

Nieuwe release DB2 is vergaand Oracle compatibel

Code Name COBRA

Klaas Brant

Na enige maanden intensief bètatesten gaf IBM op vrijdag 19 juni 2009 de nieuwste versie van DB2 on Linux, UNIX and Windows (DB2/LUW) vrij. Toen ik de eerste bètaversies in handen kreeg, dacht ik serieus dat ik met DB2/LUW versie 10 te maken had. Zo veel nieuwe functionaliteit, dat moest wel een major release zijn. Niet dus, het is release 9.7 geworden.

Het had allemaal zo mooi kunnen zijn: IBM zou Sun kopen en tegelijk Oracle nog een tik op de neus geven met een nieuwe release van DB2 die 'Oracle compatible' zou zijn. Hoe anders was de harde realiteit, Oracle ging met Sun aan de haal en de open source database Postgres is IBM net een stap voor door ook 'Oracle compatible' te zijn. In dit artikel beschrijf ik een paar nieuwe zaken van de DB2 LUW release 9.7 die tijdens de ontwikkeling en testen de codenaam Cobra had. Het is geen compleet overzicht, want er is erg veel nieuws. Wie een totaal-overzicht wil hebben kan het beste de What's new? manual (247 pagina's) lezen.

SQL compatibility enhancements

Een van de doelstellingen van de Cobra release was om Oracle compatible te zijn. DB2 kan niet 100 procent compatible met Oracle zijn, want de twee databases zijn diep van binnen 'by design' anders. Het beste wat DB2 kan doen is het gedrag van Oracle zo veel mogelijk nabootsen. En dat gaat ver, heel ver! Zo is er een SQL*PLUS interface gekomen en zijn er views op de DB2 catalog gemaakt die Oracle data dictionary compatible zijn. Er zijn system modules aanwezig die ook standaard aanwezig zijn in Oracle. Deze modules zijn interfaces naar job scheduling, alerts management, LOB-manipulatie, utility's en nog veel meer. Allemaal compatible met hun Oracle-broertjes. De grootste schok is natuurlijk de PL/SQL engine die nu in DB2 aanwezig is. In DB2 worden de PL/SQL statements door een compiler in DB2 opgeslagen op een soortgelijke manier zoals dat in Oracle gebeurt. Ook hier gaat de compatibiliteit weer erg ver. Zelfs PL/SQL anonymous blocks kunnen worden opgeslagen. Maar met PL/SQL alleen ben je er natuurlijk niet. Oracle kent ook andere datatypes zoals NUMBER, VARCHAR2 en DATE met een time component. Ook die zijn allemaal aan DB2 toegevoegd. Ook zijn er erg veel functies aan DB2 toegevoegd die kenmerkend

voor Oracle zijn: LEAST, GREATEST, NVL, DECODE en nog veel meer. Er is zelfs een nieuw ISOLATION LEVEL dat CURRENTLY COMMITTED heet. U raadt het al, op deze manier gaat DB2 in een modus die sterk aan de Oracle versioning doet denken. Al deze zaken worden niet geëmuleerd maar zijn native in DB2 geïmplementeerd.

Prachtig allemaal, wat is het 'slechte nieuws'? De compatibiliteit moet expliciet aangezet worden en dat gebeurt op instance niveau! Met DB2SET DB2_COMPATIBILITY_VECTOR="ORA" komt DB2 in een Oracle modus en worden zelfs elementen van SQL veranderd om de typische zaken uit het Oracle dialect te ondersteunen; in dezelfde instance dus geen native DB2 database in combinatie met een Oracle compatible database. Maar als DB2 gebruiker wil je natuurlijk best gebruik maken van sommige mooie Oracle-zaken. Gelukkig is er goed nieuws voor de DB2-gebruiker; vele zaken (zoals typische Oracle-functies en de isolation currently committed) werken altijd en in alle versies. DB2's Oracle compatibiliteit kan niet overweg met Oracle's complex data type. Dus PL/SQL met daarin een TYPE declaratie kan niet gecompileerd worden. Aangezien Oracle zijn XML op een complex data type heeft geïmplementeerd, kan DB2 (nog) niet de Oracle XML-zaken ondersteunen. Verder mag een PL/SQL trigger geen FOR EACH STATEMENT bevatten.

