

Praktisch voorbeeld verwerkt niet-optelbare gegevens

Doorbreek de grenzen van OLAP

Bas Pruijn

OLAP is een krachtige manier om snel complexe vragen te kunnen beantwoorden. Met OLAP kan op eenvoudige wijze vanuit de verschillende dimensies naar de feiten gekeken worden. OLAP-technieken als slice-'n-dice, roll-up en drill-down bieden een gestructureerde manier om data te analyseren.

OLAP brengt echter ook duidelijke grenzen met zich mee. De gegevens die door middel van OLAP geanalyseerd worden dienen optelbaar te zijn. Meestal is dit geen praktische beperking. Er zijn echter situaties waarbij deze grens van OLAP tot problemen kan leiden (zie het kader kennisportal).

Een concreet voorbeeld hiervan is het analyseren van prognoses. Deelprognoses van afdelingen zijn namelijk zelden optelbaar naar een totaalprognose voor de organisatie. Wanneer er ook nog deelprognoses voor gedeeltes van afdelingen toegevoegd worden, wordt het geheel nog onoverzichtelijker. Op het eerste gezicht lijkt OLAP geen geschikte methode om niet-optelbare gegevens te analyseren. Maar is er een oplossing?

Introductie tot de oplossing

De stap-voor-stap oplossing voor het met behulp van OLAP analyseren van niet-optelbare gegevens gaat uit van het volgende principe. Wanneer er een sub totaal bekend is, moet deze altijd een optelling zijn van de onderliggende detailgegevens. Wanneer de optelling van de detailgegevens niet klopt met het sub totaal, moeten er dus detailgegevens ontbreken. Om dit op te lossen worden er in deze aanpak details gecreëerd die er voor zorgen dat de details wel optellen tot het sub totaal. Hierna volgt een theoretische uitwerking die wordt toegepast op een praktisch voorbeeld. Er zal een aantal stappen genomen moeten worden om tot een totaaloplossing

te komen van het genoemde probleem. We gaan uit van een situatie met n dimensies. Dit houdt in dat de detailgegevens bestaan uit een situatie waarbij alle n dimensies bekend zijn.

Stap 1: zoek alle subtotalen met 1 dimensie onbekend.

Bij elk van de subtotalen waar n-1 dimensies bekend zijn, zoeken we de bijbehorende detailgegevens. Dit zijn die detailgegevens waarbij de dimensies uit de details overeenkomen met die uit het sub totaal. Bepaal nu de verschillen tussen de subtotalen en de details. Voeg voor elk van deze verschillen een detailgegeven toe met daarin de n-1 overeenkomende dimensies gevuld en voor de laatste dimensie de waarde 'onbekend'. Het resultaat is een set met aangeleverde detailgegevens, aangevuld met in deze stap gecreëerde detailgegevens.

Stap 2: zoek alle subtotalen met 2 dimensies onbekend.

Bij elk van de subtotalen waar n-2 dimensies bekend zijn, zoeken we de bijbehorende detailgegevens. Hierbij worden ook de in stap 1 gecreëerde detailgegevens in ogenschouw genomen. Bepaal nu de verschillen tussen de subtotalen en de details. Voeg voor elk van deze verschillen een detailgegeven toe met daarin de n-2 overeenkomende dimensie gevuld en voor de laatste twee dimensies de waarde 'onbekend'. Het resultaat is een set met detailgegevens uit stap 1, aangevuld met in deze stap gecreëerde detailgegevens.

Stap 3 tot stap n-1.

Volg de hierboven beschreven procedure voor de subtotalen op elk niveau. Men moet hierbij rekening houden met de volgorde: eerst drie dimensies onbekend, dan vier dimensies onbekend etcetera.

Stap n.

In deze stap vergelijken we het totaal generaal met de detailgegevens zoals die zijn aangeleverd, aangevuld met de gecreëerde detailgegevens uit alle voorgaande stappen. Ook hier bepalen we

OLAP

OLAP staat voor On-Line Analytical Processing. Dit is natuurlijk geen definitie. Edgar F. Codd formuleerde in 1993 12 regels waaraan OLAP-databases zich dienen te houden. In 1995 werden deze regels samengevat met een nieuwe begrip: FASMI (Fast Analytical Shared Multidimensional Information).

	Nederland		België	
	zakelijk	particulier	zakelijk	particulier
pennen	Nederland zakelijk pennen	Nederland particulier pennen	België zakelijk pennen	België particulier pennen
sleutel- hangers	Nederland zakelijk sleutelhangers	Nederland particulier sleutelhangers	België zakelijk sleutelhangers	België particulier sleutelhangers

Afbeelding 1: De organisatiestructuur.

