

VELDWIJK

It's the spreadsheet, stupid!

In mijn vorige column schreef ik over de kracht van het combineren van gestructureerde gegevens (bron: databases) en ongestructureerde gegevens (bron: Internet). Ik heb het punt daar niet uitgewerkt, maar je kunt je afvragen of er iets tussen deze categorieën van gegevens in ligt. Sommigen onder ons zullen zich misschien herinneren dat er al eens een poging is ondernomen om het antwoord op deze vraag te geven.

Zo'n tien jaar terug ontstond de gedachte dat semi-gestructureerde gegevens kunnen worden gevangen in complexe objecten en dat rdbms'en zouden moeten worden uitgebreid met faciliteiten om deze objecten op te slaan en methoden om ze te ontsluiten. Sommigen meenden dat de idee om de noodzakelijke objectgeoriënteerde uitbreidingen op te nemen in rdbms'en een doodlopende weg was en dat we het rdbms zouden moeten laten vallen voor het zuivere oodbms. Na enkele jaren waaide de hype over zonder veel sporen na te laten in de database praktijk. En sindsdien is het stil. Database toepassingen zijn vandaag de dag nog steeds bijna allemaal tamelijk stomme opslagplaatsen voor *text and number* data.

Ondertussen is er nog steeds een wereld van niet-helemaal-gestructureerde data die we graag zouden willen opslaan in databases en ze daarmee beschikbaar maken voor vormen van geautomatiseerde gegevensverwerking die verder gaan dan *opgoogelen*. Dat dergelijke gegevens niet passen in de starre structuren die zo kenmerkend zijn voor (r)dbms'en, betekent anno 2004 niet dat ze niet elektronisch worden vastgelegd. Dat gebeurt wel degelijk en niet alleen in Word- of HTML-formaat. De praktijk in heel veel organisaties laat zien wat de database is voor al die toepassingen die niet uit de voeten kunnen met starre en dure administratieve, op database-technologie gebaseerde, toepassingen: Microsoft Excel!

Wie niet werkt met heel grote hoeveelheden gegevens, wie geen behoefte heeft aan het gelijktijdig kunnen wijzigen van gegevens en wie wel behoefte heeft aan flexibiliteit, die mijdt database-toepassingen als de pest en stopt zijn gegevens in een spreadsheet. En jazeker, die semi-gestructureerde gegevens die in al die miljoenen werkbladen ligt opgeslagen, is voor 99,9 procent de stomme *text and number* informatie die we tien jaar terug achter ons wilden laten. En nee hoor, geen enkel artikel over spreadsheets in dit en andere vakbladen op Database-gebied. Ongehoord!

De aanleiding voor deze column is niet eens zozeer dat ik wel eens met Excel werk. U moet weten dat mijn bedrijf de kern van haar administratie in Excel voert en ondergetekende die administratie destijds heeft opgezet. Wij voldoen namelijk aan

de drie criteria die hierboven zijn genoemd en dus laten wij vooralsnog de database-beker aan ons voorbij gaan – en dat terwijl we ons profileren als gegevensspecialisten *par excellence*. We roepen tegen elkaar dat we eens goed de tijd moeten nemen om een flexibele toepassing voor onszelf te bouwen zodra we het wat minder druk hebben, maar helemaal eerlijk is dat niet. Ondertussen doen we uitgebreid ervaring op met de enorme nadelen die het werken met een product voor semi-gestructureerde gegevens oplevert en die nadelen zijn niet gering: kopieerfouten, *overrulen* van formules en dergelijke leiden tot veel extra inspanning en irritatie en zonder allerlei organisatorische maatregelen zouden we zeker echte schade hebben gehad. Ondertussen kom ik bij onze opdrachtgevers dezelfde 'Excel-lende' tegen, vaak wel met echte schade. Structureel werken met spreadsheets is een ramp, maar omdat echte automatisering veelal een nog grotere ramp is, gebeurt het op zo'n grote schaal.

Excel-fans die DB/M lezen (zijn die er?) zullen natuurlijk tegenwerpen dat er mogelijkheden zijn om verkeerd gebruik van spreadsheets tegen te gaan. Dat argument maakt op mij geen indruk omdat het gerommel in de marge is en tegelijk vormt dit de harde aanleiding voor deze column. Onlangs zag ik dat Excel zelfs (data)constraints ondersteunt! "Nog even en we krijgen een *rollback*-mechanisme en *grid computing* aangeboden," dacht ik in eerste instantie. Daarna realiseerde ik me dat dit precies de autistische arrogantie is die het database-wereldje zo vaak kenmerkt. De blijvende gedachte is dat Excel ons op twee niveaus een les leest. De gebruikers laten ons met hun miljoenen werkbladen en eindeloos geploeter zien hoe ver ze willen gaan om uit de klauwen van ons dataspecialisten te blijven. En Microsoft laat ons zien dat ze moeite willen nemen om de kloof van hun klant te verkleinen.

Ikzelf ben arrogant genoeg om te geloven dat aangepaste database-technologie meer perspectief biedt dan nieuwe Excel versies, maar eerst moeten dan onze ogen open. Denk er eens over na: het antwoord op de uitdaging van semi-gestructureerde gegevens is niet het oodbms of het oo-rdbms maar het **p**-rdbms, waarbij de p staat voor *permissive*. Ik hoop nog eens daarover in DB/M te schrijven. Gebeurt dat niet dan ben ik waarschijnlijk bezig met het bouwen van een database-toepassing voor onze administratie.

René Veldwijk

Dr. R.J. Veldwijk (rene.veldwijk@faapartners.com) is werkzaam bij FAA Partners.