijk niet in dit overzicht thuis. Wanneer we verder kijken dan de pure relationele producten, zijn er eveneens enkele projecten die voor de BI wereld erg interessant zijn.

Neem bijvoorbeeld Palo, dat een in-memory OLAP database met write-back mogelijkheid biedt voor Excel. Binnen Palo kunnen tot 256 dimensies gedefinieerd worden, wat ruim voldoende is voor welk soort business-analyse dan ook. Data kunnen geïmporteerd worden door middel van flat files, ODBC query's of via Excel-invoer. Voor controllers en analisten die op zoek zijn naar een krachtig en gratis hulpmiddel voor (financiële) analyse, planning & budgetting, zonder hierbij het vertrouwde Excel te hoeven verlaten, is dit een onweerstaanbare aanvulling.

Tenslotte mag MonetDB niet onvermeld blijven. In DB/M is dit product al diverse malen besproken, recentelijk nog als onderdeel van het interview met Martin Kersten, de geestelijk vader van MonetDB. In DB/M 4 van juni 2005 heeft Rick van Rein een

uitvoerig artikel gepubliceerd waarin MonetDB letterlijk binnenste-buiten wordt gekeerd. Door de radicaal andere benadering van het DBMS maakt MonetDB optimaal gebruik van moderne hardware, waardoor het product in de meeste benchmarks gehakt maakt van de competitie.

## Conclusie

De ontwikkelingen op het gebied van OSDB's gaan razendsnel. Waar de meer traditionele DBMS'en nu nog tekortschieten in functionaliteit wordt, als we de roadmaps mogen geloven, het gat met de closed source concurrentie met rasse schreden gedicht. Veel tools voor database-modellering en -beheer kunnen al met het grootste deel van de OSDB's overweg, wat een eventuele migratie of naast elkaar gebruiken van verschillende producten eenvoudiger maakt. Dit, in combinatie met de concurrerende prijsstelling en geboden supportprogramma's, maakt dat OSDB's op geen enkele shortlist meer mogen ontbreken. Hoe het op dit moment gesteld is met de marktacceptatie en -penetratie hoop ik u op het komende Database Systems congres te kunnen vertellen. Er breken interessante tijden aan.

*Met dank aan Roland Bouman, van MySQL ab, voor zijn waardevolle aanvullingen en opmerkingen.*

### Jos van Dongen

Jos van Dongen (jvdongen@tholis.com) is Senior Consultant bij Tholis Consulting.

Online archief Database Magazine

Online archief

trefwoorden:  zoek Zoektip

U bent op dit moment niet ingelogd. [ inloggen ]

Extra zoekcriteria:

Database Magazine

Alle magazines

Zoek in:

Alle velden

Titel

Auteur

Omschrijving

Jaar

Bladnummer

Datum:

van:

tot:

Aantal artikelen per pagina:

[Array Publications ©](#) | [Disclaimer](#) | [Privacy statement](#)

Database Magazine-lezer opgelet! Artikelen over onderwerpen als Datawarehousing, SQL, ETL, Business Intelligence, Relationele databases, modellering en nog veel meer vindt u in het Online Archief van Array Publications. Vaktijdschriften als Storage Magazine, Database Magazine, IT Service Magazine, Java Magazine en ons Oracle vakblad Optimize hebben hun artikelenarchief online gezet. Met een Google-achtige zoekstructuur vindt u snel wat u zoekt op [www.dbm.nl](http://www.dbm.nl)