U begrijpt dat IBM veel tijd en geld heeft besteed om DB2 te veranderen. Om te voorkomen dat mensen gaan converteren van Oracle naar de gratis (Express-C) en de goedkope (Express en Personal) edities van DB2, zijn de Oracle-faciliteiten in deze versies uitgeschakeld. IBM heeft hoge verwachtingen en verwacht dat mensen van Oracle naar DB2 zullen converteren. Het kostaspect zou de grote motivator moeten zijn. Donald Feinberg van Gartner schrijft in zijn publicatie The Cost Optimization Reality of DBMS Consolidation: "Although DBMS migrations can result in cost savings, the time frame for realizing these savings is longer than many enterprises expect or desire". Een groot deel van de kosten van migratie gaat normaal gezien op aan code re-write. Door de nieuwe compatibility mode van DB2 is dat niet meer nodig. Blijven nog over de kosten voor het operationeel beheer en procedures zoals database backup en reorganisatie. En dan is er natuurlijk de menselijke factor, tegen een Oracle-man zeggen dat we op DB2 gaan draaien is net zoiets als vloei-

ken in de kerk. Een draagvlak vinden in de organisatie zal niet gemakkelijk zijn.

PureXML

XML in DB2 was en is het grote thema in eerdere releases van DB2 versie 9. De naam pureXML is door IBM gekozen om aan te geven dat de hybride engine (SQL en XQUERY) de XML native (pure) opslaat en verwerkt. Dit in tegenstelling tot de grote concurrent Oracle, die XML afbreekt en omzet in een complex data type. Er zijn vele wegen die naar Rome leiden en het is moeilijk te zeggen wat 'beter' is. Zelfs vele XML-only databases hanteren of de ene of de andere methode. Bij IBM is men natuurlijk erg trots dat hun eigen Don Chamberlin, die aan de wieg stond van SQL, ook weer aan de wieg heeft gestaan van XQuery. Net als bij SQL, is het ook nu weer IBM die de eerste zeer geavanceerde optimizer voor XQuery heeft voor query processing. Maar heel eerlijk: ik heb wel eens het idee dat XML in databases kenmerkend zo'n onderwerp is waar veel mensen over praten maar dat niemand echt gebruikt. Als ik dat hardop zeg dan krijg ik natuurlijk een stortvloed van voorbeelden over me heen van gebruikers die het wel gebruiken: nog steeds een druppel op een gloeiende plaat. Maar om te voorkomen dat ik wordt uitgemaakt voor oud en *grumpy* vertel ik u dat XML in de database erg hip is en bijna iedereen het gebruikt of wil gebruiken (u valt natuurlijk in de laatste categorie).

Hoe dan ook, de vorige releases 9.1 en 9.5 (waar is de tijd dat een leverancier normaal kon nummeren) beschikten over een vrij volledige implementatie van XQuery, maar hadden nog een paar beperkingen. Die beperkingen worden door release 9.7 netjes weggewerkt. XML mag nu ook in partitioned tabellen gebruikt worden, evenals in Multi Dimensional Clustering (MDC) tabellen. Partitioned databases (DPF) en temporary tabel-

