

Leerboek Business Intelligence 3^e druk

Boom

Peter ter Braake



Leerboek Business Intelligence

Leerboek Business Intelligence

Peter ter Braake

Boom

**inclusief
website!**

Met behulp van onderstaande unieke activeringscode krijg je via **www.boomstudent.nl** toegang tot de website met extra materiaal. Deze code is persoonsgebonden en gekoppeld aan de 3^e 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. De code is eenmalig te gebruiken.

Opmaak binnenwerk: Nu-nique grafische vormgeving, Goor

Basisontwerp omslag: Dog & Pony, Amsterdam

Omslagontwerp: Haagsblauw, Den Haag

Beeld omslag: 4Max/Shutterstock

© ter Braake & Boom uitgevers Amsterdam, 2022

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 het overnemen van (een) gedeelte(n) uit deze uitgave in bijvoorbeeld een (digitale) leeromgeving of een reader in het onderwijs (op grond van artikel 16, Auteurswet 1912) kan men zich wenden tot Stichting Uitgeversorganisatie voor Onderwijslicenties, Postbus 3060, 2130 KB Hoofddorp, www.stichting-uvo.nl.

No part of this book may be reproduced in any form, by print, photoprint, microfilm or any other means without written permission from the publisher.

ISBN 9789024441662

ISBN e-book 9789024441679

NUR 173

www.boomstudent.nl

www.boomhogeronderwijs.nl

Inhoudsopgave

Voorwoord	11
1 Inleiding	15
1.1 Wat is Business Intelligence?	15
1.1.1 Definitie	15
1.1.2 Van wie is Business Intelligence?	17
1.1.3 Voor wie is Business Intelligence?	19
1.1.4 Self-service BI	21
1.1.5 Waarom Business Intelligence?	25
1.1.6 Conclusie	26
1.2 Andere terminologie en definities	26
1.2.1 Datawarehouse	27
1.2.2 Data Lake	27
1.2.3 Extract, Transform, Load	28
1.2.4 Staging, ODS	29
1.2.5 Kubussen, modellen	29
1.2.6 Datamining en machine learning	31
1.2.7 Big Data	33
1.3 Tot slot	34
2 Business Intelligence in de organisatie	37
2.1 Business Intelligence-volwassenheidsmodel	37
2.2 De Business Intelligence Life Cycle	42
2.2.1 De grote BI-cyclus	43
2.2.2 De kleine BI-cyclus	47
2.3 Business Intelligence-volwassenheidsmatrix	50
2.4 Business Intelligence implementeren	54
2.4.1 Business Intelligence scope	56
2.4.2 Business Intelligence-doelen	58
2.4.3 Informatiebehoeften	60
2.4.4 Meetwaarde en KPI's	62
2.4.5 Bronnen, architectuur, organisatie	66
2.5 Business Intelligence-projecten	66
2.5.1 Geen gewoon IT-project	67
2.5.2 Informatieanalyse	73
2.5.3 Scrum	74
2.5.4 De Business Intelligence-consultant	77
2.6 Beveiliging	79
2.7 Tot slot	81

3	Waarom een datawarehouse?	83
3.1	Het datawarehouse	83
3.1.1	Waarom een datawarehouse?	84
3.2	Performance van rapporten	85
3.2.1	Normaliseren	86
3.2.2	Indexen	89
3.3	Schrijven van query's	92
3.3.1	Wie maakt de rapporten?	93
3.4	Rapportage-impact op het primaire proces	94
3.4.1	Resource-gebruik	94
3.4.2	Concurrency	96
3.5	Kwaliteitsproblemen	97
3.5.1	Dubbele gegevens	98
3.5.2	Ontbrekende gegevens	99
3.5.3	Foutieve gegevens	101
3.5.4	Inconsistente gegevens	101
3.6	Verschillende operationele systemen	102
3.7	Historische gegevens	104
3.8	Tot slot	106
4	Datawarehouse design	109
4.1	Dimensioneel modelleren	109
4.1.1	Inleiding	109
4.1.2	Modelleren	111
4.1.3	Dimensies	117
4.1.4	Slowly Changing Dimensions (SCD)	127
4.1.5	Conformed dimensions en snowflakes	131
4.1.6	Overige dimensie-overwegingen	133
4.1.7	Feiten	135
4.1.8	Tot slot	139
4.2	Bill Inmon	139
4.2.1	Kritiek op Kimball	139
4.2.2	De Inmon-methodiek	141
4.2.3	Kritiek op Inmon	143
4.2.4	Samenvatting	144
4.3	Data Vault	145
4.3.1	Definitie	145
4.3.2	Hubs, Links and Satellites	146
4.3.3	Voordelen Data Vault	148
4.3.4	Samenvatting	149
4.4	Kimball, Inmon en Data Vault vergeleken	150
4.5	Tot slot	151

