

ROI ziet er zeer aantrekkelijk uit

# Open Source database-systemen zijn realiteit

Geert Batterink en Jim Mlodgenski

**Nu IT-afdelingen in de huidige economische situatie kostenbesparingen moeten realiseren, gaat de adoptiesnelheid van Open Source software en Open Source Databases bij bedrijven snel omhoog. Hoewel het kostenaspect altijd een groot voordeel is, zijn er tevens andere factoren die een belangrijke rol spelen bij de acceptatie van Open Source database-systemen in grote organisaties.**

Vanwege de flexibiliteit van het Open Source model, kan deze technologie op steeds meer plaatsen ingezet worden waarmee het aantal toepassingsgebieden sterk toeneemt. Deze trend zal zich voortzetten en impact hebben op de technologie die bedrijven de komende jaren zullen gebruiken. Steeds meer bedrijven in Nederland en andere delen van Europa nemen op Open Source gebaseerde oplossingen in gebruik en deze trend beperkt zich niet tot test- en acceptatieomgevingen in bedrijven. Er zijn steeds meer voorbeelden van bedrijven die ook voor business kritische toepassingen in de productieomgevingen gebruik maken van Open Source database-oplossingen. Voor andere bedrijven is de inzet van Open Source Business Intelligence-oplossingen vaak een geschikte eerste stap om gebruik te maken van Open Source database-technologieën.

## Redenen voor keuze

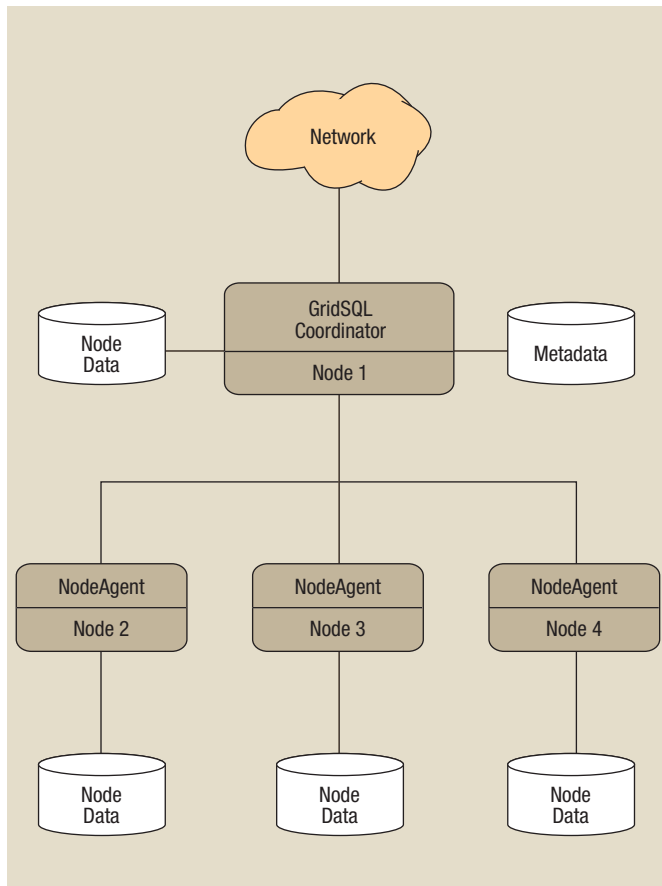
Tot voor kort waren er verschillende factoren die het grootschalige gebruik van Open Source database-systemen belemmerden. Het *nerdy* en 'zolderkamergerichte' beeld van Open Source software dat tot voor kort bij veel organisaties heerste, blijkt inmiddels hopeloos achterhaald. Open Source leveranciers zijn in veel gevallen net zo professioneel, zo niet professioneler, dan traditionele leveranciers van software. In veel gevallen hebben de medewerkers en het management van een commerciële Open Source leverancier veel ervaring opgedaan bij de traditionele closed source leveranciers van soortgelijke oplossingen. Open Source oplossingen hebben hun volwassenheid, schaalbaarheid en gebruiksvriendelijkheid inmiddels bewezen in de markt en vormen geen belemmering meer voor de invoering van Open Source database-systemen in grote organisaties. Verder zijn steeds meer getrainde en gecertificeerde IT-specialisten beschikbaar op de markt die gespecialiseerd zijn in Open Source oplossingen. Leveranciers realiseren zich dat