len ondersteunen nu ook XML. Zoals gesteld gaat het hier over opheffen van beperkingen. Wat wel echt nieuw is zijn de verbeteringen aan de XQuery optimizer. Dit is natuurlijk een soort van 'history repeats'; soortgelijk als we dat van SQL gewend waren, zal bij iedere release de optimizer steeds slimmer worden. Release 9.7 zal sommige filterexpressies (zoals XPATH expressies) slimmer gebruiken op een lager niveau (predicate pushdown). Hierdoor worden indexen op XML beter benut. Maar verwacht nog geen wonderen van XML processing, vergeleken met SQL is het een slak als het gaat om grote hoeveelheden en complexe statements. Dat is bij vele andere XML-implementaties niet anders. Zelfs de meest geavanceerde XML databases moeten het qua snelheid echt afleggen tegen SQL databases. Waren sommige XML zaken in vorige releases nog tegen betaling, alle pureXML zaken in DB2 zijn volledig gratis in alle varianten van DB2. Ook in DB2 release 9.5 wordt alle pureXML gratis met fixpack 3b.

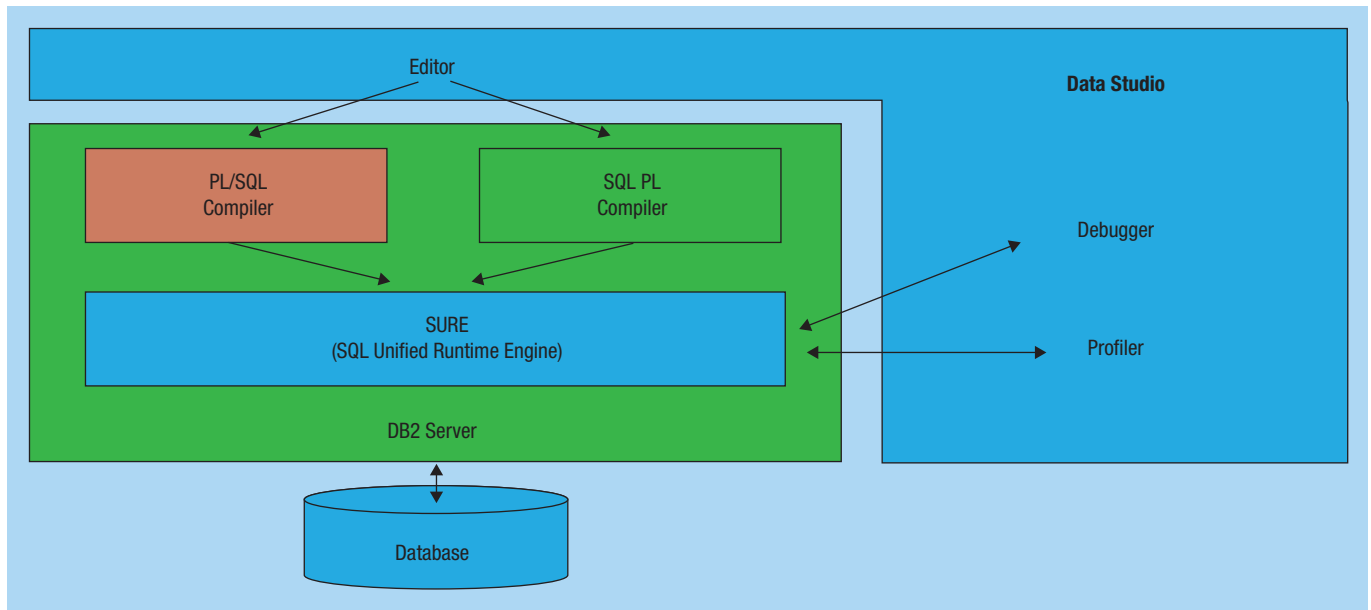
Een verwarring die nog wel eens wil optreden is tussen pureXML en PureQuery. PureQuery is een object oriented data access methode van IBM voor Java- en .NET-omgevingen. PureQuery kan gezien worden als een 'concurrent' voor Spring of Hibernate. Het is onderdeel van Optim Development studio (nieuwe naam voor IBM Data Studio Developer). PureQuery kan PureXML gebruiken, maar eigenlijk hebben ze dus niets met elkaar te maken.

