



Centrale administratie van stamgegevens is noodzaak

The single point of WHAT?

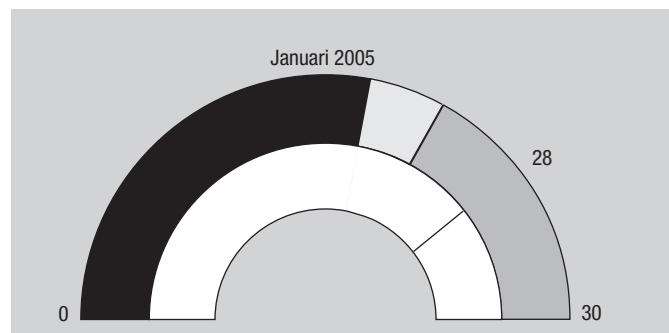
Karien Verhagen

Architecten besteden veel aandacht aan het 'Single Point', maar of daar nu ook de *waarheid* te vinden is? Uit de in dit artikel beschreven case blijkt dat je er nog niet bent met de identificatie van een leading system.

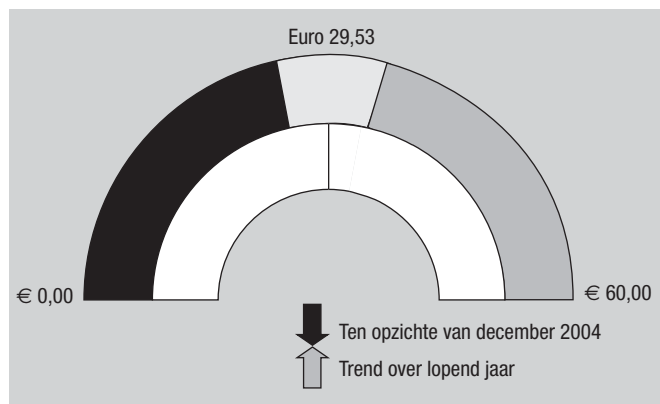
Implementaties van BI-systemen brengen de noodzaak voor centrale stamgegevens steeds duidelijker aan het licht. Centrale stamtabellen of master-data zijn tabellen met de actuele codes, omschrijvingen en attributen die vaak worden gebruikt als referentie, dus in een BI-systeem als dimensie. Ze bevatten bijvoorbeeld de adresgegevens van alle klanten, de indeling van alle artikelen, de regiocodes met hun omschrijving en dergelijke. Een bestand met referentiegegevens dat door veel systemen gebruikt wordt moet betrouwbaar zijn. In de applicatiearchitectuur worden daarom *leading systems* benoemd en gegevens eigenaren geïdentificeerd.

Onverwachte crisis

Het fictieve bedrijf dat in deze (hypothetische maar representatieve) case beschreven wordt, heeft eindelijk alle tegenslagen en weerstanden overwonnen. Je zou bijna kunnen spreken van een intelligente organisatie. Dan volgt toch nog de deceptie: de prachtige overzichten waar ze al jaren op sturen kloppen niet! Gartner noemt dit de 'reliability' crisis: de (tot nog toe) laatste crisis in een organisatie op weg naar Business Intelligence en die tevens de meest onverwachte is.



Afbeelding 1: Het aantal nieuwe klanten in de gerapporteerde periode.



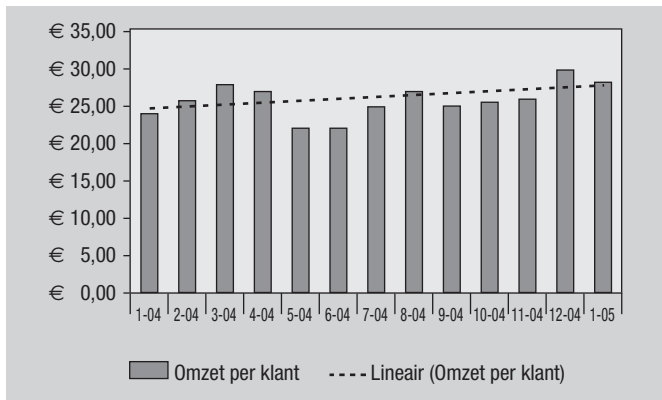
Afbeelding 2: De gemiddelde omzet per klant.

