

Testen is niet echt de tofste job voor de doorsnee ontwikkelaar. En als de aanvankelijke frisse tegenzin overwonnen is en men er daadwerkelijk aan begint, wordt het echte probleem al snel duidelijk. Grondig testen zonder gegevens gaat eenvoudig weg niet. Het intokkelen van AAAA's, BBBB's en zoverder en 1234... cijfercombinaties om een beperkt aantal records op te vullen in de database is lastig, weinig productief en ronduit vervelend. Het alternatief is te proberen om de databank op een intelligente(re) manier te gaan opvullen. En dat gaat niet zonder slag of stoot.

bespreking

Data genereren met Turbodata

Sterk niche tool met schoonheidsfoutjes

Goede data genereren is minder eenvoudig dan het lijkt. Dat was voor de RAD Race editie 2001 business case niet anders. Die case was op zich stevig uit de kluiten gewassen, maar de jury wil ook kunnen beoordelen hoe de applicatie omgaat met gegevens in een "echte", - lees opgevulde - database. In het beperkte RAD Race tijdsbestek kan een team natuurlijk niet méér aanmaken dan de voor eigen gebruik bestemde "test" record, zodat de wedstrijdleiding dan maar een "kant en klare" dataset aflevert. Die wordt opgemaakt aan de hand van een datamodel in Access (omdat iedereen dat wel op zijn PC heeft staan, u hoeft hier verder niets achter te zoeken) dat zowat 90% "af" en uitgemodelleerd is. Meer uitwerken zou de teams teveel kunnen beperken en impliciet een bepaalde oplossing doordrukken en dat is niet de bedoeling. Elk team kan de data binnentrekken in het eigen RDBMS en vervolgens naar hartelust anders gaan modelleren. Tegelijk wordt ook deze competentie getest.

ZINVOL VULLEN Het punt is niet de tabellen te maken, maar wél om die zinvol op te vullen. Handig knippen en plakken in combinatie met zoeken en vervangen in Excel is aanvankelijk een goed idee, tot je op de "not enough memory" barrière stoot. Ingeven in een (Access MDB) databank is niet echt een optie vanwege manueel, er is dus meer nodig. Tijdverslindende zoekoperaties op het internet leverden weinig resultaat op: dit is een blijkbaar weinig gevraagde functionaliteit;

testgeneratoren blijken dun gezaaid. Grote leveranciers als CA en Platinum bieden datageneratie vooral aan als onderdeel van hun (dure) test suites. Veel kleine spelers zijn er ook al niet en Turbodata van het Canadese Canam Software viel daarbij positief op.

Turbodata komt in drie versies: de Light (gratis te downloaden en beperkt tot 10 records zonder vervaldatum op de werking van het programma); de Personal editie (1000 rijen per tabel – 99 US \$) en de Professional versie (geen enkele beperking) aan 499 US \$ die door SRM werd getest in versie 3.00.0040.

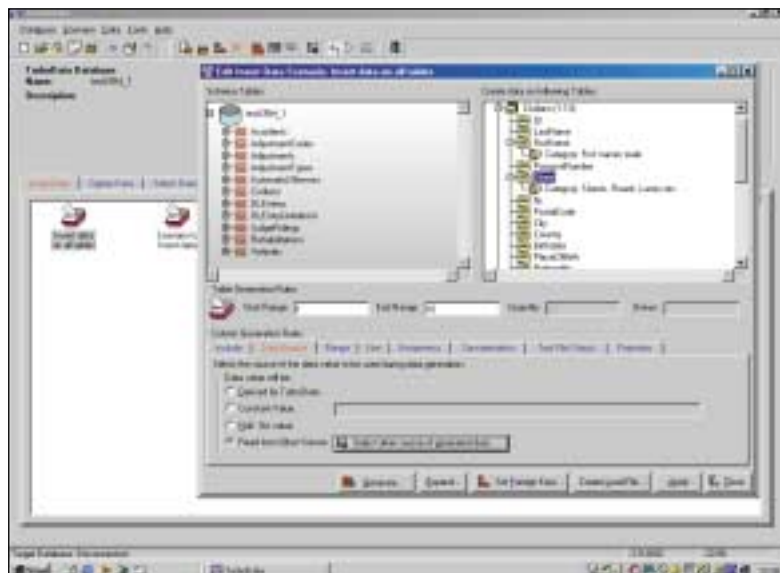
TREE CONTROL Het product is duidelijk op maat van programmeurs gesneden. Het tool start op de meest logische manier door een extern database schema in de eigen Turbodata omgeving binnen te trekken. Daarvoor wordt vanuit Turbodata een rechtstreekse database connectie opgebouwd naar DB2 Universal Database, MS SQL Server, Oracle (alle varianten), SQL Anywhere, NonStop SQL, Sybase of een ander DBMS (met in te vullen connectiestrings); onderliggend wordt native of via ODBC gewerkt zonder dat er nog bijkomende LIB of DLL's dienen bijgeïnstalleerd te worden. De Microsoft Jet engine is ook van de partij en kan via Data Access Objects of het nieuwere ActiveX Data Objects aangepakt worden. Zodra dat gebeurd is, wordt dit schema in een tree control voorgesteld en is het aan de gebruiker om de eisen, limieten en beperkingen voor de generatieslag