Meer data

Begin jaren tachtig was ik computer operator bij Philips in Eindhoven en onze machine had 16 MB geheugen en 25 GB diskgeheugen. Mijn huidige PC heeft 4 GB geheugen en 1 TB aan diskgeheugen. De snelheid is niet te vergelijken (waarschijnlijk 500 keer sneller). Alles is niet een beetje groter, maar heel veel groter geworden. Ik schrik al niet meer als mijn klanten zeg-

Oracle Database	→	DB2	
Concurrency Control	→	Native support	Not Emulation
SQL	→	Native support	
PL/SQL	→	Native support	
Packages	→	Native support	
Built-in packages	→	Native support	
OCI	→	Native support	
JDBC	→	Native support	
Online schema changes	→	Native support	
SQL*Plus Scripts	→	Native support	

Afbeelding 1: Oracle native support binnen DB2.



Afbeelding 2: Data Studio.

gen dat ze dingen 'forever' willen bewaren. Dat wil zeggen dat de maximum omvang van een tabel ook groter en groter moet worden. In de nieuwe versie is dit maar liefst 64 TB geworden. Voor de doorsnee gebruiker is dat 'groot genoeg'. Wat niet wil zeggen dat in volgende releases ook deze grenzen zullen sneuvelen, zie afbeelding 3.

Maar veel data bezitten geeft ook een ander probleem, namelijk I/O. Het scannen van data (zelfs in parallel) kan lang duren. Daarom biedt DB2 sinds versie 9 compressie; of 'deep compression' zoals de IBM marketingmensen graag zeggen. In tegenstelling tot wat velen denken is dit niet bedoeld om disk space uit te sparen (mooi meegenomen maar disk space kost bijna niets meer), maar om in 1 I/O (één page) nog meer data te lezen. Compressie zorgt namelijk dat er meer rijen op een page passen. Voor de CPU is dat geen probleem, die is erg snel geworden. In versie 9.7 kunnen nu ook indexen, temp tables, large objects (LOB's) en XML gecomprimeerd worden. Deze waren in de vorige release nog onmogelijk. Ook de beperking dat data die onderdeel zijn van replicatie niet gecomprimeerd kunnen worden is nu weg: *alles* kan gecomprimeerd worden. Natuurlijk kan de disk space besparing ook geld opleveren, hoewel professionele gebruikers allemaal een SAN gebruiken dat reeds in de hardware aan compressie doet. De grote besparing is dus bij I/O en niet bij disk space zoals zo veel publicaties melden.

Page size	Version 9.5	Version 9.7
4 KB	2 TB	8 TB
8 KB	4 TB	16 TB
16 KB	8 TB	32 TB
32 KB	16 TB	64 TB

Afbeelding 3: Maximum tabelomvang.

In DB2 versie 8.2.2 heeft een nieuw soort opslag zijn intrede gedaan in DB2: automatic storage. Een mooie en geavanceerde manier om data op te slaan, en een methode die helemaal van deze tijd is. Automatic storage denkt niet in disks maar in storage paths en dat sluit veel beter aan bij het moderne gebruik van opslag in een SAN. Toch waren er nog wat vervelende tekortkomingen bij deze manier van werken. Het grootste probleem was dat bestaande storage (DMS en SMS containers) niet naar automatic storage geconverteerd kon worden. Met de komst van release 9.7 is deze beperking weg. Met een ALTER TABLESPACE command kan een tablespace gemakkelijk geconverteerd worden. Een alternatief is een backup gevolgd door een redirected restore. Ook een van de mooiste features van DMS containers, namelijk een automatische rebalance van data, is nu van toepassing op automatic storage. Als een nieuw storage path wordt toegevoegd aan een automatic storage database (bijvoorbeeld een nieuwe LUN set in de SAN) dan zullen de data over de bestaande storage paths verdeeld worden. Ook een reclaim van unused storage is sterk verbeterd. Hoewel dit laatste best nog wel complex is .