Persoon	Land	Markt	Product	Prognose
CEO	Totaal	Totaal	Totaal	95
productmanager pennen	Totaal	Totaal	pennen	60
productmanager sleutelhangers	Totaal	Totaal	sleutelhangers	55
landmanager Nederland	Nederland	Totaal	Totaal	80
landmanager België	België	Totaal	Totaal	10
marktmanager Nederland zakelijk	Nederland	zakelijk	Totaal	65
marktmanager Nederland zakelijk pennen	Nederland	zakelijk	pennen	50
marktmanager Nederland particulier	Nederland	particulier	Totaal	5
marktmanager België zakelijk	België	zakelijk	Totaal	2
marktmanager België particulier	België	particulier	Totaal	7

Afbeelding 2: Prognoses van de managers.

Land	Markt	pennen	sleutelhangers	Totaal
Nederland	zakelijk	50		65
	particulier			5
	Totaal			80
België	zakelijk			2
	particulier			7
	Totaal			10
Totaal	Totaal	60	55	95

Afbeelding 3: Prognoses van managers in dimensioneel model.

het verschil tussen het totaal generaal en al deze detailgegevens. Dit verschil levert een laatste detailgegeven op, waarbij alle dimensies de waarde 'onbekend' krijgen.

Het resultaat van de voorgaande stappen is een set met detailgegevens die optelbaar zijn en leiden tot de subtotalen die als basis zijn aangeleverd.

Voorbeeldorganisatie

We gaan uit van de volgende versimpelde situatie: een multinational is gevestigd in Nederland en België. Deze multinational verkoopt pennen en sleutelhangers. De kopers van de producten zijn in te delen in twee groepen: zakelijke en particuliere klanten. Om de multinational effectief aan te kunnen sturen zijn de volgende functies gedefinieerd:

- Per land (Nederland, België) is er een landenmanager benoemd;
- Per markt (zakelijk, particulier) is er een marktmanager benoemd. Elk land heeft zijn eigen marktmanagers;
- Per product (pennen, sleutelhangers) is er een productmanager benoemd. Productmanagers werken over de landsgrenzen heen.

De organisatiestructuur, feitelijk een matrixorganisatie, staat weergegeven in afbeelding 1. Naast de eerder genoemde managers is er speciaal voor de zakelijke markt pennen in Nederland nog een manager aangetrokken. Nu wordt elk van de managers

subtotalen stap 1

Land	Markt	Product	Prognose
Nederland	zakelijk	Totaal	65
Nederland	particulier	Totaal	5
België	zakelijk	Totaal	2
België	particulier	Totaal	7

Afbeelding 4: Subtotalen en details voor stap 1.

gevraagd een prognose voor het komende jaar te geven.

Daarnaast maakt de CEO zijn eigen overall-prognose voor het gehele bedrijf om aan de aandeelhouders te communiceren. Deze prognoses zijn weergegeven in afbeelding 2. Wanneer we deze prognoses weergeven in een dimensioneel model levert dat afbeelding 3 op. Een korte blik op de weergegeven prognoses leert ons dat deze duidelijk niet optelbaar zijn. We gaan de beschreven theoretische oplossing toepassen op dit voorbeeld.

Voorbeelduitwerking

Stap 1: zoek alle subtotalen met 1 dimensie onbekend.

Neem alle gegevens op het diepste detailniveau, oftewel alle gegevens waarbij alle dimensies bekend zijn. Neem alle gegevens waarbij slechts één van de dimensies de waarde 'Totaal' bevat. Dit zijn de subtotalen voor stap 1. We gaan deze vergelijken met het diepste detailniveau, zie afbeelding 4.

Nu gaan we vergelijken. Voor Nederland, zakelijk hebben we een totaal van 65. We hebben echter alleen details (product pennen) ter waarde van 50. We missen dus detailgegevens ter waarde van $(65 - 50 =) 15$. We creëren nu een detailgegeven voor Nederland, zakelijke markt, productgroep 'onbekend' ter waarde van 15. Let op dat we hier dus niet kunnen concluderen dat deze missende waarde van 15 bij de productgroep sleutelhangers zal horen! Voor Nederland particulier, België zakelijk en België particulier hebben we helemaal geen detailgegevens. Hiervoor maken we ook detailgegevens aan met de productgroep 'onbekend'. Het resultaat van deze stap staat weergegeven in afbeelding 5. Nu we alle subtotalen van stap 1 hebben gecontroleerd en voor de niet-kloppende subtotalen detailgegevens hebben gecreëerd, gaan we verder met stap 2.

Stap 2: zoek alle subtotalen met twee dimensies onbekend.

