



Leerboek Business Intelligence

PETER TER BRAAKE

+ ONLINE
MATERIAAL
VOOR
STUDENTEN
EN DOCENTEN



Boom

Tweede druk

Leerboek Business Intelligence

Leerboek Business Intelligence

Peter ter Braake

Boom

**inclusief
website!**

Met behulp van onderstaande unieke activeringscode kunt u toegang krijgen tot **www.leerboekbusinessintelligence.nl** voor extra materiaal. Deze code is persoonsgebonden en gekoppeld aan deze tweede druk. Na activering van de code is de website twee jaar toegankelijk. De code kan tot zes maanden na het verschijnen van een volgende druk geactiveerd worden.

© 2018 Boom uitgevers Amsterdam

Opmaak: Nu-Nique grafische vormgeving, Goor
Omslagontwerp: Carlito's Design, Amsterdam; Studio Bassa, Culemborg

ISBN: 9789024419258 (paperback)
ISBN: 9789024419265 (e-book)
NUR: 123/982

Alle rechten voorbehouden. Alle intellectuele eigendomsrechten, zoals auteurs- en databankrechten, ten aanzien van deze uitgave worden uitdrukkelijk voorbehouden. Deze rechten berusten bij Boom uitgevers Amsterdam en de auteur.

Behoudens de in of krachtens de Auteurswet gestelde uitzonderingen, mag niets uit deze uitgave worden veelevoudigd, opgeslagen in een geautomatiseerd gegevensbestand of openbaar gemaakt in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voor zover het maken van reprografische veelevoudigingen uit deze uitgave is toegestaan op grond van artikel 16 h Auteurswet, dient men de daarvoor wettelijk verschuldigde vergoedingen te voldoen aan de Stichting Reprorecht (Postbus 3051, 2130 KB Hoofddorp, www.reprorecht.nl). Voor het overnemen van gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16 Auteurswet) dient men zich te wenden tot de Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, Postbus 3060, 2130 KB Hoofddorp, www.cedar.nl/pro). Voor het overnemen van een gedeelte van deze uitgave ten behoeve van commerciële doeleinden dient men zich te wenden tot de uitgever.

Hoewel aan de totstandkoming van deze uitgave de uiterste zorg is besteed, kan voor de afwezigheid van eventuele (druk)fouten en onvolledigheden niet worden ingestaan en aanvaarden de auteur(s), redacteur(en) en uitgever deswege geen aansprakelijkheid voor de gevolgen van eventueel voorkomende fouten en onvolledigheden.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the publisher's prior consent.

While every effort has been made to ensure the reliability of the information presented in this publication, Boom uitgevers Amsterdam neither guarantees the accuracy of the data contained herein nor accepts responsibility for errors or omissions or their consequences.

Voorwoord

Voor je ligt de tweede druk van *Leerboek Business Intelligence*. Business Intelligence is een zeer populair vakgebied. Jarenlang hebben organisaties gegevens verzameld in databases. Het zinvol gebruiken van deze gegevens was echter vaak een achtergebleven gebied. In al die gegevens zit evenwel een schat aan informatie die organisaties tegenwoordig graag zouden willen gebruiken. Het doel is er beter van te worden.

Nieuwe ontwikkelingen rond Big Data en Advanced Analytics maken Business Intelligence extra interessant. Met goede toepassing ervan kan een organisatie beter inspelen op de huidige markt. Of wellicht kan er wel ergens een voordeel behaald worden ten opzichte van concurrenten. Deze tweede druk belicht deze zaken uitgebreider dan de eerste druk.

In dit boek bespreken we Business Intelligence van A tot Z. We beginnen met het definiëren van het begrip. Wat is het en waarom zou je het moeten toepassen? Wat zijn de doelen?

In de volgende hoofdstukken komen alle onderdelen van Business Intelligence aan bod. Wat is een datawarehouse? Hebben we dat nodig? Hoe vullen we een datawarehouse met gegevens en waar moeten we dan rekening mee houden? Wie gebruiken het datawarehouse en hoe doen ze dat? Vervangt Big Data het klassieke datawarehouse? Wat brengt Advanced Analytics ons? Deze uitgave geeft antwoord op deze en nog veel meer vragen.



Om de materie te verduidelijken worden enige voorbeelden uitgewerkt. Deze paragrafen kun je lezen op de website bij dit boek: www.leerboekbusinessintelligence.nl. De daarin gebruikte bestanden en scripts kun je daar ook downloaden. Op deze website kun je bovendien je opgedane kennis toetsen door het maken van meerkeuzevragen.

Het *Leerboek Business Intelligence* is geschreven voor studenten die in aanraking gaan komen met Business Intelligence, dus voor de bedrijfskundige studies, via bedrijfskundige informatica tot informatica. Omdat het boek echter zowel op technische als niet-technische lezers is gericht, is het zeker ook geschikt voor mensen uit het bedrijfsleven die een beter begrip willen krijgen van wat Business Intelligence nu precies is.

Veel plezier met lezen!

Peter ter Braake
Januari 2018

Inhoud

Voorwoord	v
1 Inleiding	1
1.1 Wat is Business Intelligence?	1
1.1.1 Definitie	1
1.1.2 Van wie is Business Intelligence?	3
1.1.3 Voor wie is Business Intelligence?	4
1.1.4 Self-service BI	6
1.1.5 Waarom Business Intelligence?	9
1.1.6 Conclusie	10
1.2 Andere terminologie en definities	10
1.2.1 Datawarehouse	10
1.2.2 Extract, Transform, Load	11
1.2.3 Staging, ODS	12
1.2.4 Kubussen, modellen	13
1.2.5 Datamining en machine learning	14
1.2.6 Big Data	15
1.3 Tot slot	16
1.3.1 Verder	17
2 Business Intelligence in de organisatie	19
2.1 Business Intelligence als competentie	19
2.1.1 Business Intelligence-volwassenheid	19
2.1.2 De Business Intelligence Life Cycle	22
2.1.3 Business Intelligence: doelen en ambitie	25
2.1.4 Business Intelligence-implementatie	27
2.1.5 Business Intelligence scope	29
2.1.6 Business Intelligence-doelen	31
2.1.7 Informatiebehoeften	32
2.1.8 Tot slot	36
2.2 Business Intelligence-projecten	37
2.2.1 Geen gewoon IT-project	37
2.2.2 Informatieanalyse	42
2.3 Tot slot	43
3 Waarom een datawarehouse?	45
3.1 Het datawarehouse	45
3.1.1 Waarom een datawarehouse	46
3.2 Performance van rapporten	46
3.2.1 Normaliseren	47
3.2.2 Indexen	49

3.3	Schrijven van query's	52
3.3.1	Wie maakt de rapporten?	53
3.4	Rapportage-impact op het primaire proces	53
3.4.1	Resource-gebruik	54
3.4.2	Concurrency	55
3.5	Kwaliteitsproblemen	57
3.5.1	Dubbele gegevens	57
3.5.2	Ontbrekende gegevens	58
3.5.3	Foutieve gegevens	60
3.5.4	Inconsistente gegevens	60
3.6	Verschillende operationele systemen	61
3.7	Historische gegevens	63
3.8	Conclusie	64
4	Het datawarehouse	65
4.1	Dimensioneel modelleren	65
4.1.1	Inleiding	65
4.1.2	Modelleren	67
4.1.3	Dimensies	72
4.1.4	Slowly Changing Dimensions (SCD)	80
4.1.5	Conformed dimensions en snowflakes	83
4.1.6	Overige dimensie-overwegingen	85
4.1.7	Feiten	87
4.1.8	Tot slot	90
4.2	Bill Inmon	90
4.2.1	Kritiek op Kimball	91
4.2.2	De Inmon-methodiek	92
4.2.3	Kritiek op Inmon	94
4.2.4	Samenvatting	95
4.3	Data Vault	95
4.3.1	Definitie	95
4.3.2	Hubs, Links and Satellites	96
4.3.3	Voordelen Data Vault	99
4.3.4	Samenvatting	100
4.4	Kimball, Inmon en Data Vault vergeleken	100
4.5	Tot slot	101
5	Het fysieke datawarehouse	103
5.1	Inleiding	103
5.2	Technisch ontwerp	104
5.2.1	Aantal records	105
5.2.2	Gemiddelde recordlengte	106
5.2.3	Page size	108
5.2.4	Grootte van de database	110
5.2.5	Overige factoren	111
5.2.6	Grootte van je oplossing	112