een overstap naar Open Source database-systemen voor klanten eenvoudig gemaakt moet worden en dat database administrators hun bestaande vaardigheden ook in de nieuwe oplossing moeten kunnen inzetten. Nu het overstappen eenvoudig kan en geen grote investeringen nodig zijn voor herscholing van database administrators, neemt de adoptie snel toe. Tot slot zorgt de beschikbaarheid van partijen die trainingen en formele support verzorgen ervoor dat ook de laatste drempel voor adoptie op grote schaal is weggenomen.

## Voordelen

Organisaties hebben op verschillende manieren baat bij het gebruik van Open Source database-systemen. In de meeste gevallen profiteren ze van een sterke reductie in licentiekosten. Voor Open Source database-systemen hoeft veelal niet geïnvesteerd te worden in licenties, en volstaat een jaarlijks supportcontract, waarmee een piekinvestering voorkomt. Daarnaast kan op deze manier ook de zogenaamde *vendor lock-in* worden voorkomen, omdat er meerdere supportpartijen beschikbaar zijn. Tot slot zien we dat steeds meer bedrijven een voorkeur voor Open Source software in het algemeen opnemen in hun enterprise-architectuurbeleid en dat het gebruik van een Open Source oplossing actief wordt gestimuleerd. Het gebruik van Open Source database-systemen sluit hier uiteraard goed bij aan.

Naast bovenstaande voordelen zijn er enkele wellicht minder in het oog springende zaken die voor het gebruik van Open Source software pleiten, zoals flexibiliteit, security en auditability. Een database-systeem kan op een flexibele wijze worden aangepast aan de architectuur van een oplossing (door de broncode te wijzigen), waar in andere gevallen vaak de applicatie aan de kenmerken en beperkingen van het database-systeem aangepast moet worden. Een voorbeeld hiervan is het IP-adres



**Afbeelding 1:** Schematische GridSQL architectuur.

datatype dat gebruikt wordt in de communicatie-industrie. De beveiliging van een oplossing en de claims daarover door de leveranciers, kunnen in een Open Source oplossing eenvoudig worden gecontroleerd, net als de kwaliteit van de code. In traditionele oplossingen kan het voorkomen dat de leverancier claimt dat een bepaalde beveiligingskwestie is opgelost, terwijl deze in feite alleen is gemaskeerd. Bij een Open Source oplossing kan in de broncode gekeken worden of dit probleem inderdaad correct en compleet is verholpen.

## Twee volwassen systemen

De meest gebruikte en technisch meest volwassen Open Source database-systemen zijn MySQL en PostgreSQL. Beide platformen hebben hun eigen technische karakteristieken en voordelen voor het gebruik in verschillende business situaties.

MySQL heeft veel technische voordelen voor het gebruik in de web-omgeving van een organisatie. Met een lage overhead voor eenvoudige query's, is de performance van MySQL veelal beter dan van commerciële database-platformen. Bovendien zorgt de flexibiliteit van verschillende back-end storage engines voor extra technische voordelen. Dit concept van meerdere storage engines zorgt ervoor dat een DBA kan bepalen welk mechanisme door een tabel wordt gebruikt om gegevens van en naar de fysieke disk te lezen en schrijven.

Indien een database-tabel bijvoorbeeld gegevens bevat die

vooral worden gelezen en bijna nooit wijzigen, dan is de MyISAM storage engine een goede keus. Als de gegevens in de tabel echter data bevat die vaak wijzigen, is een transactionele storage engine zoals InnoDB een betere keuze.

PostgreSQL is een database voor algemeen gebruik in organisaties met technische voordelen in een groot aantal toepassingsgebieden. Het meest in het oog springend is de flexibiliteit die standaard in de PostgreSQL architectuur is opgenomen.

