

SAS en Teradata starten diepgaande technische samenwerking

In-database analytics

Hans Lamboo

SAS en Teradata hebben een strategisch partnership gesloten en werken samen aan een concept dat zij In-database Analytics noemen. Keith Collins, CTO van SAS, licht een tipje van de sluier op.

SAS, of SAS Institute zoals het bedrijf formeel en een beetje stoffig heet, is altijd al een buitenbeentje geweest in de BI-markt. Lange tijd werd SAS door sommigen niet erkend als pure play BI-vendor, en ten tweede is SAS *privately owned*, dus niet aan de beurs genoteerd. De aardverschuiving op de BI-markt is in elk geval aan SAS voorbij gegaan. CEO en oprichter Jim Goodnight gaf al eens eerder te kennen "niet te hoeven verkopen." SAS vaart in de BI-markt een eigen koers en sloot een strategisch partnership met Teradata, dat zich net dit jaar had losgemaakt van moederbedrijf NCR.

De samenwerking betekent onder meer dat SAS en Teradata hun producten technisch integreren en dat er gezamenlijke marketing-, verkoop- en services-activiteiten worden ondernomen. Het doel: het ontwikkelen van in-database analytics om zo enorme tijdwinsten te kunnen realiseren.

Embedded

Volgens IDC's Dan Vesset speelt in-database analytics een sleutelrol in de ontwikkeling van business analytics-oplossingen. Enerzijds zorgt het voor minder dataverplaatsing, anderzijds verbetert het de performance. Ook voor Keith Collins, CTO van SAS, komt de verloving met Teradata niet als een verrassing. "SAS heeft al sinds jaar en dag een goede technische relatie met alle database vendors. Door middel van API's helpen we ze met

bulk-loads en bulk-extracties. Al begin jaren negentig vroegen we de database-leveranciers of ze dachten dat de tijd rijp was voor analytics. Maar hun klanten waren toen nog helemaal gericht op transacties, op ERP. Pas nadat men dat redelijk op orde had, begonnen de bedrijven zich te realiseren dat de beschikbare cijfers moesten worden gegroepeerd en geanalyseerd, kwam er interesse in analytics en ontstond wat we nu BI noemen. Nu zijn we aanbeland op het moment dat we met de vendors in gesprek zijn over technieken en instrumenten die nodig zijn om analytics in databases te kunnen embedden", verhaalt Collins. "Dat maakt alles veel sneller. Als erfenis uit het verleden hebben mensen bij het woord 'sneller' altijd het idee dat dat alleen kan met *propriety* technologie. Dat was vroeger, maar is nu niet meer aan de orde. We schrijven nu direct bovenop het systeem, daar is niets *propriety* aan."

Als je de database behandelt als een grid krijg je lineaire schaalbaarheid

Volgens Collins was Teradata de eerste database-leverancier die deze beweging in de markt begreep en zijn ze bovendien de grootste leverancier van datawarehouses met *partial scheduling* en ondersteuning van grootschalige query's voor analytics. "Ze zijn de absolute top. En als je partner van SAS wilt zijn, moet je tot de top behoren. Dan doel ik niet zozeer op commerciële relaties, maar op partnerships waarbij je diep in elkaars technologie duikt. We zijn nu verloofd, maar een trouwerij staat nog niet gepland", grapt Collins. "Samen willen we in-database analytics ontwikkelen. De eerste twee stukken technologie zullen met name de data-filtering enorm verbeteren."

De eerste stap: modellering

Collins beschrijft een roadmap, die uit vier stappen bestaat. Als eerste brengt het tandem de functionaliteit onder in de database, hetgeen een enorme performance-winst oplevert over de gehele suite van datamigratie en -integratie, Business Intelligence en analytic intelligence. Als tweede deel van de technologie pakt

SAS op de DB/M website

In de onlangs gestarte poll op onze website www.dbm.nl werd aan de bezoekers gevraagd naar hun mening over de toekomst van SAS. Een kleine 41 procent meent dat SAS zelfstandig blijft, ruim 30 procent denkt dat SAS wordt overgenomen door een grote partij en 28 procent ziet SAS fuseren met andere kleinere partijen.

De poll blijft online tot in het nieuwe jaar; vanaf 2 januari 2008 is er elke maand een ander onderwerp.



Keith Collins, CTO van SAS: "Als je partner van SAS wilt zijn moet je tot de top behoren".

men de overgang van modellering naar in productie nemen aan. Collins: "Vandaag de dag wordt bij het in productie brengen van model-scores meestal een keuze gemaakt uit twee mogelijkheden. De een is om alle data op te halen, het tegen de modellen te laten draaien, en de scores terug te koppelen. De andere mogelijkheid is om het model te herschrijven in SQL, C of Java, en het in de database te laten draaien. Dat laatste is een ontzettend tijdrovende en ingewikkelde aangelegenheid. Ik ken een aantal praktijkvoorbeelden, waarbij men drie maanden bezig was met het hele proces van het herschrijven van de modellen, het testen ervan en het in productie brengen. Wat gebeurt er als wij dat proces kunnen terugbrengen van drie maanden tot een paar dagen?"

De tweede stap: 4GL in de database

Als tweede noemt hij de toepassing van databases als data grid. "De performance-winst op de eerdergenoemde score-activiteit is enorm", zegt Collins. "Daartoe hebben we de kern van de 4GL in de database gebracht. Daarmee beschikt de database over een

krachtige transformatietaal, waarmee omzettingen naar andere modellen in de database plaatsvinden".