Dit zijn alle subtotalen die in twee dimensies 'Totaal' bevatten. Ook deze subtotalen gaan we weer vergelijken met de detail-

detailgegevens

Land	Markt	Product	Prognose
Nederland	zakelijk	pennen	50

Citaten over OLAP op www.kennisportal.com

“Het belangrijkste doel van OLAP-toepassingen is het eenvoudiger toegankelijk maken van informatie. Deze informatie dient relevant te zijn voor het nemen van beslissingen en moet daarom op het juiste moment en voor de juiste gebruiker beschikbaar zijn. OLAP-toepassingen worden voornamelijk door verkoop- en marketing-afdelingen gebruikt. Dit geeft aan waar OLAP-toepassingen met name geschikt voor zijn: het analyseren van verkoopgegevens per klantengroep, product, regio, enzovoort.”

“OLAP-systemen zijn minder geschikt voor het verzamelen van budgetteringsgegevens en het maken van de benodigde P&L-berekeningen. In principe bieden OLAP-systemen weinig of helemaal geen bedrijfseconomische calculatielogica. Hierdoor is het implementeren van een OLAP-systeem voor het budgetteringsproces een tijdsintensieve en dus kostbare zaak.”

detailgegevens

Land	Markt	Product	Prognose
Nederland	zakelijk	onbekend	15
Nederland	zakelijk	pennen	50
Nederland	particulier	onbekend	5
België	zakelijk	onbekend	2
België	particulier	onbekend	7

Afbeelding 5: Resultaat stap 1.

gegevens. In deze vergelijking nemen we ook de detailgegevens mee die we in stap 1 gecreëerd hebben, zie afbeelding 6.

Nu gaan we het totaal van Nederland vergelijken met de detailgegevens van Nederland. Hierbij nemen we bewust de in stap 1 gecreëerde detailgegevens mee. Voor Nederland zien we dat de details tezamen (15 + 50 + 5 =) 70 vormen. Om aan het totaal voor Nederland van 80 te komen missen we dus een detailgegeven ter grootte van 10. Van dit detailgegeven weten we niet

subtotalen stap 2

Land	Markt	Product	Prognose
Nederland	Totaal	Totaal	80
België	Totaal	Totaal	10

Afbeelding 6: Subtotalen en details voor stap 2 (deel 1).

subtotalen stap 2

Land	Markt	Product	Prognose
Totaal	Totaal	pennen	60
Totaal	Totaal	sleutelhangers	55

Afbeelding 7: Subtotalen en details voor stap 2 (deel 2).

detailgegevens

Land	Markt	Product	Prognose
Nederland	onbekend	onbekend	10
Nederland	zakelijk	onbekend	15
Nederland	zakelijk	pennen	50
Nederland	particulier	onbekend	5
België	onbekend	onbekend	1
België	zakelijk	onbekend	2
België	particulier	onbekend	7

Afbeelding 8: Resultaat stap 2.

bij welke markt of bij welk product dit thuishoort. Deze dimensies worden dus met 'onbekend' gevuld. Op dezelfde manier wordt er voor België een detailgegeven ter grootte van 1 gecreëerd.

Daarnaast zijn er nog meer subtotalen voor stap 2, waarbij zowel de land- als de marktdimensie 'Totaal' bevatten, zie afbeelding 7. Ook hier gaan we op basis van de detailgegevens uit de voorgaande stap de verschillen bepalen: we hebben detailgegevens voor pennen ter grootte van 50, terwijl het subtotale van pennen 60 bedraagt. Dit houdt in dat we detailgegevens toevoegen met land en markt 'onbekend'. Voor sleutelhangers hebben we geen detailgegevens. Daarom voegen we detailgegevens toe met land en markt 'onbekend' en product 'sleutelhangers'. Het resultaat van deze stap staat weergegeven in afbeelding 8.

Stap 3.

Dit is het moment om het totaal generaal te vergelijken met alle aangeleverde en gegenereerde detailgegevens, zie afbeelding 9. Nu worden we geconfronteerd met de situatie waarbij we detailgegevens hebben ter waarde van 155, terwijl het totaal generaal, de prognose van de CEO slechts 95 is. Dit houdt in dat er een detailgegeven ter waarde van -60 moet worden toegevoegd. Het resultaat van deze stap staat weergegeven in afbeelding 10.

Hoe nu verder?

Nu hebben we een hoop detailgegevens gecreëerd. Wanneer deze detailgegevens worden opgeteld via de dimensies zien we dat elke combinatie inderdaad precies overeenkomt met de opgegeven

detailgegevens

Land	Markt	Product	Prognose
Nederland	zakelijk	onbekend	15
Nederland	zakelijk	pennen	50
Nederland	particulier	onbekend	5
België	zakelijk	onbekend	2
België	particulier	onbekend	7

detailgegevens

Land	Markt	Product	Prognose
Nederland	zakelijk	pennen	50

Totaal stap 3

Land	Markt	Product	Prognose
Totaal	Totaal	Totaal	95

Afbeelding 9: Totaal en details stap 3.