Voor de grotere datawarehousingomgevingen heeft DB2 zijn Data Partition Feature (DPF). Deze feature bestaat al een tijdje (de oude naam voor release 8 was DB2 EEE). Veel mensen zeggen dat door de komst van grotere machines met vele krachtige CPU's, SAN's en de DB2 table partitioning feature de noodzaak van DPF is afgenomen. DPF is een shared nothing omgeving waarbij meerdere machines, ieder met hun eigen DB2, samen één grote database vormen. Het probleem is het verdelen van de data (welke machine krijgt welk deel van de data). In release 9.7 is het geheel iets flexibeler gemaakt en kunnen nu machines (nodes) dynamisch aan het geheel worden toegevoegd. Volgens IBM is DPF nog steeds absoluut high-end als het gaat om grote hoeveelheden data.

In versie 9.1 werd table partitioning geïntroduceerd, een krachtig feature dat een roll-in en roll-out van data in een datawarehousingomgeving mogelijk maakt. Deze feature is in release 9.7 verbeterd door indexen ook te partitioneren en de roll-in en roll-out nog flexibeler te maken.

Innovatie

Is het nog mogelijk om iets te verbeteren aan een Saab, Audi, BMW of Mercedes? Veel mensen denken dat databases klaar zijn en alles wat uitgevonden kon worden, nu zo wel uitgevonden is. Dan bent u bij IBM aan verkeerde adres, hun research komt regelmatig met nieuwe zaken die hun weg vinden in DB2 (denk aan de Multi Dimensional Clustering, MDC). Ook versie 9.7 heeft weer een paar mooie trucjes die de concurrent nog moet uitvinden.

Scan Sharing. Als een SQL statement bezig is met een scan van de data dan kan een tweede statement de scan *sharen* met de eerste. Dit is iets anders dan handig gebruik maken van de data in de bufferpool. De data worden maar één keer door de scan processor gehaald (denk aan decompressie). Als het eerste statement klaar is dan ontkoppelt DB2 de twee statements en kijkt of het tweede statement nog extra data moet scannen.

More Efficient Query's with Literals. In het verleden werden query's opnieuw geëvalueerd door de optimizer als ze niet 100 procent gelijk waren. In release 9.7 wordt nu gekeken waar in de query's literals staan en de statements worden opgeslagen in de cache alsof er host variables stonden. Hierdoor zullen aanzienlijk minder statements in de cache komen en dus veel langer in de cache bewaard kunnen worden. Ook de snelheid neemt toe omdat er minder query optimalisaties nodig zijn. Met name de vele pakketten die hun SQL dynamisch samenstellen en geen gebruik van host variables maken, zoals bijvoorbeeld Siebel, zullen veel baat van deze feature hebben.

Migratie

Migratie naar release 9.7 is alleen mogelijk van een andere versie 9 release (gratis voor gebruikers met een onderhoudscontract) of een versie 8 release. Mensen die nog steeds versie 7 (die inmiddels niet meer ondersteund wordt) gebruiken zullen dus eerst naar versie 8 moeten overstappen. De gratis variant van DB2 (Express-C) kan gedownload worden, maar heeft minder 'Oracle features'. Daarmee schaart de Oracle compatibiliteit zich bij de 'enterprise' features zoals HADR en Federated Data. Voordat men een database migreert, moet eerst goed gekeken worden of er geen oude type-1 indexen meer zijn, want deze kunnen niet gemigreerd worden. PureXML is sinds februari 2009 gratis en daarmee is de oude XML extender komen te vervallen in release 9.7.

Met iedere release komt er altijd een lijst van zaken waarvan IBM aangeeft dat ze in een toekomstige release niet meer zullen werken of niet meer onderhouden worden (de zogenaamde deprecated features). In deze release is deze lijst nogal lang en bevat een aantal 'schokkende' dingen. Zo wordt de hele toolset



Don Chamberlin, IBM Fellow, Professor Emeritus at Almaden Research Center.