Als voorbeeld noemt Collins een test met een eenvoudig regressiemodel voor een creditcard. Het model werd op een gewone database gedraaid, die was verbonden met een Teradata-database. De hoogst mogelijke score was 16.000 leesen schrijftransacties per seconde. Vervolgens werd de test herhaald, maar nu rechtstreeks op de node van de Teradata-box. Het aantal liep op naar 77.000. Daarna werd de Teradata-toepassing met 4GL gebruikt over zes nodes en het resultaat liep op naar 750.000 transacties per seconde. "Dat gebeurt er als je de database behandelt als een grid: je krijgt lineaire schaalbaarheid", legt hij uit.

Maar er is meer, wat Collins omschrijft als *scoring in batch*. "Het sjabloon van de transformatietaal kun je niet alleen embedden in een database, maar ook in een MQ-engine, een workflow- of een BPM-engine. Het modelleringsproces kan daardoor toegepast worden via het web, via de workflow, via de database – allemaal werken ze met hetzelfde model".

De eerste twee stappen heeft het tandem inmiddels ontwikkeld en volgens Collins is men hard op weg om ook de laatste fase van de modelleerproblematiek op te lossen.

De derde stap: encapsulation

Als derde stap willen Teradata en SAS het security framework naadloos maken. Dat zou het einde betekenen van het gebruik van userID's en passwords tussen het BI-systeem en de database. "Daarvoor willen we van de Temp Space een *shared managed resource* maken", zegt Collins. "Alle tijdelijke data bevinden zich dan zowel in de SAS-processen als in de database-processen. Daardoor ontstaat een naadloze security, eenvoudiger management en governance van de security, terwijl je nog steeds flexibiliteit biedt aan de business". Vervolgens wordt in deze context gekeken naar het concept van *stored procedures*. "Je kunt van elke SAS-functionaliteit een *encapsulation* maken", licht Collins toe. "Autorisatie, een forecast, al die ingewikkelde dingen die we niet konden doen maar nu wel, met analytics op de nodes. Dan kijk je naar zo'n encapsulation: een naam en parameters, precies de elementen waaruit een stored procedure bestaat. We laten de SQL zich gedragen als stored procedure en vervolgens als SAS stored process."

De vierde stap: modelleren over het grid

De vierde stap van de roadmap ligt nog in het verschiet. Het gaat om de technologie waarbij modelleeractiviteiten door het gehele data grid worden verbeterd. "Daar moet nog een hoop aan gebeuren, zowel bij SAS als bij Teradata", zegt Collins. "De databases van nu hebben te weinig functionaliteit, gebruiken een vorm van aggregatie waar we weinig mee kunnen en hebben geen goede caching-techniek."

SAS blijft doorgaan met het bij elkaar brengen van drie dingen: data-integratie, analytics en information delivery. "In de BI-markt zou een van de doelen op lange termijn het volledig elimineren

van rapportages moeten zijn", stelt Collins. "De technologie achter Master Data Management biedt de mogelijkheid om Activity Based Modeling in processen onder te brengen, om analytics in te bakken. We zouden dus in staat moeten zijn om in te breken in het proces op het punt waar het zich voordoet. Stop het proces en wijzig het systeem. Maar dat staat nog op onze verlanglijst."

Toekomst

De interesse in MDM-technologie is groot bij SAS. "We willen het voor onze klanten mogelijk maken de data te cleansen bij de invoer. Ik geloof er heilig in dat de dingen die we nu met een operationele datastore doen, binnen afzienbare tijd in de database zullen plaatsvinden. Het real-time opschonen van databestanden bijvoorbeeld. Maar daarvoor moeten we eerst beter worden in MDM", voorspelt Collins.

Binnen SAS werken nu ruim 60 mensen samen met technici van Teradata aan de eerste twee stukken technologie. "De samenwerking verloopt in een prima werksfeer", zegt Collins. "We werken in twee labs, SAS in Cary en Teradata in San Diego. Je ziet en hoort technici met elkaar bellen, mailen, vergaderen. De mensen in beide labs realiseren zich dat er nog zoveel te ontdekken is en werken daar enthousiast aan. Voor de levering mikt SAS op versie 12 van haar platform".

Vooralsnog is SAS een van de laatste pure play BI-vendors; niet aan de beurs genoteerd, in tegenstelling tot bijvoorbeeld MicroStrategy, Business Objects en (nog) Cognos. SAS en Teradata denken met de ontwikkeling van in-database analytics hun marktpositie te verstevigen. SAS en Teradata hebben nu 240 wederzijdse klanten, de gezamenlijk te ontwikkelen technologie voor in-database analytics moet dat aantal verveelvoudigen. Daarnaast ontwikkelt SAS nog op andere terreinen binnen analytics nieuwe technologie, zoals het real-time fraudedetectieplatform dat samen met de Engelse bank HSBC werd ontwikkeld en succesvol is geïmplementeerd.

**We laten de SQL zich gedragen
als stored procedure en vervolgens
als SAS stored process**

Met een sterke persoonlijkheid als Jim Goodnight aan het roer, die bovendien niet in is voor een verkoop, is het vertrouwen bij SAS op een mooie toekomst groot.

Hans Lambou is hoofdredacteur van Database Magazine.



Ben jij de BI specialist die weet wat-ie wil?

Dat mag jij ons dan komen vertellen!

Join FourPoints !!

Dé specialist in Data Warehousing & Business Intelligence