

Een business case zorgt voor de juiste afweging

Wat is datakwaliteit waard?

Nico Klaassen

In de vorige artikelen is aangegeven dat de datakwaliteit nog wel eens te wensen over laat. Sterker; datakwaliteit kan (nog) beter. In dit laatste artikel wordt behandeld wat datakwaliteit nu waard is voor een organisatie en hoe dit om te zetten in een business case.

Doel van zo'n business case is om de mogelijkheden van datakwaliteit-meting en -schoning aan te tonen en vooral om inzicht te krijgen in kosten, baten en risico's. Hierbij wordt minder gekeken naar de details rondom het hoe.

Business case

De business case is erop gericht om het management te overtuigen van nut en prijs. Voor vele technisch ingestelde medewerkers die verantwoordelijk voor de data worden geacht lijkt dit niet zo'n interessant issue, maar uiteindelijk moeten ook deze technici hun oordeel geven over analyse en schoning. Hun bijdrage is vaak: kan het wel gemeten worden; kunnen we er iets aan doen, en zo ja wat heb je ervoor nodig.

"Verantwoordelijk geacht voor de kwaliteit van de data", het is toch de eindgebruiker die alles invoert. De beleving door het management is vaak anders, want de Database Administrator is voor hen toch de persoon die over de data gaat.

Door onderzoek bij diverse organisaties naar de kwaliteit van gegevens, zijn allerlei problemen ontdekt, zoals gegevens die niet conform de definities zijn, niet plausibel en niet consistent zijn op basis van de opgegeven business rules. Nadat de problemen in kaart waren gebracht, is gekeken naar de mogelijke oplossingen en de kosten van schoning. Uiteindelijk zijn ook de risico's in kaart gebracht van de oplossing. Alles bij elkaar vormt dit de basis voor een business case.

De opbrengsten voor de organisatie in kaart brengen is vaak een complexer taak dan gedacht. Het management wil geconfronteerd worden met feiten. En hoe breng je nu het risico van problemen en imagoschade aan het licht.

Enkele voorbeelden

Een reisbureau wil een brochure naar potentiële klanten sturen. Klanten kunnen via internet een formulier invullen. Het matchen van gegevens om te checken of deze klant al eerder is inge-

schreven (of in een ander systeem bij het reisbureau zit vanwege eerdere boeking) gebeurt echter niet. Dit leidt ertoe dat regelmatig klanten dubbel vermeld zijn in het adressenbestand. Het onderzoek naar dubbele vermeldingen (combinaties van genormaliseerde naam en adres, postcode en geboortedatum) toont een vervuiling van 20 procent op ongeveer 100.000 klanten. De feitelijke klanten-database bestaat dus maar uit 80.000 relaties. Bovendien bevatte de klantendatabase niet alleen dubbele gegevens, maar ook niet adresseerbare gegevens (bijvoorbeeld verzonden adressen).

Bij het opstellen van de business case moet men uitgaan van de kosten per te verzenden brochure, dit bedraagt € 7,50 per brochure inclusief portokosten. Door het niet versturen van brochures naar deze dubbel vermelde en niet adresseerbare klanten, bespaart het reisbureau dus $7,50 \times 20.000 = € 150.000$. Daarnaast kan eventueel ook het imago een rol spelen. Potentiële klanten ontvangen niet twee keer de brochure, dat komt het beeld van de efficiency van het reisbureau ten goede.

De kosten van deze schoning kunnen worden ingeschat op ongeveer € 40.000 aan externe consultancy en tooling. De organisatie echter heeft hier nog eens 1,5 maand aan interne schoning moeten besteden (onderzoeken en ontdebellen). De uiteindelijke besparing bedraagt dus bijna € 110.000.

Uitgaande van meerdere brochures of mailings per jaar levert dit een nog hogere besparing op.

Soms echter zijn de baten niet altijd even hoog, maar geeft het een duidelijk beeld van hoe men als organisatie werkt. Een advocatenkantoor staat op het punt een datamigratie uit te voeren naar

Wat is datakwaliteit?