(onder andere het fameuze Control Center) vervangen door de, eveneens gratis, IBM Data Studio tool. Of alle gebruikers (alweer) een verandering van de toolset zullen toejuichen valt nog te bezien. Ook de Health monitor is vervangen door een nieuw tool: Data Studio administration console. IBM raadt het gebruik van DB2 Governor en Query Patroller af en wijst naar de DB2 Workload manager als vervangende tool. Hoofdstuk 17 van de What's new? manual bevat alle details en het is verstandig deze goed door te nemen bij een migratie.

Conclusie

De Cobra release is een grote release met heel veel bruikbare features. Of mensen zullen migreren van Oracle naar DB2 valt nog te bezien. Migratie is meer dan een database die compatibel is, maar het is een stap in de goede richting voor degenen die een migratie overwegen en in tijden van recessie kan geld een motivatie zijn. Wellicht dat daarom IBM er voor gekozen heeft deze release 9.7 te noemen en niet 10.1.

Hoe dan ook, een migratie naar 9.7 moet goed voorbereid worden, want naast het vervangen van software moeten er bij deze release ook wel enkele procedures en tools vervangen worden. De nieuwe toolset is overigens niet gebonden aan versie 9.7, dus het is verstandig om deze te implementeren bij uw huidige release voordat u over gaat op release 9.7

Klaas Brant (kbrant@kbce.nl) is DB2-specialist en directeur van KBCE b.v.

The screenshot shows the DBM website homepage with a red header. The main navigation bar includes 'Home', 'Het blad', 'Speciaal', 'Archief', 'Database Tools', 'Bl-matrix', 'Abonnementen', 'Advertenties', 'English', 'Agenda', 'Contact', and 'Vacatures'. The 'Actueel' section features news items such as 'HP lanceert nieuwe versie Noolver 2-6-2008', 'Nieuwe gepatenteerde technologische innovaties voor real-time BI', and 'Cognos reikt Partner Sales Excellence Award uit'. The 'Zoeken' section has a search bar and filters for 'Zoek op de site', 'Zoek vacature', and 'Top jobs'. The 'Online archief' section provides access to past issues of Database Magazine. The 'Vacatures Database' section includes a search form and a list of job openings. The 'BI-matrix' section is a large matrix of BI-related products and services. The 'DBM Tools' section offers a search for database tools and leveranciers. The footer contains contact information for Array Publications and a list of featured speakers.

De website van Database Magazine is een belangrijke bron voor database professionals en wordt dagelijks bijgewerkt door de redactie. U vindt hier ook blogs en artikelen uit het blad. Verder biedt de site onder andere:

Actualiteit

Dagelijks verzorgt de redactie het nieuws uit de markt met zowel productnieuws als bedrijvnieuws. Dit nieuws kunt u ook via de online nieuwsbrief rechtstreeks in uw mailbox ontvangen.

Online Archief

Onze artikelendatabase bevat beschrijvingen van alle artikelen die zijn verschenen in het blad vanaf de prille start van het magazine. De artikelen vanaf 2002 zijn bovendien als PDF beschikbaar waarbij abonnees op Database Magazine volledige downloadrechten genieten.

DBM Vacatures

Uit de database van VNU Media (Computable en Intermediair) wordt direct een selectie getoond van vacatures voor database specialisten. Uiteraard kunt u zelf trefwoorden opgeven en via 'Geavanceerd Zoeken' kunt u filteren op datum, branche, opleidingsniveau en de plaats of regio.

Topjobs

Uw vacature op een prominente plaats altijd in beeld? Dat kan in de Topjobs module van Database Magazine.

BI-matrix

De BI-matrix is een onafhankelijk, actueel en gratis overzicht van de verschillende leveranciers van BI-software, geplaatst in een doorzoekbare matrix. De Matrix bevat bovendien een 'matchmaking' module waarbij u kunt selecteren op vele criteria.

Database Tools

Een database met Database Tools waarin op basis van selectiecriteria kan worden gezocht. U vindt hier beschrijvingen van de tools en van de leveranciers op het gebied van onder andere OLAP, ETL en Database beheer.