5	Big Data en Data Lakes	153
5.1	Big Data	153
5.1.1	Volume	154
5.1.2	Variety	155
5.1.3	Velocity	157
5.1.4	Variability	158
5.1.5	Andere kenmerken	158
5.1.6	Samenvatting	159
5.2	Data Lake	160
5.2.1	Gevolgen voor Business Intelligence	160
5.2.2	Wat is een Data Lake?	161
5.2.3	Architectuur met een Data Lake	163
5.2.4	Structuur van een Data Lake	165
5.3	Tot slot	166
6	Het fysieke datawarehouse	169
6.1	Inleiding	169
6.2	Technisch ontwerp	171
6.2.1	Aantal records	172
6.2.2	Gemiddelde recordlengte	173
6.2.3	Page size	175
6.2.4	Grootte van de database	177
6.2.5	Overige factoren	178
6.2.6	Cloud	179
6.2.7	Grootte van je oplossing	180
6.2.8	Data Lake	180
6.3	Hardware en software	181
6.3.1	CPU-capaciteit	182
6.3.2	Overige hardwarespecificaties	183
6.3.3	Appliances	184
6.3.4	Cloud-overwegingen	185
6.4	Performance features	186
6.5	Tot slot	186
7	ETL	189
7.1	Inleiding	189
7.2	Master Data Management	190
7.2.1	Verschillende soorten gegevens	191
7.2.2	Definities	193
7.2.3	Implementatie	195
7.2.4	Voorbeeld met Microsoft MDS	199
7.2.5	Tot slot	199

7.3	Implementeren van het ETL-proces	199
7.3.1	Resultaten van informatieanalyse	200
7.3.2	Architectuur	203
7.3.3	Documenteren	206
7.3.4	Tools en technieken	209
7.4	ETL, ELT en Big Data	210
7.4.1	Big Data-clusters	211
7.4.2	Big Data ETL-tools	212
7.5	Datakwaliteit	213
7.5.1	Slechte gegevens	213
7.5.2	Data Cleansing	214
7.5.3	Voorbeeld met Microsoft DQS	214
7.6	Scenario's	214
7.7	Tot slot	215
8	Semantische modellen	217
8.1	Inleiding	217
8.1.1	Semantisch model	217
8.1.2	Performance	220
8.1.3	Query-performance en Big Data	222
8.1.4	OLAP query engine	223
8.1.5	Tot slot	224
8.2	OLAP-kubus	224
8.2.1	Wat is een kubus?	224
8.2.2	Waarom kubussen?	226
8.2.3	Aggregaties	229
8.2.4	Model of database?	232
8.2.5	Tot slot	234
8.3	Voorbeeld in-memorymodel	234
8.4	Tot slot	235
9	Het front-end	237
9.1	Inleiding	237
9.2	Analytics	238
9.2.1	Basic Analytics	239
9.2.2	Advanced Analytics	240
9.2.3	Top-down versus bottom-up	242
9.2.4	Tot slot	244
9.3	Machine learning en datamining	244
9.3.1	Definities	244
9.3.2	Machine learning-scenario's	245
9.3.3	Tot slot	249