5.3	Hardware en software	113
5.3.1	CPU-capaciteit	114
5.3.2	Overige hardwarespecificaties	114
5.3.3	Appliances	116
5.4	Performance features	117
5.5	Tot slot	117
6	ETL	119
6.1	Inleiding	119
6.2	Master Data Management	120
6.2.1	Verschillende soorten gegevens	121
6.2.2	Definities	122
6.2.3	Implementatie	124
6.2.4	Voorbeeld met Microsoft MDS	128
6.2.5	Tot slot	128
6.3	Implementeren van het ETL-proces	128
6.3.1	Resultaten van informatieanalyse	128
6.3.2	Architectuur	131
6.3.3	Documenteren	135
6.3.4	Tools en technieken	136
6.4	Datakwaliteit	137
6.4.1	Slechte gegevens	137
6.4.2	Data Cleansing	138
6.4.3	Voorbeeld met Microsoft DQS	138
6.5	Scenario's	139
6.6	Tot slot	139
7	Big Data en Advanced Analytics	141
7.1	Introductie Big Data	141
7.1.1	Beperkingen van klassieke datawarehouses	143
7.2	Definitie van Big Data	144
7.2.1	Volume	145
7.2.2	Variety	146
7.2.3	Velocity	148
7.2.4	Andere kenmerken	148
7.2.5	Tot slot	149
7.3	Advanced Analytics	149
7.3.1	Inleiding Data Analytics	149
7.3.2	Basic Analytics	150
7.3.3	Advanced Analytics	151
7.3.4	Top-down versus bottom-up	152
7.3.5	Tot slot	154
7.4	Machine learning en datamining	154
7.4.1	Definities	155
7.4.2	Machine learning-scenario's	156
7.4.3	Tot slot	159

7.5	Big Data Tools - Data Lake - noSQL	160
7.5.1	Data Lake	160
7.5.2	noSQL	162
7.5.3	Big Data-databases	163
7.6	Big Data en datawarehousing	164
7.7	Big Data en ethiek	166
7.7.1	Hoe wenselijk is Big Data Analytics?	166
7.7.2	Wet- en regelgeving	167
7.7.3	Tot slot	169
7.8	Voorbeeld van datamining met Excel	169
7.9	Tot Slot	169
8	Semantische modellen	171
8.1	Inleiding	171
8.1.1	Semantisch model	171
8.1.2	Performance	174
8.1.3	Query-performance en Big Data	176
8.1.4	OLAP query engine	176
8.1.5	Tot slot	177
8.2	OLAP-kubus	177
8.2.1	Wat is een kubus?	177
8.2.2	Waarom kubussen?	179
8.2.3	Aggregaties	182
8.2.4	Model of database?	184
8.2.5	Tot slot	185
8.3	Voorbeeld in-memorymodel	186
8.4	Tot slot	186
9	Het front-end	189
9.1	Inleiding	189
9.2	Wie maakt en wie gebruikt rapporten?	189
9.2.1	Farmers	190
9.2.2	Tourists	192
9.2.3	Explorers	194
9.2.4	Miners	195
9.2.5	Tools en acceptatie	195
9.3	Rapporten, scorecards en dashboards	196
9.3.1	KPI	196
9.3.2	Scorecard	198
9.3.3	Dashboard	200
9.3.4	Rapporten	202
9.3.5	Security	203

9.4	Ontwerpen van (goede) rapporten	204
9.4.1	Inleiding rapportontwerp	204
9.4.2	Attentieve verwerking	206
9.4.3	Rapportrichtlijnen	210
9.4.4	Tabel versus grafiek	218
9.5	Power View	219
9.6	Tot Slot	219
	Index	221

Hoofdstuk 1

Inleiding

Business Intelligence is een veel gebruikte term die niet altijd hetzelfde betekent. Dit komt omdat het niet door iedereen op dezelfde manier wordt gebruikt. Voordat je verderop in dit boek alle onderdelen die een rol spelen bij Business Intelligence leert kennen, krijg je met dit hoofdstuk eerst een goed beeld bij het begrip Business Intelligence en bij enkele andere veel gebruikte begrippen. Ook leer je waarom Business Intelligence voor bijna alle bedrijven belangrijk is of belangrijk zou moeten zijn.

1.1 Wat is Business Intelligence?

Voor je ligt een leerboek over Business Intelligence, vaak afgekort tot BI. Voor het gemak ga ik er even van uit dat je al een klein beetje een idee hebt wat Business Intelligence inhoudt. Het doel van dit boek is dat idee te concretiseren, te verbreden en te verdiepen. Voor een belangrijk deel wil ik dat doen door concrete zaken te belichten. Theorie en praktijk moeten hand in hand gaan. Maar voordat we echt beginnen: neem eens de tijd om voor jezelf te benoemen wat Business Intelligence nu eigenlijk inhoudt. Probeer eens een voor jou werkbare definitie te formuleren.

1.1.1 Definitie

Volgens Gartner staat Business Intelligence voor *'an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance'* (gartner.com/it-glossary/business-intelligence-bi). Dat is meteen een hele mond vol. Vrij vertaald luidt deze definitie:

Business Intelligence is een overkoepelende term waarmee applicaties, infrastructuur en hulpmiddelen en aangeraden werkwijzen worden bedoeld, die als doel hebben om gegevens beschikbaar te stellen ten einde de juiste beslissingen te kunnen nemen op basis van correcte, betrouwbare informatie.

Business Intelligence is een soort paraplueterm geworden waar heel veel onder kan vallen. Het maken van rapportages om inzicht te krijgen in de kostenstructuur van een organisatie valt al onder Business Intelligence. Maar zo beschouwd doet elk bedrijf aan Business Intelligence omdat in elke organisatie wel rapporten gebruikt worden. De belangrijkste woorden in bovenstaande definitie zijn:

1. Informatie (information)
2. Beslissingen (decisions)

Business Intelligence draait om het beter in staat zijn beslissingen te nemen op basis van informatie. Het Engelse Intelligence betekent in de context van Business Intelligence niet iets als 'intelligentie' maar simpelweg 'informatie', net zoals de I in CIA.

Volgens Wikipedia staat Business Intelligence voor het verzamelen van gegevens binnen de eigen handelsactiviteit. Het woord 'gegevens' is ook een onderdeel van de definitie, maar staat niet bij de twee benoemde kernwoorden. Op Wikipedia valt verder dan ook te lezen:

'Business Intelligence kan omschreven worden als het proces van gegevens omzetten in informatie, dat vervolgens zou moeten leiden tot kennis en het aanzetten tot adequate actie.'

Een belangrijke toevoeging van deze omschrijving ten opzichte van de definitie van Gartner is dat hier gesproken wordt over het omzetten van gegevens, kleine brokjes data zoals gegevens van klanten en de orders die zij bestellen, naar informatie (wat betekenen die gegevens eigenlijk en hoe kunnen ze mij verder helpen?). Dat klinkt eenvoudig, maar hoe worden gegevens informatie?

Making Better Decisions Faster

Jaren geleden las ik een boek over Business Intelligence dat als ondertitel had: Making Better Decisions Faster. Als korte heldere werkdefinitie is dat misschien wel de beste omschrijving die ik tot nu toe ben tegengekomen. Business Intelligence draait er om de mensen in een organisatie beter in staat te stellen hun werk te doen. Een beetje afhankelijk van je precieze functie, hoort het nemen van besluiten bij je werk. Mensen zijn beter in staat die besluiten te nemen als ze goed geïnformeerd zijn. Bovendien moet die informatie tot hun beschikking staan op het moment dat ze het besluit moeten nemen. En dat is heel wat anders dan te zeggen dat ze het besluit moeten nemen pas als ze de informatie tot hun beschikking hebben. Als het enkele weken duurt voordat een andere afdeling binnen het bedrijf die informatie heeft verzameld, kan het wel al eens te laat zijn.

Het doel van Business Intelligence is:

■ Op het juiste moment op basis van de juiste informatie beslissingen nemen.

Voorbeeld

Denk bijvoorbeeld aan een winkelier die in een nieuw winkelcentrum een winkelpand kan huren voor een nieuwe vestiging. Gezien de investering die dat vergt, is dat een grote beslissing met mogelijk veel impact. De winkelier denkt zijn klanten te kennen: wat voor soort mensen kopen wat voor soort producten? Welke marge valt er met die producten te halen en hoe lang is er dan nodig om de initiële investering terug te verdienen?