op te bouwen in een scenario. Turbodata heeft aparte "scenario" panels (tabs) voor Insert, Delete, Select en Update waar dus eigenlijk telkens een specifieke SQL generator achter de schermen zit. Die SQL output wordt naar de database gestuurd waarmee de gegevens moeten worden gecreëerd via een connection string. Die kan verschillen van de eerdere "binnentrek" connectie en dat biedt aardig wat toepassingen. De SQL kan ook weggeschreven worden voor handmatige aanpassing of latere uitvoering. Maar er kan ook direct een delimited ASCII text bestand aangemaakt worden. Turbodata biedt op heel wat plaatsen SQL-matige ondersteuning door een SQL Builder "opbouw" interface.

WAAIER VAN MOGELIJKHEDEN Dan begint het echte werk. Turbodata herkent het datatype van het veld en gaat er vanuit dat daarop standaard gegevens worden gegenereerd stijl "Naam123" et cetera. De kunst bestaat er nu precies in om exact aan te geven hoe men de gegevens anders aangemaakt wil zien. De gebruiker heeft daarbij een waaier van mogelijkheden. Het bereik (range), de lengte (size, al dan niet met aanvullen van blanco's) en het al dan niet uniek zijn van een waarde in de kolom kan ingesteld worden. Een specifieke kolom kan uitgesloten worden bij generatie, maar er kan evengoed NULL of een constante waarde in gezet worden. Vooral het feit dat er een aantal standaard opsommingen van veel voorkomende types gegevens wordt meegeleverd is handig. Een lijst van steden in Texas, kleuren, vrouwelijke voornamen (in het Engels), et cetera: het is allemaal al voorzien. Ook het interval waarmee een time-date stamp wordt gemaakt is instelbaar. Handig is ook dat de gebruiker elk mogelijk datum-tijd formaat kan instellen; het moment waarop de dbms een commit moet uitvoeren tijdens het genereren maar evengoed de opvulcarakters of het gedrag in geval de kolom niet breed genoeg blijkt.

De gebruiker kan ook rechtstreeks een tabel "live" uit een ander dbms (via de connection string), een Excel sheet of een txt bestand als "bron" gegevens kan aanwenden en volledig of slechts gedeeltelijk gebruiken. Een dergelijke bron kan ook in Turbodata geïntegreerd worden en vormt dan een extra keuzemogelijkheid. Ze kunnen sequentieel, volgens bepaalde intervallen of volkomen willekeurig ingezet worden. Met relatief weinig, maar relevante brongegevens kan u dus serieus aan de slag. Als de parameters op de velden zowat goed staan, volstaat het aangeven van een globale record count op Tabel niveau en het genereren kan starten. Het is deze waarde die afhankelijk van de versie wordt gelimiteerd, verder niets.