Het identificeren van een leading system is een stap vooruit ten opzichte van al de deeladministraties. Toch lijkt een waarschuwing op zijn plaats anders rekent men zich te snel rijk. Pas een gedegen gegevensaudit in combinatie met een ingericht organisatieonderdeel dat de waarheid bewaakt, kan betrouwbaarheid garanderen. Zo'n gegevensaudit is veel meer dan een papieren onderzoek naar de definities van velden en een onderzoek naar afwijkende veldwaarden. Om het gebruik van velden in systemen echt te doorgronden moet je letterlijk naast de gebruiker gaan staan, met twee voeten in de operatie. Dan wordt duidelijk hoe subtiele verschillen ontstaan tussen de oorspronkelijke opzet van het systeem en de praktische invulling.

De case is een bundeling van praktijkervaringen die qua ernst ongeveer het midden houdt tussen worst case en best case.

De case start met een directie die zich realiseert dat zij beleid 'op los zand bouwt'. Zij wil haar beslissingen met betrouwbare cijfers kunnen onderbouwen. De bedrijfsresultaten wil men vervolgens meten om te zien of de genomen beslissingen juist zijn.

De bedrijfsleiding stort zich op de kernwaarden, de strategie en de



Afbeelding 3: De ontwikkeling van de gemiddelde omzet per klant over het lopende jaar.

prestatie-indicatoren. Het wordt duidelijk dat men zich moet focussen op de productportefolio. Producten die niet verkocht worden, blijven te lang in het assortiment en goed verkopende producten die nieuwe klanten genereren, worden te laat als zodanig herkend. Ook is het klantenbestand te versnipperd: er zijn te veel kleine klanten, die in verhouding hoge kosten genereren. Er moet in absolute zin meer omzet per klant worden gerealiseerd. De prestatie-indicatoren krijgen vorm in aantal dashboard-elementen. Per gerapporteerde periode wil de directie het aantal nieuwe klanten zien (afbeelding 1). Over diezelfde periode wil ze de gemiddelde omzet weten per klant, zie afbeelding 2, het verloop van de gemiddelde omzet ten opzichte van de voorgaande periodes en de trend, zie afbeelding 3 (het betreft hier niet-seizoensgevoelige producten). Wat ook handig lijkt is een top-10 van de producten die de meeste nieuwe klanten genereren (afbeelding 4). Tot slot wil de directie een lijst met alle artikelen, aflopend gesorteerd op lengte van de termijn dat ze niet zijn verkocht, zie afbeelding 5.

De gegevenshuishouding is in orde

In dit bedrijf is de gegevenshuishouding niet ingewikkeld en dik in orde. Er zijn slechts twee afdelingen die bovendien dezelfde producten voeren. Voor de klantgegevens is de debiteurenadministratie leidend. Een factuur mag de deur pas uit als bij de centrale administratie een debiteurnummer is aangevraagd. De decentrale

Ranking op afzet			Ranking op omzet		
Artikel	Afzet	Omzet	Artikel	Afzet	Omzet
G 7845	6574	€ 6.475,34	P 0987	4333	€ 9.078,00
K 8976	5346	€ 2.345,32	M 0098	677	€ 6.734,90
P 0987	4333	€ 9.078,00	N 8976	322	€ 6.723,89
M 0098	677	€ 6.734,90	G 7845	6574	€ 6.475,34
J 8956	544	€ 563,89	J 7845	23	€ 5.634,45
V 9987	342	€ 4.523,08	V 9987	342	€ 4.523,08
N 8976	322	€ 6.723,89	K 8976	5346	€ 2.345,32
M 9853	122	€ 345,56	G 6734	12	€ 1.876,65
O 0970	89	€ 178,60	J 8956	544	€ 563,89
L 0845	54	€ 356,90	L 0845	54	€ 356,90

Afbeelding 4: Ranking artikelen verkocht aan nieuwe klanten in januari 2005.

afdelingen leveren dan hun adresgegevens. De centrale administratie voert ze in en de nieuwe of gewijzigde debiteur staat veilig in het leading system. Het debiteurenbestand bevat circa twaalfduizend nummers. De decentrale administraties van de afdelingen A en B bevatten respectievelijk zesduizend en achtduizend debiteuren, subsets van het centrale debiteurenbestand. De decentrale afdelingen factureren zelf en periodiek boeken ze de omzet per artikelgroep en per debiteur naar de centrale administratie door. Op die overboeking zit een valide controle, geen enkele omzetregel van de decentrale afdelingen verdwijnt. De controllers hebben solide vierkantstellingen geëist en gekregen.