De winkelier zou er erg bij gebaat zijn als hij zijn ervaring (hij denkt zijn klanten te kennen) zou kunnen staven aan beschikbare gegevens. Klanten krijgen korting met een klantenkaart en hoeven alleen een korte vragenlijst in te vullen om zo'n klantenkaart te krijgen. De zo verkregen gegevens worden via de

klantenkaart gekoppeld aan de geregistreerde aankopen. Daarmee weet de winkelier wat voor soort klanten welke producten kopen. Via het CBS (Centraal Bureau voor de Statistiek) kan de winkelier demografische gegevens opvragen over de wijk waar zijn mogelijke nieuwe winkel komt. Hoe komen de gegevens van het CBS overeen met zijn eigen klantgegevens? Door zijn eigen gegevens te koppelen aan de gegevens van het CBS kan de winkelier beter inschatten of de mogelijke nieuwe winkel de juiste soort mensen zal aantrekken en dus of die winkel winstgevend kan zijn. Uiteraard is dat geen garantie voor succes, de winkelier kan nog steeds de verkeerde beslissing nemen. Ook kan er iets veranderen waardoor de beslissing anders uitpakt dan gedacht. Er kan een economische crisis uitbreken of er komt een nieuwe concurrent bij, waardoor de situatie anders wordt dan voorzien. Business Intelligence draait om het nemen van gefundeerde (en daarmee betere) beslissingen. Daar is informatie voor nodig.

1.1.2 Van wie is Business Intelligence?

Laten we nog eens teruggaan naar de definitie uit de vorige paragraaf. Een gevaar van de definitie die we in paragraaf 1.1.1 gegeven hebben, is de expliciete benoeming van applicaties en infrastructuur. De applicaties zijn de programma's waarmee mensen werken. Dat kan Microsoft Excel zijn, een van de meest gebruikte Business Intelligence-applicaties, maar ook andere programma's zoals rapportageprogramma's. Ook de achterkant (back office) hoort daarbij met databases, datawarehouses en kubussen (later in dit boek leer je daar meer over). Met de infrastructuur wordt dan bedoeld welke hardware, welke computers, er gebruikt worden. Hoeveel servers hebben we nodig, aan welke specificaties moeten deze servers voldoen (hoeveel geheugen, hoeveel processoren, hoeveel opslagcapaciteit), wat voor netwerk hebben we nodig?

In de rest van de paragraaf zien we geen enkele verwijzing naar hardware terug. Ook worden applicaties en hulpmiddelen niet verder genoemd. Uiteindelijk moeten we de zaken concreet maken en mensen hulpmiddelen, tools, geven om hun werk beter te doen. Maar de tools zijn van ondergeschikt belang. Business Intelligence draait als het goed is niet om tools en datawarehouses. Het draait om mensen en de beslissingen die zij nemen. Het heet ook niet voor niets *Business Intelligence*. Als het om datawarehouses draaide had het wel *Techniek Intelligence* of iets dergelijks geheten. De bovenstaande constatering heeft als consequentie dat Business Intelligence niet van ICT is.

■ Business Intelligence draait om mensen en de beslissingen die zij nemen.

De meeste Business Intelligence-projecten zijn ICT-projecten. Dat is wel begrijpelijk omdat de concrete oplossingen die we bouwen veel IT-gerelateerde componenten bevatten. Er gaat immers software ingezet worden om het doel te bereiken. Die software draait op hardware die, voor een succesvolle implementatie, aan de juiste specificaties

moet voldoen. Zowel de hardware als de software moet beheerd worden zodra de implementatie af is. Allemaal werk voor IT en dus speelt ICT een belangrijke rol bij elke Business Intelligence-implementatie.

Het gevaar zit niet in het feit dat ICT een rol speelt. Het gevaar zit in het feit dat de rol zo groot en dominant wordt, dat het een project van alleen de afdeling ICT wordt. Het doel van het project moet duidelijk voor ogen gehouden worden. Het doel is niet mooie technische oplossingen bouwen met de nieuwste technieken.

■ Business Intelligence heeft als doel competitief voordeel te creëren en organisaties slimmer te kunnen laten werken.

Het doel is zelfs niet mensen in een organisatie beter beslissingen te laten nemen. Beide zijn slechts middelen om tot het doel te komen. Het echte doel kan variëren. Wikipedia zegt: 'Business Intelligence heeft als doel competitief voordeel te creëren en organisaties slimmer te kunnen laten werken.' Dit moet voor echte projecten concreet gemaakt worden. Bijvoorbeeld: we moeten kosteneffectiever werken door slimmer in te kopen en/of door de supply chain (hoe, via welk proces, komt een product van de leverancier bij de klant) efficiënter in te richten.

Merk op dat dit genuanceerder is dan het doel dat in paragraaf 1.1.1 is benoemd. Het doel van paragraaf 1.1.1 is meer het middel om dit 'echte' doel te bereiken.

Als de informatie waarin een Business Intelligence-oplossing voorziet, niet helpt bij het maken van de juiste beslissingen (ICT heeft iets gemaakt wat de mensen op de werkvloer niet kunnen gebruiken), is het project mislukt, ook als het vanuit een technisch perspectief een perfecte implementatie is. Net zo goed is het mislukt als de beoogde gebruikers er niet mee gaan werken. Misschien zetten zij hun hakken in het zand omdat ze bang zijn voor hun baan. Misschien willen ze gewoon niet leren werken met nieuwe tools. Het kan ook zijn dat ze de informatie die uit het systeem komt niet vertrouwen, al dan niet terecht. De achterliggende reden maakt eigenlijk niet uit, als de gebouwde Business Intelligence-oplossing niet gebruikt wordt, is het project mislukt (of je bent nog niet klaar).

Bovenstaande is als het goed is een open deur. Iedereen snapt dat er meerwaarde moet zijn en dat die er alleen is als een oplossing ook echt zinvol gebruikt wordt. Het risico dat uw project bovenstaand lot beschoren is, wordt groter naarmate de betrokkenheid van de gebruikers kleiner is gedurende het project. De eindgebruikers moeten van het begin aan betrokken zijn. Sterker: het moet (zo veel mogelijk) hun project zijn.

■ Business Intelligence is van de hele organisatie, niet van (alleen) ICT.

1.1.3 Voor wie is Business Intelligence?

Na de vorige paragraaf, waarin ik betoog dat Business Intelligence van de business (van de beoogde gebruikers) is, lijkt het duidelijk voor wie Business Intelligence

is. Voor diezelfde gebruikers natuurlijk. Maar wie zijn die gebruikers? Is Business Intelligence alleen voor grote bedrijven met grote budgetten? Of kunnen kleine bedrijven ook profiteren van Business Intelligence? Is Business Intelligence iets voor financiële instellingen zoals banken en verzekeraars of kunnen productiebedrijven en fabrieken er ook baat bij hebben?

Naast onderscheid in soorten bedrijven is de vraag ‘voor wie is het eigenlijk?’ ook zinvol te stellen binnen één organisatie. Is Business Intelligence typisch iets voor de directie? Welke managementlagen hebben profijt van Business Intelligence? Kunnen ook mensen op de werkvloer er hun voordeel mee doen?

Om het antwoord op die vragen kort te houden: Business Intelligence is voor iedereen. Microsoft kwam rond de introductie van SQL Server 2005, in het jaar 2005, met de slogan ‘*BI for the masses*’. Uiteraard is dat niet meer dan een slogan uit de marketing-fabriek van Microsoft en in dat licht moet je de zin ook plaatsen. Met deze slogan markeert Microsoft wel een nieuwe periode binnen het denken over Business Intelligence.

■ Business Intelligence is voor iedereen. Het is zowel interessant voor grote als voor kleine bedrijven. Het is in elke branche interessant. Binnen een bedrijf biedt het meerwaarde aan alle soorten functies, van hoog tot laag.