REFERENTIEEL CORRECT En dan val je op het diepere, conceptuele probleem. Een normale, in de tijd opgebouwde database die gebruik maakt van de input



FIGUUR 1. Het volledige Insert Data scenario geopend op de RAD Race structuur (links) met rechts de reeds ingestelde opties voor het genereren (vanuit een specifieke bron)

schermen van het bijhorende programma, schrijft gegevens over meerdere tabellen weg, rekening houdend met de geldende samenhang tussen de tabellen (primary en foreign keys) en de afgedwongen regels op referentiële integriteit. Om een database correct op te vullen in een parent-child relatie volstaat het dus niet om de twee tabellen onafhankelijk van elkaar zomaar wat waarden te laten genereren. Turbodata pakt dit in zijn meest recente versie zo automatisch mogelijk aan. De bestaande relaties, afgeleid uit de primary-foreign keys worden maximaal aangewend.

In de bijgeleverde Customer - Order - Orderline - Products Access demo mdb lukte dat volledig. Maar in geval van nood moet de gebruiker selectief de gerelateerde tabellen bijtrekken via de 'Expand' knop en eventueel de relaties scherper stellen in een soort van E-R diagram view. Voor de RAD Race opgave diende was dit bijvoorbeeld wel nodig. Waarna het aantal childs per parent moet vastgelegd en het genereren kan starten. Het "referentieel correct" correct genereren neemt overigens behoorlijk wat tijd in beslag en belast de machine aanzienlijk. Toch weegt dat niet op tegen het "manuele" alternatief. Vervelend is dat een generatieslag niet onderbroken en herstart kan worden.

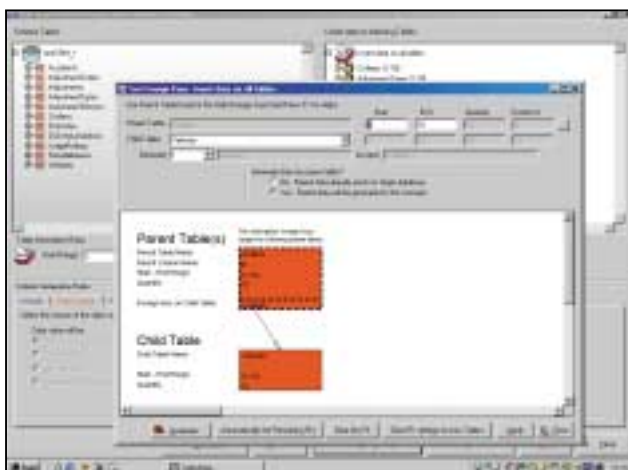
VERSIEHISTORIE Vooraleer de gemiddelde gebruiker zover raakt, is er waarschijnlijk gradueel opgebouwd door een basistabel binnen trekken, de generatieparameters aan te passen, de tabel te genereren en te kijken hoe die er uit komt. Turbodata biedt daarvoor een ingebouwde mogelijkheid om het geleverde werk direct te viewen zonder naar een andere tool te moeten omschakelen. Vervolgens gaat de gebruiker iteratief de parameters voor de tweede tabel opgeven in het scena-