9.4	Big Data en ethiek	249
9.4.1	Hoe wenselijk is Big Data Analytics?	249
9.4.2	Wet- en regelgeving	250
9.4.3	Tot slot	252
9.5	Wie maakt en wie gebruikt rapporten?	253
9.5.1	Farmers	254
9.5.2	Tourists	255
9.5.3	Explorers	257
9.5.4	Miners	258
9.5.5	Tools en acceptatie	259
9.6	Rapporten, scorecards en dashboards	259
9.6.1	KPI	260
9.6.2	Scorecard	262
9.6.3	Dashboard	264
9.6.4	Rapporten	266
9.6.5	Security	268
9.7	Ontwerpen van (goede) rapporten	269
9.7.1	Inleiding rapportontwerp	269
9.7.2	Attentieve verwerking	270
9.7.3	Rapportrichtlijnen	275
9.7.4	Tabel versus grafiek	283
9.8	Tot slot	285
	Register	287
	Literatuurlijst	291
	Illustratieverantwoording	293

Voorwoord

Voor je ligt de derde 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 om 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.

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? Hoe verhoudt een Data Lake zich tot een 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 via www.boomstudent.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. Daarmee is het ook interessant voor mensen die veel met gegevens in Excel hebben gewerkt en nu de stap willen gaan zetten richting Power BI of een andere self-service BI-tool.

Veel plezier met lezen!

Peter ter Braake
Januari 2022



Verification mmn 45f73

xxx xxxxxx xxx000mm

00--00101---000

tpX



Business Intelligence is een veelgebruikte 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 veelgebruikte 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 is. 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). 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 teneinde de juiste beslissingen te kunnen nemen op basis van correcte, betrouwbare informatie.



Business Intelligence is een paraplueterm 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. Gartner noemt Business Intelligence tegenwoordig niet meer apart, maar zegt dat bovenstaande de definitie is van **Analytics and Business Intelligence**. Analytics is het actief analyseren van gegevens en het communiceren van de resultaten. De eenvoudigste vorm van analytics is het werken met rapportages. Maar zo beschouwd doet elk bedrijf aan Business Intelligence omdat in elke organisatie wel rapporten gebruikt worden. Elk bedrijf heeft immers wel lijstjes met omzetgegevens, ziekteverzuim van medewerkers, gemaakte kosten, enzovoort.

De belangrijkste woorden in de definitie van Business Intelligence 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 van wat Business Intelligence is. Business Intelligence draait erom 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 pas het besluit moeten nemen 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.



Laten we aan de hand van een kort voorbeeld bekijken wat Business Intelligence is en hoe het een bedrijf kan helpen:

Voorbeeld 1.1



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? Maar je klanten denken te kennen en dat kunnen staven met gegevens is nog wel iets anders.

De winkelier zou er erg bij gebaat zijn als hij zijn ervaring (hij denkt zijn klanten te kennen) zou kunnen staven aan beschikbare gegevens. Daarom maakt de winkelier gebruik van een klantenkaart. 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. 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. Hij weet nu wat voor soort mensen zijn klanten zijn en wat voor soort mensen in de wijk wonen. Met deze informatie kan hij/zij een weloverwogen afweging maken. 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. Door deze formulering kan het idee ontstaan dat Business Intelligence een onderdeel is van ICT. De applicaties zijn immers de computerprogramma's waarmee mensen werken. Dat kan Microsoft Excel zijn, een van de meestgebruikte Business Intelligence-applicaties,

maar ook andere programma's zoals rapportageprogramma's. Ook de achterkant (back office) hoort bij die applicaties en infrastructuur. Daarmee worden databases, datawarehouses en kubussen (later in dit boek leer je daar meer over) bedoeld waar we de gegevens in opslaan, zodat we de gegevens kunnen gebruiken om er rapporten mee te maken. 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?) en 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 moeten 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 om 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 kosten-effectiever werken door slimmer in te kopen en/of door de supply chain (hoe, via welk proces, een product van de leverancier bij de klant komt) 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 nemen 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 je project bovenstaand lot beschoren is, wordt groter naarmate de betrokkenheid van de gebruikers kleiner is gedurende het project. De eindgebruikers moeten van het begin af aan betrokken zijn. Sterker: het moet (zoveel mogelijk) hun project zijn.