In vroegere implementaties, toen de term Business Intelligence nog niet in zwang was, werd er voornamelijk gesproken over **Decision Support Systems (DSS)**. In die tijd waren projecten vaak langdurig en duur. Daarmee waren Decision Support Systems voornamelijk voorbehouden aan grote bedrijven met grote budgetten. De informatie die uit deze systemen kwam, was vaak hoog geaggregeerd. Je kunt dan bijvoorbeeld denken aan kwartaalomzetten in plaats van aan gedetailleerde informatie over individuele orders. Dat maakt dat de informatie die deze systemen leverde eigenlijk alleen bruikbaar was als stuurinformatie hoog in de organisatie. Ze leverden informatie voor het hoger management.

Wat Microsoft bedoelde te zeggen met ‘*BI for the masses*’, is dat Business Intelligence-projecten geïmplementeerd kunnen worden met SQL Server (het databaseplatform van Microsoft) en alle bijbehorende functionaliteit zoals SQL Server Analysis Services. Dat houdt in dat er geen speciale dure software nodig is, met name wanneer het bedrijf in kwestie toch al SQL Server heeft draaien. In kleine omgevingen, waar alles op een en dezelfde server draait, zijn er zelfs geen extra softwarelicenties nodig. In grote (scale-out) scenario's gaat dat niet meer op, maar dan hebben we het ook over grotere bedrijven met meer budget. Daarmee komt Business Intelligence binnen het bereik van zo'n beetje elke organisatie, ook de kleinere bedrijfjes met kleine budgetten. Uiteraard zijn er naast Microsoft ook andere leveranciers van betaalbare software voor Business Intelligence-oplossingen.

Naast het feit dat Business Intelligence voor alle soorten bedrijven interessant is, zowel inhoudelijk als qua benodigde budgetten en kennis, is Business Intelligence ook al lang niet meer voor alleen het hoger management. Iedereen binnen een organisatie moet wel eens (of heel vaak) een beslissing nemen. Andere beslissingen hebben een ander soort informatie nodig (veel detail of juist niet) maar in alle gevallen zullen de

genomen besluiten beter gefundeerd genomen kunnen worden als er betrouwbare informatie aan ten grondslag ligt. Daarmee wordt de organisatie slimmer en beter.

En dat is het doel van Business Intelligence. En dus is Business Intelligence voor iedereen.

1.1.4 Self-service BI

De term self-service BI is de laatste tijd erg hot. Self-service BI komt voort uit de noodzaak tot acceptatie van de gebouwde oplossing door de gebruikers. Binnen een Business Intelligence-project zal je minimaal twee zaken moeten overwinnen:

1. Is de Business Intelligence-oplossing die gebouwd wordt wel wat de gebruikers nodig hebben?
2. Komt de oplossing wel op tijd?

Self-service BI, waar de gebruikers zelf maken wat ze nodig hebben, kan hier bij helpen.

■ **Self-service Business Intelligence** betekent dat eindgebruikers hun eigen rapporten en analyses ontwerpen en in gebruik nemen, gebruikmakend van goedgekeurde tools en een ondersteunde BI-architectuur.

Om de eerste vraag positief te kunnen beantwoorden moet je een goede informatie-analyse doen. Voor het tweede punt moet je voornamelijk snel nieuwe functionaliteit kunnen ontwikkelen. Beide punten zijn makkelijker gezegd dan gedaan, zoals hieronder duidelijk wordt.

Informatieanalyse

Een van de belangrijkste stappen in het hele Business Intelligence-project is het achterhalen van de wensen van de gebruikers ('requirements gathering' ofwel informatieanalyse). Waar heeft welke gebruiker wanneer last van? Je moet altijd een gedegen informatieanalyse uitvoeren. Hoe verloopt een proces, wat zijn de beslispunten binnen dat proces en welke informatie hebben de mensen nodig op het moment dat de beslissing genomen moet worden?

■ **Informatieanalyse** is het achterhalen wat de behoeftes, wensen en eisen van de beoogde eindgebruikers van een systeem zijn.

Een voor de hand liggende en veel toegepaste manier om de informatieanalyse te doen is door middel van interviews. Belangrijk bij die interviews is het stellen van open vragen. Open vragen zijn vragen in de trant van: Wat heb je nodig? Welke informatie gebruik je wanneer, hoe en waarom? Een gesloten vraag zou zijn: Heb je een omzetrapportage nodig?

Als je geen open vragen stelt, krijg je je eigen ideeën uit het interview terug omdat je te sturend bent geweest bij je vragen. Alleen leiden open vragen vaak tot open antwoorden. En schiet je daar wel voldoende mee op?

Je zou bijvoorbeeld kunnen vragen: waar ondervindt u in het dagelijkse werk problemen? Waar hebt u meer informatie of kennis nodig om uw werk goed te kunnen doen? Dit zijn immers open vragen en vanuit jouw perspectief zeer relevante vragen.

Het antwoord dat je in de praktijk helaas vaak krijgt als je open vragen stelt, luidt: ik heb geen problemen, ik kan mijn werkzaamheden prima uitvoeren!

Er bestaan veel redenen waarom je dit soort antwoorden krijgt. Enkele redenen kunnen zijn:

- toegeven dat je je werk niet optimaal doet is moeilijk;
- angst voor het onbekende;
- onwetendheid over de mogelijkheden;
- miscommunicatie.

Toegeven dat je je werk niet optimaal doet is moeilijk

Bewust of onbewust hebben mensen het idee dat ze moeten toegeven hun werk niet goed te doen als ze iets anders zeggen dan: ik heb geen problemen, ik kan mijn werkzaamheden prima uitvoeren! Ze doen hun werk misschien al jaren op dezelfde manier. Dus nu zeggen dat ze hier of daar meer informatie nodig hebben, is een beetje alsof ze moeten toegeven al jaren hun werk niet goed te doen.

Angst voor het onbekende

De angst voor verandering kan bij veel mensen een rol spelen. Ik weet waar ik nu aan toe ben, maar straks, als er een nieuw systeem komt, hoe zal mijn werk er dan uitzien? Wordt mijn functie overbodig met het nieuwe systeem?

Onwetendheid over de mogelijkheden

Een andere veel voorkomende reden dat open vragen tot onbevredigende antwoorden kan leiden, is dat mensen zich helemaal niet realiseren waar er knelpunten bestaan en/of wat de mogelijkheden zijn.

Voorbeeld

Ik heb ooit iemand horen zeggen dat een query op een database die acht uur duurde geen enkel probleem was. Deze persoon was zo gewend aan een doorlooptijd van acht uur dat hij 's avonds om vijf uur, voor het naar huis gaan, de query startte, zodat hij de volgende morgen zijn werk kon doen: die gegevens analyseren. Hier heeft iemand zijn dagindeling en werkwijze aangepast aan de omstandigheden. Na wat database-tuning kon de query teruggebracht worden naar een doorlooptijd van nog geen minuut!

Miscommunicatie

Los van bovenstaande kunnen er nog problemen van een heel andere orde optreden.

Voorbeeld

Ik heb ooit in een vergadering gezeten waar een request-for-change (verzoek een verandering aan te brengen in een bestaande applicatie) besproken. Uit de organisatie was een verzoek gekomen de applicatie op enkele punten te wijzigen. Degene die kwam uitleggen wat ze precies bedoelden met de gevraagde verandering, vertelde een verhaal in de trant van: en dan klik ik hier en dan klik ik daar. Ze was op dat moment niet in staat uit haar handeling-denken te komen en het proces wat ze eigenlijk uitvoerde te verwoorden. De programmeur die bij de vergadering zat en de wijzigingen moest gaan implementeren, antwoordde in termen als: dan moet ik een variabele bijmaken, dan moet ik de stored procedure aanpassen,... Hij gebruikte allemaal technische termen. Ook hij kwam niet los van zijn dagelijkse jargon en sprak dus niet dezelfde taal als de gebruiker.

Doordat beide mensen in dit verhaal in hun eigen jargon bleven praten en niemand even boven de dagelijkse praktijk kon uitstijgen, ontstond er onbegrip, bijna ruzie. Het ergste daarvan was eigenlijk nog dat ze het eens waren met elkaar maar dat zelf niet doorhadden.

De moraal hiervan is: zelfs als je de goede vragen stelt, en zelfs als de gebruiker goede antwoorden geeft, is er een (niet geringe) kans dat je elkaar ergens onderweg niet goed verstaat. Je denkt te begrijpen wat de gebruiker zegt, maar uiteindelijk heeft hij iets anders bedoeld. Als dat niet vroegtijdig aan het licht komt, levert de Business Intelligence-oplossing straks de verkeerde informatie op.