Allereerst kunnen ontwikkelaars dankzij server-side taalondersteuning gebruik maken van meer dan tien talen, inclusief Java, Perl, Ruby en PHP, om eigen functies en triggers te implementeren. Hiermee kunnen de specifieke voordelen en sterke kanten van een programmeertaal worden benut om een oplossing te implementeren die naadloos aansluit bij het op te lossen business vraagstuk. Verder kunnen business requirements goed ingevuld worden door gebruik te maken van specifieke datatypes die via extensies beschikbaar zijn in PostgreSQL. Een voorbeeld hiervan is een netwerk-monitoring applicatie, waarbij het native IP-adres formaat slechts in beperkte mate schijfruimte gebruikt, terwijl geavanceerde zoektechnieken, zoals filtering op bepaalde subnetten, heel goed mogelijk zijn.

In veel gevallen groeien de datavolumes en behoeftes erg sterk, waardoor het nodig is om horizontaal uit te schalen in de database-laag en niet langer een enkelvoudige database-instantie te gebruiken. Zowel MySQL en PostgreSQL bieden hiervoor technisch zeer gevanceerde oplossingen.

Met MySQL clustering is een mechanisme beschikbaar waarmee data horizontaal over het geheugen van verschillende servers in een *high-availability* omgeving worden verdeeld. Hierdoor kan MySQL goed ingezet worden in read-intensieve situaties, zoals veel gebruikte websites. PostgreSQL levert horizontale schaalbaarheid in BI-oplossingen met GridSQL. Met GridSQL wordt een grote parallelle processing-omgeving gemaakt, waarmee intensieve query's door meerdere servers parallel kunnen worden afgehandeld. In BI-toepassingen kunnen hiermee aanzienlijke performance-verbeteringen worden bereikt.

## In de meeste gevallen profiteren organisaties van een sterke reductie in licentiekosten

GridSQL dankt zijn goede performance aan de architectuur waarbij op basis van commodity hardware de binnenkomende requests worden verdeeld over de beschikbare nodes in het cluster door de GridSQL Coördinator. Elke node in het cluster bevat een gedeelte van de hele dataset, die verdeeld is over alle nodes met behulp van een hash-algoritme. GridSQL maakt gebruik van een kostengebaseerd optimalisatiemechanisme,

waarbij kennis over de data en partitionering wordt gebruikt om maximale parallelle processing in te zetten bij het beantwoorden van query's op de database. Deze parallelle technologie zorgt ervoor dat met relatief goedkope commodity hardware, eenvoudige servers met meerdere CPU's, uitstekende performance wordt gerealiseerd.

### GridSQL maakt gebruik van een kostengebaseerd optimalisatiemechanisme

Hoewel Open Source database-systemen in een enterprise context voor velen een relatief nieuw verschijnsel zijn, bestaan deze database-systemen al geruime tijd en in veel gevallen net zo lang als hun commerciële tegenhangers. PostgreSQL is in 1986 ontstaan aan de Universiteit van Californië in Berkeley en gemaakt door Michael Stonebraker. Hij wilde de beperkingen van het relationele model van Ingres, een eveneens aan Berkeley ontwikkeld database-systeem, adresseren. Ingres heeft aan de basis gestaan van veel moderne commerciële database-systemen die tegenwoordig worden toegepast, zoals Sybase, SQL Server en Informix. Postgres behoort tot dezelfde familie van gevestigde database-systemen en is gemaakt om in brede enterprise context ingezet te kunnen worden.