Adv.
Furore

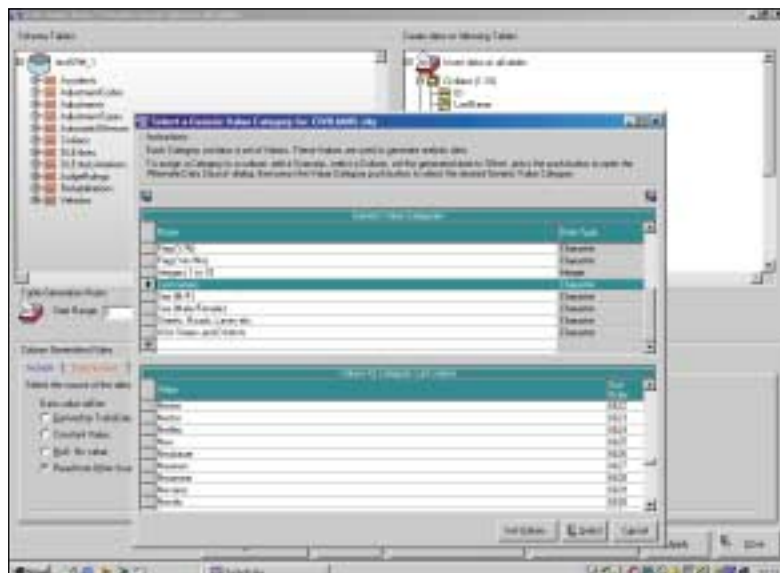
rio en opnieuw genereren. Maar dan loopt het natuurlijk mis. De tabellen zijn daarnet immers gemaakt en moeten eerst leeggemaakt worden omdat er anders dubbele gegevens voor index, primaire sleutel of relatie (aldus de ODBC feedback) kunnen uitrollen. Dat kan door naar de Delete tab en het dito scenario te gaan. En vervolgens terug opnieuw genereren. Standaard wordt weer één volledig delete scenario en één zelf op te bouwen delete scenario aangeleverd. Mocht het Delete scenario nu een (automatische) “schaduw” van het “Insert scenario in opbouw” kunnen zijn, dan zou dit veel manipulatiefouten en bijhorende foutboodschappen met lange nummercodes vermijden. Een soort van “in opbouw mode” staat hoog op het verlanglijstje voor de volgende versie. Maar dat moet lukken. De versiehistorie op de website leert dat er ongeveer om de drie à vier maanden een nieuwe release uitkomt die nuttige features oplevert. Canam telt “traag” (van 3.00.37 in november 2001 naar .40 in augustus 2002) en dat is zeker op prijs te stellen in deze hectische tijden. Men zou er van gaan geloven dat het nog kan: een product dat getest wordt vooraleer het op de markt komt en geen drie ‘toeters & bellen’ major releases per jaar kent.

RUWE KANTJES Een van de verbeteringen is dat er nu ook een interface is ingebouwd om Turbo Reports, de eigen COBOL rapport generator, aan te sturen. Een goede aanwijzing dat Canam ook in grotere systemen schijnt thuis te zijn.

Aan het werken met scenario's is ook geschaafd maar er blijven ruwe kantjes. Hernoemen en aanpassen ervan is vlot uitvoerbaar, net zoals met de (basis) database schema's die vanuit Turbodata direct kunnen benaderd worden en waarop een “database definition rapport” in HTML kan uitrollen. Anderzijds is een scenario met alle achterliggende generatieparameters dan weer niet documenteerbaar. Terwijl een print-out, gezien de insert-delete slag die interactief moet geleverd worden, erg



FIGUUR 2. Het instellen van de Foreign Keys voor het referentieel correct genereren



FIGUUR 3. Beschikbare categorieën met daarin de gehanteerde waarden

handig kan zijn om de parameters op te schrijven en dan in éénmaal in te voeren. Turbodata kan ook een conversietabel aanmaken waarmee tabelwaarden omgezet worden op het moment dat de gegevens gegenereerd worden en transformeert daarmee ook (aanwezige) inputgegevens.

Leuk om weten is dat het Canadese Canam Software eerder al Data Composer ontwikkelde dat als onderdeel van de COOL:Gen suite (vroeger van Texas Instruments, nu intussen CA) door het leven ging. Dat stuk is nu omgewerkt als plug-in voor Advantage Gen van CA. XML ondersteuning wordt momenteel volop voorbereid.

CONCLUSIE Turbodata bezet een interessante plaats in een specifieke nichemarkt en doet dat verre van slecht. De scherpe prijsstelling en de doorgroeimogelijkheden zijn daar niet vreemd aan. Canam levert een goede ondersteuning die kort op de bal speelt en waar de focus op techniek en niet op marketing ligt. Het gevoel is dat Canam niet echt groot is en geen enorme pool programmeurs ter beschikking heeft, maar toch 200% zijn best doet; gestadig verder blijft werken en schaven en pal achter het product staat. Dat maakt dat de balans positief uitvalt als u kan leven met een aantal schoonheidsfouten, waarvan de meesten op de “to do” lijst staan die hier wél geloofwaardig is. In het andere geval zal u in een ander gamma van producten moeten gaan shoppen. Of u kan natuurlijk beginnen met gegevens ingeven...

Ivan Verborgh

Verborgh is onafhankelijk IT-deskundige en doceert informatica aan de Provinciale Hogeschool Limburg (België)