Totaal stap 3

Land	Markt	Product	Prognose
onbekend	onbekend	onbekend	-60
onbekend	onbekend	pennen	10
Nederland	zakelijk	pennen	50
Nederland	onbekend	onbekend	10
Nederland	zakelijk	onbekend	15
Nederland	particulier	onbekend	5
België	onbekend	onbekend	1
België	zakelijk	onbekend	2
België	particulier	onbekend	7
onbekend	onbekend	sleutelhangers	55

Afbeelding 10: Resultaat stap 3.

prognoses. Wanneer we bijvoorbeeld willen kijken naar de totale prognose van pennen, dan selecteren we alle detailgegevens, waarbij het product gelijk is aan pennen. Nu kunnen we zien dat de totale prognose gelijk is aan $50 + 10 = 60$, waarbij we van de 50 weten dat die in de zakelijke markt in Nederland geprognosticeerd zijn en van de overige 10 weten we zowel het land als de markt niet. Dit blijkt nog duidelijker wanneer we de detailgegevens weergeven in het eerder aangehaalde dimensionale model, hierin zijn de opgegeven prognoses in vet weergegeven – zie afbeelding 11.

Is het opgelost? Ja en nee. Op de hier beschreven manier is het mogelijk om ook gegevens die niet optellen in een OLAP-kubus op te slaan. Zoals een oplettende lezer wellicht al heeft ontdekt lost deze aanpak ook de beperking van veel OLAP-tools op in het omgaan met zogenaamde 'unbalanced trees'. Maar deze aanpak

Land	Markt	pennen	sleutelhangers	onbekend	Totaal
Nederland	zakelijk	50		15	65
	particulier			5	5
	onbekend			10	10
	Totaal	50		30	80
België	zakelijk			2	2
	particulier			7	7
	onbekend			1	1
	Totaal			10	10
Onbekend	zakelijk				
	particulier				
	onbekend	10	55	-60	5
	Totaal	10	55	-60	5
Totaal	Totaal	60	55	-20	95

Afbeelding 11: Dimensionale weergave van detailgegevens.

detailgegevens

Land	Markt	Product	Prognose
onbekend	onbekend	pennen	10
Nederland	zakelijk	pennen	50
Nederland	onbekend	onbekend	10
Nederland	zakelijk	onbekend	15
Nederland	particulier	onbekend	5
België	onbekend	onbekend	1
België	zakelijk	onbekend	2
België	particulier	onbekend	7
onbekend	onbekend	sleutelhangers	55

heeft wel een prijs. De verwerkingstijd van het bepalen van alle afgeleide detailgegevens wordt langer, exponentieel met het aantal dimensies. In afbeelding 12 staat voor het aantal dimensies hoeveel subtotalen er op elk niveau kunnen voorkomen. Voor elk van deze subtotalen moet gecontroleerd worden of de detailgegevens consistent zijn met dit subtotaal en moeten eventueel detailgegevens worden toegevoegd.

Is deze oplossing dan alleen kleinschalig toepasbaar? Het antwoord luidt: nee. Bij geringe hoeveelheden gegevens kost elk van de controles nauwelijks tijd. Maar ook als de hoeveelheid gegevens groter wordt is deze methode nog prima toepasbaar. Door vooraf te controleren welke subtotalen daadwerkelijk in de gegevens aanwezig zijn, hoeven alleen deze subtotalen gecontroleerd te worden. Dit spaart enorm op de rekentijd. Men moet er dan wel op letten om de volgorde zoals in dit artikel aangegeven aan te houden, dus eerst alle subtotalen over één dimensie, vervolgens alle subtotalen over twee dimensies, enzovoort. Neem ook altijd de in een vorige stap gecreëerde detailgegevens mee.

aantal dimensies	3	4	5	10	n
totaal aantal subtotalen	7	15	31	1023	$2^n - 1$
aantal subtotalen in stap 1	3	4	5	10	
aantal subtotalen in stap 2	3	6	10	45	
aantal subtotalen in stap 3	1	4	10	120	
aantal subtotalen in stap 4		1	5	210	
aantal subtotalen in stap 5			1	252	
aantal subtotalen in stap 6				210	
aantal subtotalen in stap 7				120	
aantal subtotalen in stap 8				45	
aantal subtotalen in stap 9				10	
aantal subtotalen in stap 10				1	

Afbeelding 12: Aantal te controleren aggregaten.

Aanbeveling

Deze methode is natuurlijk niet alleen toe te passen op prognoses. Deze methode is ook toegepast op budgetten, het oplossen van 'unbalanced trees' in een organisatiestructuur, het verwerken van handmatige invoer waarbij zowel totalen als details uitgevraagd worden, enzovoort. Wees dus vooral creatief!

Bas Pruijn

Drs. B.C. Pruijn (bas.pruijn@ordina.nl) is Architect bij Ordina VisionWorks.