Het 'single point of truth' is gevonden. De gevraagde overzichten zijn niet moeilijk, er zijn eerder honderden dan tientallen tools die ze kunnen produceren. Een BI-deskundige heeft in een paar uurtjes met een extract van de live data al een opzet voor een proto-type. Over de vorm, de verspreiding van het detailniveau en de analysemogelijkheden wordt men het eens. Wanneer ook de software-installaties, toegangen en periodieke scripts veilig geïmplementeerd zijn, heeft de directie de inzichten waarom ze vroeg. Alles lijkt in orde, totdat men besluit om een gegevensaudit te doen. Die was aanvankelijk wegbezuindigd; hij lijkt immers overbodig. Een flinke tijd na de implementatie besluit de directie toch nog om er een te doen. Safety first, een voorgevoel?

Gegevensaudit en kwaliteit adressenbestand

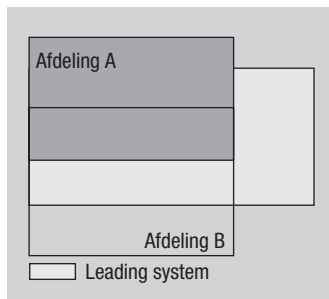
Dan volgt de ontuchtering: van de twaalfduizend klanten blijken er drieduizend zowel bij afdeling A en B als bij de debiteurenadministratie voor te komen met dezelfde adresgegevens.

Tweeduizend klanten komen alleen voor bij de afdeling A en de debiteurenadministratie en drieduizend bij de afdeling B en de debiteurenadministratie. Na enkele cleansing en vergelijkings-exercities hebben we een betrouwbaar klantenbestand van achtduizend klanten. Drieduizend zijn klant bij beide afdelingen, tweeduizend alleen bij A en drieduizend alleen bij afdeling B. De vraag wordt gesteld of er dus geen twaalfduizend maar achtduizend klanten zijn.

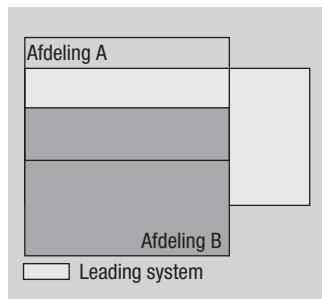
Wat is er aan de hand met de klanten, die wel degelijk omzet genereren maar ondanks de waterdichte laadprocedure geen adres-match in het debiteurenbestand hebben? Voor het grootste

Artikel	Omschrijving	Maanden
P 9860	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	24
M 0894	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	24
K 8670	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	23
H 0234	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	22
G 7007	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	22
H 0564	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	18
H 8745	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	15
G 7850	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	10
G 0673	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	7
L 9845	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	6

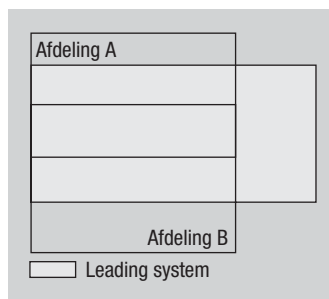
Afbeelding 5: Artikelen gesorteerd op de termijn dat ze niet verkocht zijn.



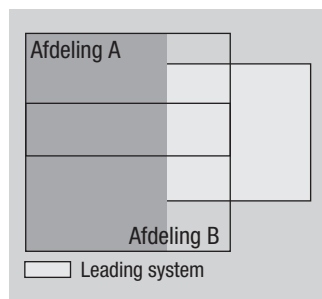
Afbeelding 6: Klantenbestand afdeling A.



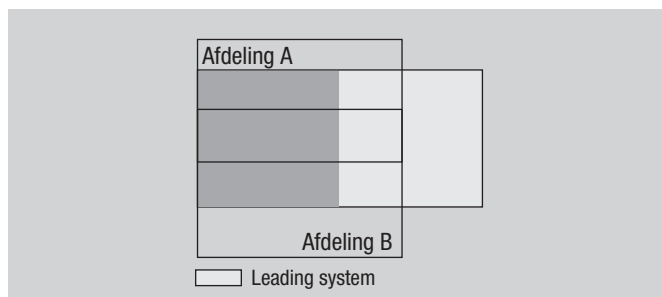
Afbeelding 7: Klantenbestand afdeling B.



Afbeelding 8: Het klantenbestand in de Single point of Truth-oplossing.



Afbeelding 9: De 'echte' waarheid; Klantenbestand na de audit.