### Praktijkvoorbeeld HI5

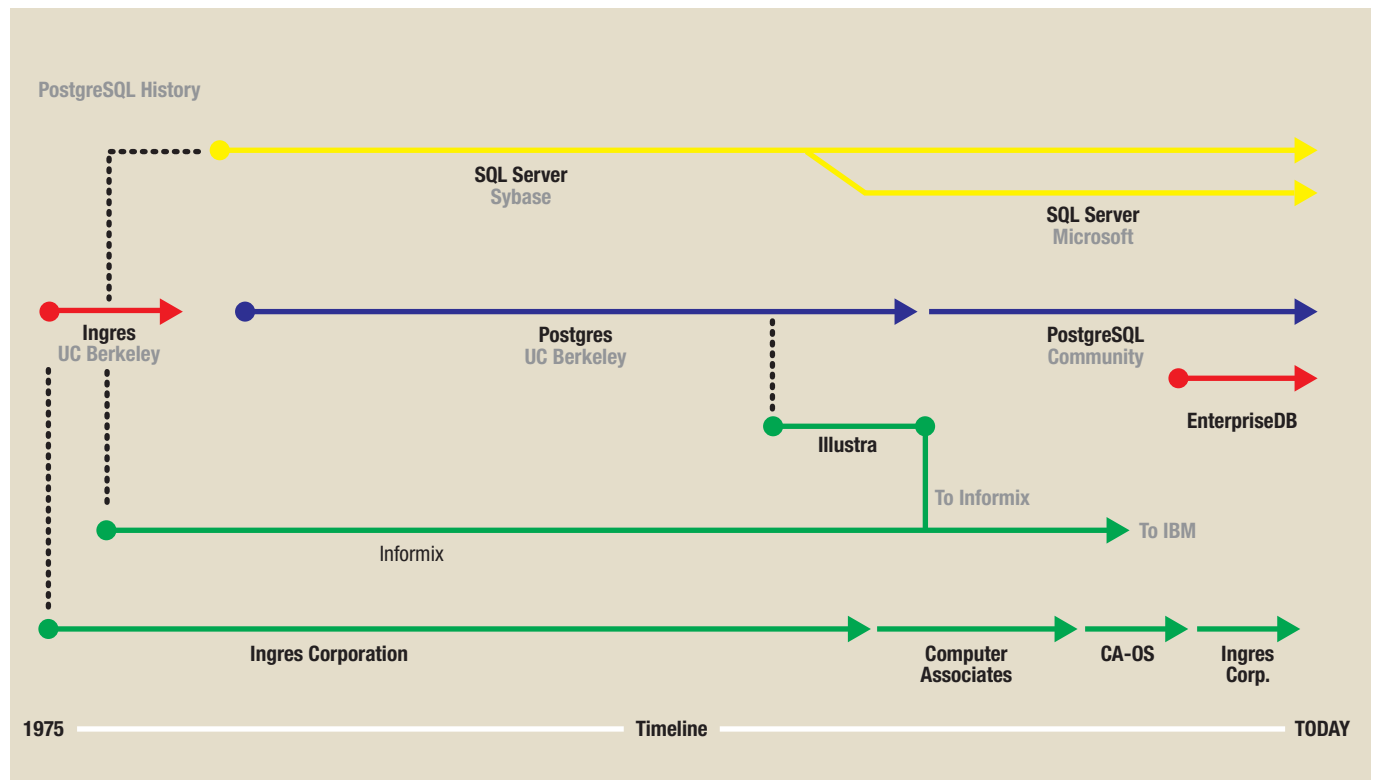
Er zijn nu talloze voorbeelden van gebruik van Open Source database-systemen in kritische bedrijfstoeepassingen. Zo maakt bijvoorbeeld de sociale netwerksite 'HI5' gebruik van een Open Source database platform. Florist' Transworld Delivery (FTD), het Amerikaanse equivalent van de in Europa bekende Interflora bloemisten bezorgservice, heeft een Open Source database-systeem ingezet om de pieken rond Moederdag en Valentijnsdag op te kunnen vangen.

De sociale netwerksite HI5, in de top-20 van meest bezochte sites ter wereld, draait de volledige database-infrastructuur op PostgreSQL. Meer dan vijftig database servers handelen het verkeer af van de duizenden gebruikers van HI5. PostgreSQL vormt de kern van de architectuur waarin dynamische profielen afgehandeld moeten worden, net als sociale interactie tussen gebruikers en de verschillende blogpostings.

Door gebruik te maken van een flexibel Open Source database-systeem kon HI5 de database-systemen schalen om aan de business requirements te voldoen. Met een traditionele oplossing had men vanuit kostenperspectief moeten kiezen voor een business model waarbij minimalisatie van database-licentiekosten de belangrijkste *driver* was.