Uit een en ander blijkt hoe belangrijk en moeilijk een goede informatieanalyse is. Stel je wel de juiste vragen en krijg je wel de juiste antwoorden? Het wordt heel anders als de gebruiker naar je toekomt met 'dit en dit heb ik nodig, kan je dat maken?'. Nog beter is misschien zelfs dat je naar de gebruiker kunt stappen en zeggen 'ik zie dat u dit en dit gebruikt, zullen we samen aan een meer gestructureerde oplossing werken want ik zie dat u iets succesvol heeft gemaakt?'

Met self-service BI kun je bovenstaande (misschien) wel bereiken. Laat de gebruikers zelf beginnen. Zij weten immers het beste wat ze nodig hebben. En tegenwoordig zijn heel veel mensen heel goed in staat met techniek om te gaan. Excel is een uitstekend voorbeeld van een applicatie waar mooie dingen mee gedaan kunnen worden. Bovendien zijn er veel mensen die met Excel kunnen werken.

Timing

Een ander probleem, naast de hierboven beschreven mismatch tussen wensen en gebouwde oplossingen, is het moment waarop de oplossing door de gebruikers gebruikt kan gaan worden. Stel dat de hier boven beschreven informatieanalyse helemaal perfect is verlopen. Alle eisen en wensen zijn op tafel gekomen en alles is precies gemaakt zoals gevraagd. Hoe lang heeft het je gekost om de oplossing te bouwen?

De wereld verandert tegenwoordig snel. Veel branches, zoals de gezondheidszorg, hebben regelmatig met nieuwe wetgeving te maken. Door bijvoorbeeld de economische crisis hebben we te maken met reorganisaties, overnames en faillissementen. Dit

soort zaken kan de informatiebehoefte in een organisatie sterk veranderen. Vragen die enkele maanden geleden heel relevant waren, zijn ondertussen achterhaald. Veel nieuwe zaken hebben de kop opgestoken en daar hoort nieuwe informatie bij.

Los van de wereld om ons heen die verandert, is er sprake van voortschrijdend inzicht binnen de organisatie zelf. Mensen leren, krijgen meer ervaring, doorzien zaken met meer nuance en beginnen zich andere zaken af te vragen. Zijn dat zaken die jouw Business Intelligence-oplossing ook kan beantwoorden?

Met self-service BI is dit probleem minder groot, tot misschien wel helemaal verholpen. Het idee van self-service BI is immers dat de mensen in de organisatie zelf hun BI doen. Op het moment dat er nieuwe vragen opkomen, kunnen de mensen die met die vragen zitten op het moment zelf aan de slag om de vragen te beantwoorden. Op een later moment kunnen we bekijken of we de ad-hoc-oplossing die dat oplevert, integreren in de bestaande centrale Business Intelligence-oplossing (in het datawarehouse) of dat het om een eenmalige exercitie ging.

Self-service Business Intelligence draait om het in staat stellen van eindgebruikers om zelf in hun informatiebehoefte te voorzien zonder anderen te vragen oplossingen voor ze te bouwen ten einde de juiste informatie op het juiste moment ter beschikking te hebben.

1.1.5 Waaron Business Intelligence?

Tot nu toe heb ik getracht een definitie te geven voor Business Intelligence en een kader te scheppen waarbinnen Business Intelligence valt. Wat is het, van en voor wie is het? Een belangrijke vraag die overblijft is: waarom zou je als organisatie iets met Business Intelligence doen? Dat anderen het doen of dat het ‘hot’ is, zijn redenen om deze vraag te stellen, niet direct redenen om het zelf te gaan implementeren.

Hopelijk hebben de vorige paragrafen de vraag indirect al beantwoord. Business Intelligence gaat er om de mensen binnen een organisatie beter hun werk te laten doen, waarmee de organisatie als geheel beter wordt. Een bedrijf moet zich altijd en continu de vraag stellen: wat is mijn onderscheidend vermogen? Waarin ben ik beter/anders dan de concurrenten? Denk bijvoorbeeld eens aan internetboekhandels. Waarom zou je een bestelling plaatsen bij bol.com en niet bij amazon.com?

Voorbeeld

Voor een bedrijf is het niet altijd gemakkelijk ergens voordeel te halen. De prijzen van Nederlandstalige boeken staan bijvoorbeeld vast, dus prijsvoordeel is er niet. Voor de meeste bedrijven is er wel aan kapitaal te komen via banken of investeerders. Daarmee hebben bedrijven gelijke kansen op het gebied van bijvoorbeeld investeren in hardware en software voor hun computerpark. De meeste webshops gebruiken dezelfde pakket-service om de bestellingen nog de volgende dag bij u thuis te bezorgen. Waar zit het verschil tussen bedrijf A en bedrijf B?

Een voor de hand liggend, maar zeer belangrijk verschil, tussen twee concurrerende bedrijven zijn de mensen die er werken. Misschien is het belangrijkste kapitaal van een bedrijf wel het human capital, de mensen. Naast dat een bedrijf gebaat is bij de beste mensen wat betreft opleiding, potentie, ervaring enzovoort, is het van belang deze mensen in staat te stellen hun werk goed en efficiënt te doen. Dat laatste is wat Business Intelligence tot doel heeft.

De vraag 'waarom Business Intelligence' is daarmee eigenlijk de verkeerde vraag. De vraag zou moeten zijn waarom níét. Goede, accurate, tijdige informatie kan het verschil maken tussen succes en faillissement. Je zou dus wel gek zijn niet iets met BI te doen.

1.1.6 Conclusie

Deze paragraaf begon met het definiëren van het begrip Business Intelligence. Die definitie bracht ons bij allemaal zaken die de context bepalen van een Business Intelligence-project. Dat allemaal samenvattend zouden we tot de volgende definitie kunnen komen:

Business Intelligence is *iedereen* in een organisatie op het *juiste moment* van de *juiste informatie* in het *juiste formaat* te voorzien met als doel competitief voordeel te behalen door verbeterde besluitvorming te realiseren.

1.2 Andere terminologie en definities

Nu je weet wat Business Intelligence is en wat het beoogde doel van Business Intelligence is, wordt het tijd om de belangrijkste componenten van een Business Intelligence-oplossing te introduceren. Het gaat er in dit hoofdstuk om de begrippen op hoofdlijnen te kunnen plaatsen. De relevante details volgen in latere hoofdstukken.

1.2.1 Datawarehouse

Bijna alle bedrijven die iets doen met Business Intelligence hebben een datawarehouse. Een datawarehouse is een centrale database die als doel heeft betere rapporten te kunnen maken en betere (data)analyses uit te kunnen voeren. De inhoud van het datawarehouse ontstaat door de gegevens van één of meer gegevensbronnen te combineren.

Het bouwen van een datawarehouse brengt een aantal voordelen met zich mee:

- Historische gegevens kunnen worden bijgehouden.
- De kwaliteit van de gegevens kan beter worden gecontroleerd.
- Gegevens uit verschillende bronnen kunnen worden geïntegreerd.
- De performance kan worden geoptimaliseerd voor het doel: rapporteren en analyseren.
- Alle rapportages komen nu vanuit één bron, wat de eenduidigheid van de rapportages bevordert.

In hoofdstuk 3 en 4 leer je meer over het hoe en waarom van datawarehouses.

■ Een **datawarehouse** is een centrale database gevuld met gegevens uit één of meer aparte bronnen met als doel het maken van rapportages en het uitvoeren van data-analyses.

1.2.2 Extract, Transform, Load

Uit dit alles komt naar voren dat een datawarehouse gevuld wordt met gegevens uit één of meer bronsystemen. Dat houdt in dat je de gegevens uit deze bronsystemen moet halen en dat je de gegevens in het datawarehouse moet inlezen. Het proces dat dit verzorgt heet het ETL-proces. De letters ETL staan voor:

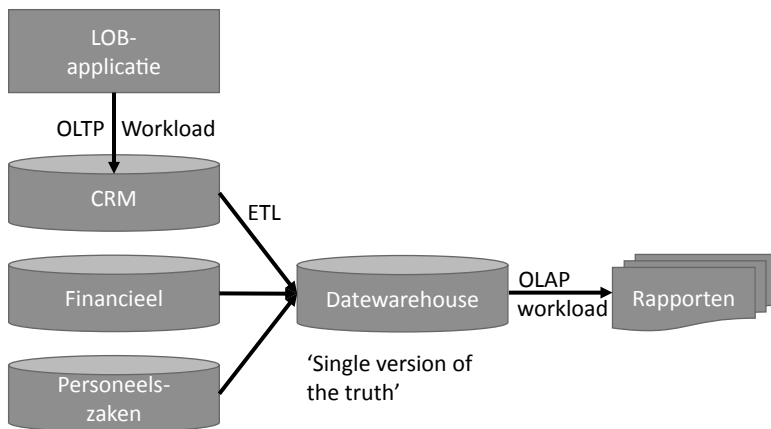
- Extract – haal gegevens uit bronsystemen;
- Transform – pas de gegevens aan aan de eisen die het datawarehouse stelt aan de gegevens;
- Load – laad de gegevens in het datawarehouse.