In een serie van drie artikelen behandelt Nico Klaassen een aantal vragen over datakwaliteit. In het eerste artikel (DB/M 7/2004) ging het om de meetbaarheid van datakwaliteit. In het tweede artikel (DB/M 1/2005) stond de vraag centraal of datakwaliteit nog beter kan. Wat kost een kwaliteitsmeting en wat levert het op? Die vragen worden in dit derde en laatste artikel behandeld.

een nieuw systeem. Bij de analyse van de gegevens worden ook open staande betalingen in kaart gebracht. Hierbij blijkt dat er diverse betalingen open staan, met als uiterste een betaaldatum in 2072. De business case tot schoning was in dit geval wat complexer. De klant moet omgaan met een factuur waarop hij al niet meer rekende. In dit geval is het corrigeren niet alleen een kwestie van wat het kost, het imago is van hoger belang dan de factuur.

Binnen het nieuwe systeem zijn door de klant extra controles gedefinieerd om vervuiling van gegevens tegen te gaan, zoals controles op het invoeren van gegevens met name rond de facturering. Daarnaast zijn ook aanvullende gegevens als het gebruik van de postcodetabel ter verificatie van adressen ingevoerd. Een fout die in het verleden is gemaakt, welke inzichtelijk maakt wat 'ondeugdelijke controles' voor gevolg kunnen hebben, leidt dus tot een beter systeem.

Vele factoren

Een goede business case geeft de diverse factoren weer die van belang zijn voor de organisatie om te besluiten een schoning-project wel of niet te starten. Hiermee kan een juiste afweging gemaakt worden. Er zijn al vele factoren in de serie genoemd, hieronder volgt een overzichtelijk resumé.

Kosten:

- Analyse (tooling-kosten, kosten extractie van data, analyse van data);
- Schoning (personele kosten, schoning programmatuur);
- Beheersing in toekomst (systeemaanpassingen, procesaanpassingen).

Baten:

- Externe factoren (imago, directe kostenbesparing, scherpere prijsstelling, wettelijke en juridische noodzaak);
- Interne factoren (minder fouten, efficiënter werken, eenduidige definities maken vergelijken mogelijk, voordelen bij migraties).

Risico:

- Schonen (wat zijn de gevolgen van foutief schonen, behouden van consistentie in de gegevensverzameling, wat zijn de business-effecten van schonen);
- Controleerbaarheid (hoe toon men beheerst schonen aan).

De ervaring leert dat er enige scepsis heerst bij aanvang van een analyse- en schoningproject. Na een (vaak korte) technische analyse wordt de scepsis vaak al weggenomen. Mogelijke problemen komen aan het licht en men ziet wat er fout gaat of zou kunnen gaan.

Met een goede business case en een goede afweging van kosten, baten en risico's kan het management worden geholpen om de juiste besluiten voor het schonen te nemen. Dit is dan tevens een goede basis voor vervolgstappen om problemen op te lossen en te voorkomen.

Conclusie

Uit de drie artikelen in het kader van datakwaliteit en verhoging ervan, is af te leiden dat bij veel organisaties datakwaliteit te wensen over laat. Er is wat aan te doen: problemen met data zijn te achterhalen, zowel met als zonder dure tools en ook in veel gevallen is slechte kwaliteit te verbeteren. Het uiteindelijke resultaat van data-analyse en datakwaliteit-verhoging is meer dan alleen een betere kwaliteit tegen een bepaalde kostprijs. De belangrijke conclusie is echter om de 'routekaart' voor succesvol schonen in acht te houden: eerst denken, dan meten, nog eens denken en dan pas doen. Het 'waarom' van de schoning-acties moet centraal staan. Zonder deze routekaart is schonen een gevaarlijke en onverantwoorde activiteit waardoor mogelijk veel schade wordt aangericht. Met weloverwogen stappen kan schonen voor vele partijen binnen een organisatie een activiteit zijn die vruchten kan afwerpen.

Nico Klaassen

Ing. Nico Klaassen (nico.klaassen@capgemini.com) is Managing Consultant bij Capgemini.