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



In de bovenstaande uitleg wordt een andere fundamentele fout gemaakt. Er wordt gesproken over Business Intelligence als over een project. Business Intelligence is geen project, het is een competentie, een werkwijze. Het gaat over hoe een proces binnen een organisatie loopt, hoe er gemonitord wordt en hoe er bijgestuurd wordt. Het is een continu aanwezig iets en dus niet een project met een doel, een deadline en een vast omlind einde. Als onderdeel van Business Intelligence worden er wel projecten uitgevoerd. Het invoeren van nieuwe Business Intelligence-software is een project met een deadline. Een multidisciplinair team zou hieraan kunnen werken voor een bepaalde tijd. Als de software succesvol is uitgerold, is het project klaar.

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 het ook zinvol om de vraag ‘voor wie is het eigenlijk?’ 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 marketingfabriek van Microsoft en in dat licht moet je de zin ook plaatsen. Maar 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 om Decision Support Systems te bouwen vaak langdurig en zeer kostbaar. 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 rapporten met 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.

In het huidige cloud-tijdperk geldt bovenstaande nog sterker. Voor enkele tientjes per maand kun je in de cloud al kleinschalige maar zinvolle Business

Intelligence-oplossingen bouwen. Je hebt geen hardware of softwarelicenties meer nodig. Je kunt vandaag nog beginnen. De cloud biedt tegelijkertijd de schaalbaarheid om op grote schaal, dus met grote gegevensverzamelingen en met veel gebruikers, met gegevens te werken. Dat verklaart mede de opkomst van Big Data.

Naast het feit dat Business Intelligence voor alle soorten bedrijven binnen handbereik 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 weleens (of heel vaak) een beslissing nemen. Andere beslissingen hebben een ander soort informatie nodig (veel details 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. Met de realisatie dat Business Intelligence voor en van iedereen is, rijst de vraag wie van de beschikbare gegevens informatie moet maken. Business Intelligence-specialisten of de eindgebruiker van de informatie zelf?

1.1.4 *Self-service BI*

Met de steeds bredere inzet van Business Intelligence binnen steeds meer bedrijven kwam self-service BI opzetten. Bij self-service BI maken gebruikers (deels) zelf wat ze nodig hebben. Self-service BI komt voort uit de noodzaak tot acceptatie van de gebouwde oplossing door de gebruikers. Binnen een Business Intelligence-project zul 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 kan hierbij helpen. Het laat bovendien zien dat Business Intelligence geen project is, maar een competentie die binnen een bedrijf toebehoort aan de werknemers. In het ideale scenario is self-service BI een competentie die werknemers constant gebruiken.

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 vraag ‘Is de Business Intelligence-oplossing die gebouwd wordt wel wat de gebruikers nodig hebben?’ positief te kunnen beantwoorden, moet je een goede informatieanalyse 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

Als we ervan uitgaan dat een Business Intelligence-specialist iets maakt, zoals een rapport of een dashboard, voor een gebruiker, dan is een van de belangrijkste taken het achterhalen van de wensen van die gebruiker ('*requirements gathering*' ofwel informatieanalyse). Waar heeft welke gebruiker wanneer last van? Of in betere woorden: welke nu ontbrekende kennis zou een medewerker in staat stellen om zijn of haar werk beter te doen? Om dit te achterhalen moet je 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 behoeften, 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 hoe het (nog) beter en/of makkelijker zou kunnen;
- 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? Soms zijn mensen uren bezig met werken in Excel-bestanden waar een goede Business Intelligence-oplossing met bijvoorbeeld een tool als QlikView dat volledig overbodig maakt. Sommige mensen voelen weerstand als ze een nieuwe tool moeten leren. Als daarmee dan ook nog een deel van hun werkzaamheden overbodig wordt, zijn ze misschien niet heel erg bereid om mee te werken aan een interview.