Met name de middelste stap, de transformaties, de aanpassingen die gedaan moeten worden, vormen hier de belangrijke stap. Het ETL-proces zorgt voor de juiste vulling van het datawarehouse en daarmee voor de kwaliteit van de rapportages die je maakt en de analyses die je doet. De aanpassingen die je doet aan de gegevens bepalen deze datakwaliteit.

In hoofdstuk 5 van dit boek kun je meer lezen over het ETL-proces, welke uitdagingen je tegenkomt, wat voor tools je kan gebruiken, et cetera.

■ Het **ETL-proces** zorgt voor de juiste vulling van het datawarehouse en daarmee voor de kwaliteit van de rapportages die je maakt en de analyses die je doet.

Figuur 1.1 laat zien hoe gegevens vanuit bronsystemen naar het datawarehouse overgehaald worden ten einde rapportages te baseren op het centrale datawarehouse.



Figuur 1.1 Datawarehouse-architectuur

1.2.3 Staging, ODS

In het plaatje in figuur 1.1 van de vorige paragraaf lijkt het net of het ETL-proces de gegevens rechtstreeks van de bronsystemen overhaalt naar het datawarehouse. In theorie is dat mogelijk, zeker met moderne ETL-tools zoals SQL Server Integration Services. In de praktijk zullen de gegevens bijna altijd eerst in een aparte laag, een aparte database, gezet worden. Dit wordt de staging area of staging database genoemd. Bij grote implementaties zijn er misschien zelfs wel meer staging-lagen voordat de gegevens het datawarehouse in gaan.

■ Een **staging database** is een database waarin gegevens tijdelijk worden opgeslagen tijdens het ETL-proces alvorens ze worden overgehaald naar het datawarehouse.

Net zo goed als dat het lijkt dat de gegevens meteen het datawarehouse in worden gepompt, lijkt het alsof de rapportages hun gegevens rechtstreeks uit het datawarehouse halen. Afhankelijk van je datawarehousestrategie is dat niet het geval. Vaak zal er nog een laag van datamarts gemaakt worden. Dit zijn als het ware kleine datawarehouse's, gemaakt voor specifieke rapport- en analysedoeleinden. Er kan bijvoorbeeld een aparte datamart gemaakt worden voor alleen de rapportages die te maken hebben met het inkoopproces. In de komende hoofdstukken van dit boek zul je dat nog verder tegenkomen.

Een veelgebruikte term naast staging en rapportagedatabase is ODS, Operational Data Store. Een ODS is de eerste plek waar gegevens landen nadat ze uit het bronsysteem zijn gehaald. Het kan dienen als tussenstop voor gegevens die naar het datawarehouse gaan. In die zin is het een staging database. Soms worden rapportages gebaseerd op deze ODS. Dan wordt het ook wel een rapportagedatabase genoemd.

1.2.4 Kubussen, modellen

Bij veel Business Intelligence-implementaties kan een datawarehouse niet voldoen aan alle wensen en eisen die er binnen een organisatie leven ten aanzien van Business Intelligence. Er kan nog een aantal problemen optreden:

- Er kunnen nog performanceproblemen zijn omdat achter veel rapporten zware (langlopende) query's schuilgaan.
- Veel rapportage- en analysehulpmiddelen (BI-tools) hebben moeite goede query's te genereren op databases.
- Eindgebruikers willen gemakkelijk en snel op een flexibele, niet vooraf gedefiniëerde wijze door de gegevens heen kunnen 'wandelen' (browsen).

Semantische modellen kunnen helpen deze problemen op te lossen. Bij het ontwerpen van een database, en dus ook van een datawarehouse, spelen technische afwegingen een voorname rol. Een semantisch model vertaalt die technische database naar voor de eindgebruikers betekenisvolle entiteiten. Betekenis en verbanden worden toegevoegd aan de gegevens. Dit stelt zowel mensen als tools in staat gemakkelijk en flexibel met de gegevens te werken.

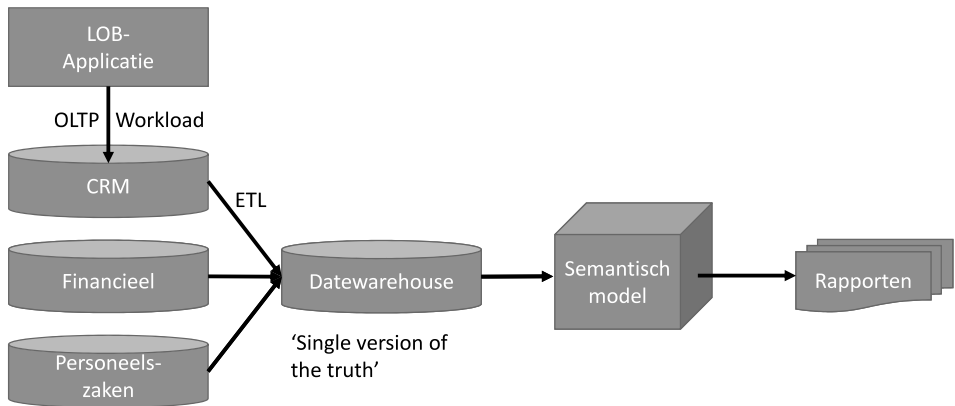
■ Een **semantisch model** is een abstractielaag die betekenis en verbanden toevoegt aan de gegevens in een database.

Een kubus is een soort database. Een kubus wordt wel een multidimensionale database genoemd. Een kubus is tegelijkertijd een model. Bij het ontwerpen van de kubus wordt rekening gehouden met de betekenis van de gegevens, worden de onderlinge verbanden beschreven, en wordt het jargon van de betreffende organisatie gebruikt. Een kubus is een soort draaitabel in het groot. Een kubus is uitermate geschikt om op een hele gemakkelijke en flexibele wijze met gegevens te 'spelen'.

■ Een **kubus** is een meerdimensionale draaitabel.

In hoofdstuk 8 kun je meer lezen over modellen en kubussen.

In figuur 1.2 zie je de logische Business Intelligence-architectuur zoals we die tot nu toe besproken hebben.



Figuur 1.2 Business Intelligence-architectuur met modellen

1.2.5 Datamining en machine learning

Datamining is ‘het gericht zoeken naar (statistische) verbanden in gegevensverzamelingen met als doel profielen op te stellen voor wetenschappelijk of commercieel gebruik.’ Met datamining ga je pas echt goed gebruikmaken van de enorme hoeveelheid gegevens die je hebt. Bedrijven zijn vooral altijd heel goed geweest in het verzamelen van gegevens. Maar wat doen we met die gegevens? Kan dat omgevormd worden tot nuttige informatie?

■ **Datamining** is het gericht zoeken naar (statistische) verbanden in gegevensverzamelingen met als doel profielen op te stellen voor wetenschappelijk of commercieel gebruik.

Voorbeeld

Een klassiek verhaal over datamining, en waar het fout kan gaan, is het verhaal van een supermarktketen die had aangetoond dat mensen die bier drinken luiers dragen. Door alle aankopen te analyseren hadden ze gezien dat er heel vaak een kratje bier op de kassabon staat als er ook luiers op de bon staan. En dus...

Wat je met datamining bijvoorbeeld goed kan doen, is het zoeken naar correlaties tussen verschillende producten op basis van kassabongegevens. Welke producten worden vaak tegelijkertijd verkocht? In een fietsenwinkel wordt vaak een bidon verkocht als er een bidonhouder verkocht wordt. Andersom is het verband veel minder sterk.