### Praktijkvoorbeeld FTD flowerists

Florist' Transworld Delivery (FTD) is marktleider in de bloemistenwereld. De database-applicaties van FTD zijn een schoolvoorbeeld van bedrijfskritische database-toepassingen, zeker gedurende bepaalde periodes zoals Kerstmis, Valentijnsdag en



Afbeelding 2: PostgreSQL geschiedenis.

Moederdag. Begin 2007 installeerde FTD een nieuw op Oracle gebaseerd administratiesysteem voor verzendingen. Op Valentijnsdag ontvangt FTD circa 25 keer meer bestellingen dan op gewone dagen. Vanwege de nieuwe databases ontstonden er aanzienlijke performance-problemen met het ordersysteem, waardoor uiteindelijk zelfs de verzendingsrapporten handmatig gemaakt moesten worden.

Omdat Moederdag naderde, moest op korte termijn een oplossing voor dit urgente probleem worden gevonden. Vanwege kostenoverwegingen was uitbreiding van het Oracle-platform in dit geval geen optie. FTD heeft samen met EnterpriseDB een op PostgreSQL gebaseerde oplossing bedacht, waarbij een replica is gemaakt van de Oracle productie-database naar Postgres Plus Advanced Server. Het administratiesysteem voor verzendingen kan vervolgens gebruikmaken van deze replica-database, waarmee de performance bottleneck was opgelost. Deze oplossing is in zes weken geïmplementeerd tegen een fractie van de kosten van de uitbreiding van het bestaande database-systeem en was zodoende voor de Moederdag-piek volledig in productie.

## Markt-trends

De ingezette adoptie van volwassen, schaalbare en betrouwbare Open Source database-systemen zal doorgaan, in het bijzonder voor Open Source database-systemen die worden ondersteund door een leverancier. Een aantal leveranciers is bezig met differentiatie van hun producten en diensten om beter aan te kunnen sluiten bij de klantbehoeften en we verwachten dat er een sterke toename zal zijn in het gebruik van Open Source database-systemen bij bedrijfskritische toepassingen.

Het bovenstaande illustreert dat de Open Source database-systemen hun volwassenheid bewezen hebben en klaar zijn voor 'het echte werk' en ook in productieomgevingen kunnen worden ingezet.

Veel organisaties maken op dit moment een begin met het inzetten van Open Source database-systemen in een BI-omgeving.

Dit is in meerdere opzichten een ideale eerste stap om Open Source database-platformen in een enterprise-omgeving in te zetten. Hoewel het belang van rapportages niet onderschat mag worden, zijn ze in de meeste gevallen niet bedrijfskritisch in de zin dat eventuele tijdelijke niet-beschikbaarheid direct impact heeft op de bedrijfsomzetten. Bovendien is deze informatie meestal niet de enige bron van informatie die door de business wordt gebruikt voor het nemen van beslissingen, waarmee een eventueel risico verder verkleind kan worden.

## De volledige omgeving kan worden opgebouwd op basis van Open Source componenten

Vanuit een breder perspectief op de BI-omgeving kan de volledige omgeving worden opgebouwd op basis van Open Source componenten. Veel leveranciers van Open Source BI-oplossingen, zoals Pentaho, Jaspersoft en Talend, leveren complete rapportage- en data-integratieproducten en -diensten die een traditionele BI-oplossing kunnen vervangen. Indien deze complete op Open Source gebaseerde oplossingen vanuit een financieel perspectief vergeleken worden met de traditionele oplossingen, dan ziet de return on investment voor Open Source oplossingen er zeer aantrekkelijk uit.

### Geert Batterink en Jim Mlodgenski

G. Batterink (geert.batterink@accenture.com) is Manager Accenture Technology Architecture Netherlands.

J. Mlodgenski (jim.mlodgenski@enterprisedb.com) is Vice President EnterpriseDB.

ARE YOU THE ONE?

[www.HierGeenNummer.nl](http://www.HierGeenNummer.nl)