Onwetendheid over de mogelijkheden

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

Voorbeeld 1.2



Ik heb ooit iemand horen zeggen dat een query met een doorlooptijd van acht uur die gegevens uit een database haalt, 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 1.3



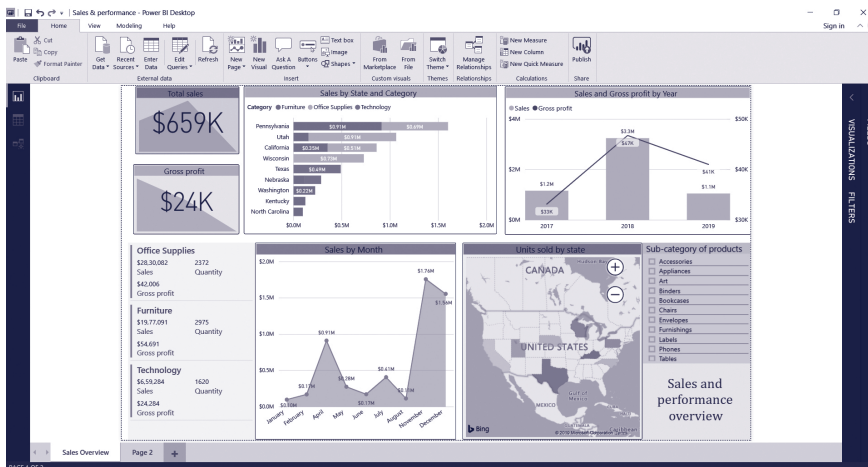
Ik heb ooit in een vergadering gezeten waar een request-for-change (verzoek om een verandering aan te brengen in een bestaande applicatie) besproken werd. Uit de organisatie was een verzoek gekomen om de applicatie op enkele punten te wijzigen. Degene die kwam uitleggen wat ze precies bedoelden met de 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 om uit haar handeling-denken te komen en het proces dat 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/zij iets anders bedoeld. Als dat niet vroegtijdig aan het licht komt, levert de Business Intelligence-oplossing straks de verkeerde informatie op.

Bovenstaande voorbeelden brengen ons terug bij self-service BI. We zien 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, kun 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 zonder een lange en moeilijke informatieanalyse. 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. Maar er zijn ook tools zoals QlikView en Power BI die laagdrempelig genoeg zijn voor veel mensen om zelf mooie dingen mee te bouwen. Je ziet een voorbeeld in figuur 1.1.



Figuur 1.1 Voorbeeld Power BI Sales-rapportage

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 hierboven 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 teneinde de juiste informatie op het juiste moment ter beschikking te hebben.



1.1.5 *Waarom Business Intelligence?*

Tot nu toe heb ik getracht om 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 erom 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 1.4**

Voor een bedrijf is het niet altijd gemakkelijk om 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 je 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 om 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 om 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 Business Intelligence. 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

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

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



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

1.2.2 Data Lake

Met de opkomst van Big Data wordt het datawarehouse steeds vaker vervangen door een Data Lake, soms ook wel Reservoir genoemd. Een Data Lake is flexibeler in het opslaan van gegevens dan een datawarehouse en kan daardoor makkelijker omgaan met verschillende soorten gegevens. Met de opkomst van IoT (Internet of Things), sociale media en het internet in het algemeen, zijn er andere soorten gegevens beschikbaar gekomen die interessant zijn om te analyseren ten behoeve van de besluitvorming dan we voorheen hadden. Een Data Lake is geoptimaliseerd om alle soorten gegevens efficiënt op te slaan. Het grootste verschil tussen een datawarehouse en een Data Lake is dat de laatste gegevens opslaat als bestanden, terwijl een datawarehouse een relationele database is waarin gegevens worden opgeslagen in tabellen.