Met datamining kunnen ook patronen in gegevens herkend worden. Schadeverzekeringen hebben te maken met een redelijk percentage onterechte claims. Om de onterecht uitgekeerde schade zoveel mogelijk te beperken, hebben ze schade-experts in dienst die claims controleren op echtheid. Echter, het is te duur om elke

claim te gaan controleren. Willekeurige steekproeven brengen te weinig frauduleuze claims aan het licht. Welke claims moeten nu gecontroleerd worden zodat een hoog percentage frauduleuze claims gedetecteerd wordt met zo min mogelijk mensen?

Met datamining kunnen patronen in claims gedetecteerd worden. Terechte claims blijken overeenkomsten te hebben. Een claim die niet aan het normale patroon voldoet, is niet per definitie frauduleus. De kans dat er sprake is van fraude is echter wel groter. Zo'n claim is dus een mooie kandidaat voor een controle.

Waar datamining voornamelijk op zoek gaat naar (nog) niet bekende verbanden binnen de gegevens, maakt machine learning gebruik van dit soort verbanden en patronen om voorspellingen te doen. We komen dan uit in het gebied van de Predictive Analytics, ofwel voorspellende analytics.

Met een klassiek rapport kan een ijswinkel zien hoeveel ijs er gister of vorig jaar is verkocht. Dat is belangrijk om bijvoorbeeld de belastingaangifte te doen of bij de aanvraag voor een nieuwe lening om verder te investeren.

Maar zeker zo belangrijk is te weten hoeveel ijs er morgen verkocht zal gaan worden. Er moet vandaag immers beslist worden hoeveel ingrediënten er ingekocht moeten worden en hoeveel personeel er ingepland moet worden. Met Predictive Analytics en machine learning bereiken we dat deze beslissingen beter gemaakt kunnen worden. Het is een mooi voorbeeld van operationele Business Intelligence: betere (operationele) beslissingen op basis van betere informatie.

■ **Machine learning** is een toepassing van artificial intelligence (AI) waarmee systemen automatisch leren op basis van ervaring (gegevens uit het verleden) zonder expliciet geprogrammeerd te worden.

In hoofdstuk 7 zullen we wat dieper ingaan op datamining, machine learning en Advanced Analytics.

1.2.6 Big Data

Tot nu toe is er in dit hoofdstuk van uitgegaan dat het datawarehouse een relationele database is. Dat wil zeggen dat de gegevens leven in tabellen die beheerd worden door een RDBMS (Relational DataBase Management System). Oracle, SQL Server en DB2 zijn voorbeelden van RDBMS'en. Een RDBMS levert veel voordelen op rond beheer van gegevens en consistentie van gegevens. De prijs die we daarvoor betalen is dat er een zekere overhead zit in het werken met gegevens. Overhead betekent performanceverlies. Bovendien moet je een databasestructuur (tabelstructuur) bedenken. Maar veel interessante data is helemaal niet zo gestructureerd!

Big Data refereert aan datasets waarbij bovenstaande een issue wordt. Over het algemeen zijn er drie aspecten aan Big Data die hier mee te maken hebben. In het Engels spreekt men van de drie v's:

1. Volume – de hoeveelheid aan gegevens;
2. Velocity – de snelheid waarmee gegevens ontstaan en opgeslagen (en verwerkt) moeten worden;
3. Variety – de diversiteit van de gegevens.

Hoewel niet noodzakelijk kun je bij Big Data denken aan informatie op het internet: Twitter, Facebook, blogs, forums enzovoort. Het aantal tweets dat verstuurd wordt, is zo immens groot dat het niet meer te doen is dat op te slaan in een tabel. De tabel zou zo groot worden dat het beheer ervan een goede DBA (database-administrator) hoofdpijn zou bezorgen.

Naast het aantal verstuurde tweets is de snelheid waarmee nieuwe tweets binnenkomen enorm. Hoe meer er tegelijkertijd binnenkomen, hoe groter de kans dat veel mensen op dat moment hun timeline bekijken. Dus de lees- en de schrijfdruk op het systeem is tegelijkertijd heel groot.

Ten slotte kun je nog concluderen dat er weinig structuur in tweets zit. Er is een zender en een datum. Verder zijn het willekeurige tekens met veel verschillende hashtags.

Bovenstaande feiten maken tweets moeilijk om te verwerken in een klassiek datawarehouse. De informatie en potentie ervan is echter enorm. Bioscopen kunnen al op basis van tweets in de week voorafgaand aan een première inschatten hoe druk het in de eerste week gaat worden. Hoe meer positieve tweets over de film, hoe beter hij zal lopen. Goed om te weten als je planner bent: welke film krijgt volgende week de grootste zaal?

■ **Big Data** refereert aan datasets waarbij de hoeveelheid aan gegevens, de snelheid waarmee gegevens verwerkt moeten worden en/of de diversiteit van die gegevens een probleem wordt als de gegevens op een klassieke manier worden behandeld.

Voor dit soort Big Data is de laatste tijd een heel scala aan nieuwe soorten databases ontstaan. Vaak worden die allemaal samengevat onder de noemer noSQL (SQL staat voor Structured Query Language). Dat is een slechte naam. Hij refereert voornamelijk aan het feit dat het niet om een relationele database gaat. Het opensourcesoftware-framework Hadoop is hier een voorbeeld van.

Hoofdstuk 7 gaat verder in op het onderwerp Big Data en de rol van Big Data in een modern datawarehouse.

1.3 Tot slot

Zoals in de eerste definitie, gegeven op de eerste bladzijde van dit hoofdstuk, al staat: Business Intelligence is een paraplueterm waar heel veel onder valt. Van simpel rapporten bouwen tot Big Data, van ODS tot kubussen en datamining. Het nadeel van deze samenvatting is dat het focust op technische termen. Het maakt Business Intelligence een technisch iets.

De twee belangrijkste lessen van dit hoofdstuk zijn:

1. Business Intelligence heeft tot doel de processen binnen een organisatie te verbeteren door de mensen betere en accuratere informatie te geven waardoor zij betere beslissingen kunnen nemen.

2. Business Intelligence is van iedereen, maar met name van de mensen die werken aan de primaire processen binnen een organisatie. Business Intelligence is niet iets van alleen ICT.

1.3.1 Verder

In hoofdstuk 2 kun je meer lezen over Business Intelligence-projecten. Wat zijn de ambities van een organisatie en welke doelen wil men met Business Intelligence bereiken? Welke valkuilen kent een BI-project en hoe omzeil je die valkuilen? Hoofdstuk 3 pakt het verhaal meer technisch op: welke strategieën zijn er om een datawarehouse op te zetten?

Hoofdstuk 2

Business Intelligence in de organisatie

In het eerste hoofdstuk heb je een beeld gekregen van wat Business Intelligence is. Naast definities is er een overzicht gegeven van veel gebruikte terminologie. Als het goed is, is het duidelijk geworden dat Business Intelligence een multidisciplinaire tak van sport is en niet iets van alleen de afdeling ICT. Te veel focus op ICT brengt zelfs een aantal gevaren met zich mee. Vandaar dat we in dit hoofdstuk ingaan op de niet-technische aspecten van een Business Intelligence-project.

Als eerste gaan we kijken naar de rol en de plek van Business Intelligence binnen de organisatie. We zullen zien dat Business Intelligence veel meer is dan een ICT-project. Sterker nog, wel zullen zien dat het meer is dan een project en dat het al helemaal niet een ICT-project is. Uiteindelijk zullen er wel een of meer Business Intelligence-projecten uitgevoerd gaan worden. Daar besteden we in paragraaf 2.2 aandacht aan.

2.1 Business Intelligence als competentie

2.1.1 Business Intelligence-volwassenheid

Bijna elke organisatie is in meer of mindere mate bezig met Business Intelligence. In hoofdstuk 1 heb je al gelezen dat het maken van rapportages al onder de term Business Intelligence geschaard kan worden. Voordat je bewust iets met Business Intelligence gaat doen, is het raadzaam te bedenken waar een organisatie staat. Hoeveel wordt er al aan Business Intelligence gedaan? Is dat bewust of onbewust? Wordt dat centraal gestuurd of zijn het losse initiatieven? Bestaat er ambitie binnen de organisatie ten aanzien van Business Intelligence?

■ Business Intelligence wordt als een waardevolle kerncompetentie beschouwd.