Afbeelding 10: Overlap tussen de Single point of Truth-oplossing en de waarheid.

deel blijkt het hier te gaan om oude, niet meer gebruikte debiteurnummers uit de deeladministraties die aan nieuwe klanten zijn toegekend. De procedure om een nieuw debiteurnummer aan te vragen, blijkt uit gesprekken met deeladministrateurs, duurt soms lang. Als de facturen de deur uit moeten lijkt deze oplossing snel en acceptabel. Het komt de consistentie in de adressenbestanden niet ten goede, maar de snelle afhandeling van de facturatie weegt zwaarder. Daarop immers wordt de medewerker door zijn directe omgeving afgerekend. Het gevolg is echter dat de omzet wordt geboekt op het verkeerde account, de verkeerde regio enzovoort.

Een ander deel van de 'foute' debiteurnummers betreft kleine, incidentele klanten die in de laadprocedure naar het centraal systeem geen matching debiteurnummer vinden. Om de vierkantstellingen van de controllers kloppend te maken worden ze geboekt als 'diverse kleine debiteuren' of op een soort verzamelnummer in het grootboek. De afspraak is dat de

afwijking later wordt getraceerd en gecorrigeerd, maar die afspraak blijkt genegeerd.

Dan zijn er nog debiteuren in de centrale administratie die wel artikelen afnemen maar blijkbaar niet via de afdelingen A en B. Dat blijken werknemers en ex-werknemers te zijn. Het is in de praktijk verboden producten rechtstreeks te verkopen aan werknemers, echter de directie heeft een vervelend precedent geschapen. De procedure wordt aangescherpt en deze debiteurnummers worden verwijderd. Verder zijn er nog klanten die zo lang niets gekocht hebben dat ze in de decentrale systemen zijn verwijderd. Ook deze klanten worden naar de historie verwezen en uit het actuele systeem gehaald.

Het klantenbestand minus de debiteuren/niet-klanten en de rustende klanten, plus de niet centraal bekende klanten uit A en B, bestaat nu uit 8250 echte klanten, zie afbeeldingen 6 tot en met 10. De overlap met de aanvankelijk gebouwde en betrouwbaar geachte single-point-of-truth oplossing is slechts vijftig procent! Wat zegt dat over de betrouwbaarheid van de uitspraak 'dit is een nieuwe klant', en hoe betrouwbaar is nog de uitspraak 'de gemiddelde omzet per klant is vijfennegentig euro'? Als dan ook nog blijkt dat het betrouwbaar geachte artikelbestand rammelt is er een crisis. Afdeling A en B verkopen weliswaar dezelfde artikelen, maar in de decentrale administratie blijkt dat afdeling A twee verpakkingen met ieder vijftien stuks van een artikel als twee stuks verkoopt. Afdeling B heeft het in dat geval over dertig stuks. Het bedrijf zou dus de afzet kunnen vervijftienvoudigen door de klanten alles bij afdeling B te laten kopen.

Conclusies

In de case is sprake van slechts twee decentrale afdelingen die dezelfde artikelen voeren. In de praktijk liggen de zaken meestal ingewikkelder. Elke nieuwe deeladministratie en elke prestatie-indicator die gebruik maakt van meer dan een soort stamgegevens zal het probleem exponentieel vergroten. Zo stapelt zich fout op fout.

Een gegevensaudit toont vaak aan dat waterdicht geachte controles lek zijn en de harde waarheden boterzacht. Daarbij spelen ook belangenconflicten een rol. Het is moeilijk om vervuiling in stamgegevens te voorkomen, omdat de decentrale afdelingen andere prioriteiten hebben dan de directie. Zij hebben geen belang bij het betrouwbaar houden van de gedeelde stamgegevens: ze moeten verkopen. De directie heeft dat belang wel. Wil zij over betrouwbare geïntegreerde bedrijfsbrede stuurinformatie beschikken dan moet ze betalen voor echt master-datamanagement.

Dat is een ingerichte centrale administratie van stamgegevens, bewaakt door geïdentificeerde gegevenseigenaren die zich door middel van Service Level Agreements (SLA's) verplichten kwaliteit te leveren. Het is niet voldoende een 'Single Point' te identificeren, een verbod op deeladministraties is vaak noodzakelijk. Pas dan zal op dit 'single point' ook 'the truth' gevonden kunnen worden.

Karien Verhagen

Drs. Karien Verhagen is directeur van 4BIS, Scholing en Advies.