Een **Data Lake** is een omgeving waar je gegevens verzamelt in hun oorspronkelijke formaat en schema.



In hoofdstuk 5 leer je meer over Data Lakes en Big Data.

1.2.3 Extract, Transform, Load

Datawarehouses en Data Lakes worden gevuld met gegevens uit operationele systemen, meestal bronsystemen genoemd vanuit het BI-perspectief. Deze bronsystemen zijn bijvoorbeeld de database achter een webshop of de database met alle personeelsgegevens. Deze gegevens moeten uit de bronsystemen gehaald worden en dan in het datawarehouse of het Data Lake ingelezen worden. 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.

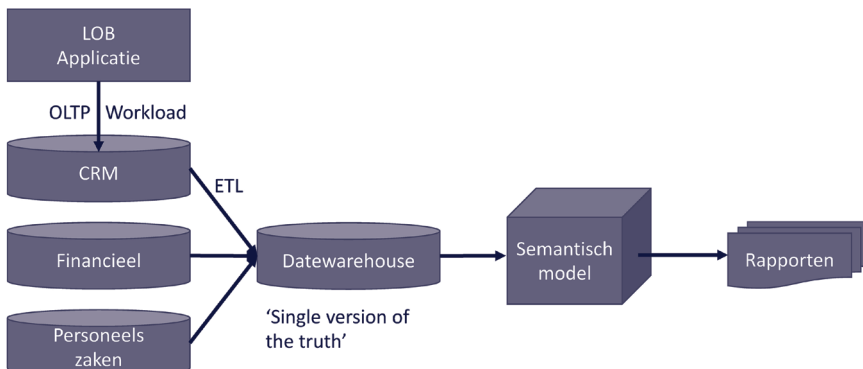
Met name de middelste stap, de transformaties, de aanpassingen die gedaan moeten worden, vormt hier de belangrijke stap. Het ETL-proces zorgt voor de juiste vulling van het datawarehouse en daarmee voor de kwaliteit van de rapportages die je maakt en de analyses die je doet. De aanpassingen die je doet aan de gegevens bepalen deze datakwaliteit. Bij een Data Lake worden de meeste transformaties pas toegepast als de gegevens uit het Data Lake worden gehaald om te gebruiken. Of je nu een datawarehouse of een Data Lake maakt, het ETL-proces vormt altijd een belangrijk onderdeel van het bouwen van een Business Intelligence-oplossing.



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.

In hoofdstuk 7 van dit boek kun je meer lezen over het ETL-proces, welke uitdagingen je tegenkomt, wat voor tools je kunt gebruiken, et cetera.

Figuur 1.2 laat zien hoe gegevens vanuit bronsystemen naar het datawarehouse overgehaald worden teneinde rapportages te baseren op het centrale datawarehouse.



Figuur 1.2 Datawarehouse-architectuur

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

Leerboek Business Intelligence 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 onbelicht blijven. Om de relatie met de praktijk te leggen worden diverse voorbeelden uitgewerkt.

Deze derde druk is geactualiseerd en aangevuld. In deze editie is meer aandacht voor bedrijfskundige aspecten en zijn meer voorbeelden en opdrachten toegevoegd.

Via www.boomstudent.nl heb je toegang tot de online-omgeving van dit boek met meerkeuzevragen en feedback, uitgewerkte voorbeelden en de daarin gebruikte bestanden en scripts. Ook zijn hier het online boek en extra paragrafen te raadplegen.

Leerboek Business Intelligence is geschreven voor studenten van bedrijfskundige studies, maar ook voor sommige informaticarichtingen. Het boek richt zich zowel op technische als niet-technische lezers en is daarom ook geschikt voor professionals in het bedrijfsleven die een beter begrip willen krijgen van Business Intelligence in de breedste zin.

Peter ter Braake is zelfstandig SQL Server en Business Intelligence docent/consultant. Hij is auteur van de boeken *Handboek Power BI* en *Database modelleren*.



9 789024 441662

www.boomstudent.nl
www.boomhogeronderwijs.nl