Wikipedia zegt over Business Intelligence dat het wordt beschouwd als een waardevolle kerncompetentie. Die stelling drukt uit wat de meest ideale rol is van Business Intelligence binnen een organisatie. Het zou mooi zijn als het beschouwd wordt als een waardevolle kerncompetentie binnen een organisatie. Het zou nog mooier zijn als dat terecht zou zijn. Maar het is geen belofte van Business Intelligence en er is al helemaal geen garantie.

Lang niet in alle organisaties wordt Business Intelligence als een kerncompetentie gezien. Maar welke rol speelt Business Intelligence dan wel binnen de organisatie? Of met andere woorden, hoe BI-volwassen is de organisatie? Steeds vaker meten bedrijven zich met **volwassenheidsmodellen**. Voor Business Intelligence bestaan deze modellen ook.

Op computerworld (www.computerworld.nl) vind je bijvoorbeeld een model dat vijf fases kent:

1. Prenatale fase
2. Peuter
3. Kind
4. Volwassen
5. Ontwikkeld

De benaming van de verschillende fases drukt uit dat het een groeiproces is om in de fase ontwikkeld te komen. Niemand wordt volwassen zonder eerst puber te zijn geweest, je doorloopt alle fases.

Prenatale fase

In deze fase is er eigenlijk helemaal geen sprake van Business Intelligence. Verschillende applicaties zullen hun eigen rapportages hebben waar de mensen mee werken om informatie te verkrijgen. Er is echter geen sprake van enige informatievergaring.

Peuter

In de peuterfase wordt in een of andere vorm gebruikgemaakt van wat wel spreadmarts genoemd wordt (zie The Data Warehouse Institute, tdwi.org). Alternatieve termen zijn datashadowsystems, IT-shadowsystems of human datawarehouses. Met deze termen wordt voornamelijk bedoeld dat individuen in de organisatie bezig zijn met het verzamelen van gegevens. Deze gegevens 'leven' vervolgens op de desktop van diegene die er mee werkt. Vaak zal Microsoft Excel gebruikt worden in dit soort situaties, vandaar de naam spreadmart (een datamart van spreadsheets of in de vorm van een spreadsheet), maar ook andere tools kunnen gebruikt worden.

■ Een **spreadmart** is een decentrale gegevensverzameling voor eigen gebruik, gemaakt door een individu, vaak opgebouwd met een spreadsheetprogramma.

Het woord shadow in de term shadowsystem, refereert aan het feit dat de gebouwde oplossing een soort schaduwboekhouding oplevert. De gegevens worden zonder centrale regie extra opgeslagen. Vaak is niet duidelijk of, en zo ja hoe vaak en wanneer, de gegevens worden bijgewerkt. Daarnaast is meestal niet bekend welke regels en definities worden gebruikt bij het transformeren van de gegevens naar informatie. Hoe wordt de ruwe data dat mooie grafiekje?

Deze fase heeft Excel een slechte naam bezorgd bij Business Intelligence-consultants. Excel groeit langzaam uit tot centrale database in plaats van een front-end-analysetool te zijn. In het laatste is Excel heel goed, daar waar Excel een belangrijke (de belangrijkste) gegevensbron of database wordt, is het einde zoek.

Kind

In de fase van kind begint er voorzichtig wat regie in de Business Intelligence te komen. Er komt een centrale datamart waar rapportages en analyses op gemaakt worden. Met datamart bedoelen we hier een database die centraal beheerd wordt en die speciaal is ingericht om Business Intelligence-achtige activiteiten op te laten plaatsvinden. IT kan de database beheren met het oog op security (wie krijgt toegang), back-up en restore en misschien zelfs (high) availability en synchronisatie (refresh) van de gegevens. Gebruikers kunnen met front-endtools, zoals Excel, connecties maken met de datamart om rapporten en analyses te maken.

Volwassen

Een datamart zoals gebruikt in de kindfase is een kleine zelfstandige database met een speciaal doel, bijvoorbeeld specifiek gericht op finance. Omdat andere afdelingen ook met rapporten en analyses bezig zijn, ontstaan er meer datamarts die onafhankelijk van elkaar zijn. Hoewel dat voordelen kan hebben, gaat het ontbreken van een centraal punt vroeg of laat opbreken. Dat centrale punt is het (corporate) datawarehouse.

Het datawarehouse is de centrale database waar alle gegevens verzameld, bewaard en beheerd worden. Deze database vormt het startpunt van alle verdere BI-activiteiten. Datamarts worden gevoed vanuit het centrale datawarehouse. Het datawarehouse wordt de *'single version of the truth'*.

Ontwikkeld

De laatste fase noemen we 'ontwikkeld'. Een organisatie die deze fase bereikt heeft, doet gevorderde analyses. Misschien worden complexe dataminingtoepassingen ingezet op grote datasets die hun oorsprong vinden zowel binnen de organisatie als ook daarbuiten.

In deze fase is Business Intelligence uitgegroeid tot een competentie. Waar Business Intelligence in het begin misschien nog wel een kostenpost is, of als zodanig wordt gezien, is het uitgegroeid tot een competentie die zijn geld meer dan opbrengt.

Hier hoort bij dat er specifieke rollen, taken en verantwoordelijkheden zijn gedefinieerd met betrekking tot zaken als beveiliging, anonimiseren van gegevens, eigenaarschap van business rules (definities van begrippen zoals cost-of-sales), eigenaarschap van gegevens en de kwaliteit van de gegevens. Dat valt allemaal onder wat we data governance noemen. In hoofdstuk 6 lees je daar meer over.

Groei

Zoals een kind niet in één keer volwassen is, maar langzaam tot volwassenheid groeit, geldt dat ook voor organisaties. Zo zal een bedrijf dat nog in de prenatale fase zit, niet met één project volwassen zijn.

Wat organisaties die in de fase ontwikkeld zitten vaak hebben, en wat organisaties die nog niet zo ver zijn kan helpen, is een zogenaamd Business Intelligence Competence Center (BICC). Dit is een speciaal team dat het liefst multidisciplinair is. Er horen niet alleen mensen van de afdeling ICT deel van uit te maken zodat het geen 'ICT-feestje' wordt. Het team houdt ontwikkelingen in de gaten, ziet kansen, enthousiasmeert andere mensen in de organisatie en houdt de overall kwaliteit van de geboden oplossingen in de gaten.



Jarenlang hebben organisaties gegevens verzameld in databases. Nu moeten zij de stap maken naar zinvol gebruik: Business Intelligence. Goed toegepast kan Business Intelligence een organisatie kosteneffectiever laten werken. Of een voorsprong geven op concurrenten.

Het boek beantwoordt de volgende vragen: wat is Business Intelligence, waarom zou je het toepassen en wat zijn de doelen? Wat is een datawarehouse en waar heb je dat voor nodig? Hoe vul je een datawarehouse met gegevens en waar moet je dan rekening mee houden? Wie zijn de gebruikers en hoe werken ze ermee? Daarbij komen ook ETL, Analytics en het front-end uitgebreid aan de orde.

Verder wordt dieper ingegaan op (Advanced) Analytics: van (het belang van) visualisaties tot en met toepassingen van Data Science en Machine Learning. Uiteraard kan de rol van de cloud hierbij niet achterblijven. Om de relatie met de praktijk te leggen worden diverse voorbeelden uitgewerkt.

Peter ter Braake is zelfstandig SQL Server docent/consultant. Hij is MCT sinds 2002 en was SQL Server MVP van 2012 tot en met 2017. Hij werkt sinds 2004 met Microsoft SQL Server Business Intelligence en is auteur van de boeken *Aan de slag met Reporting Services 2012* en *Handboek Power BI*.

Deze tweede druk is geactualiseerd en uitgebreid met een hoofdstuk over Big Data. Is het een *game changer* in de wereld van de Business Intelligence? Waarom en hoe kunnen bedrijven er hun voordeel mee doen?

In dit boek zit een code voor www.leerboekbusinessintelligence.nl. Daar vinden studenten en docenten extra materiaal, zoals meerkeuzevragen met feedback, uitgewerkte voorbeelden en de daarin gebruikte bestanden en scripts.

Leerboek Business Intelligence is geschreven voor studenten van bedrijfskundige studies, maar ook voor (bedrijfskundige) informatica. Omdat het boek zowel op technische als niet-technische lezers is gericht, is het ook geschikt voor professionals in het bedrijfsleven die een beter begrip willen krijgen van Business Intelligence in de breedste zin.

