

Grip op data-architectuur

Grip op data-architectuur

Bert Dingemans

Schrijver: Ir. Ing. Bert Dingemans
Coverontwerp: Rudy Barten
ISBN: 9789465209265
© Ir. Ing. Bert Dingemans

Voorwoord

Heeft een boek specifiek over data-architectuur in de praktijk toegevoegde waarde? Data krijgt op dit moment volop aandacht vanuit allerlei gezichtspunten, zoals technisch, maatschappelijk, management en ethisch. Die vraag is daarom met een volmondig ja beantwoorden.

Of je nu werkt op het vlak van datamanagement, kunstmatige intelligentie, data gedreven werken, enterprise-architectuur of modellering er is altijd een relatie met data. Het structureren van al deze relaties met data en van de gegevens onderling is daarbij het werkveld van de data-architect. Alle activiteiten op het vlak van data-architectuur kunnen daarmee toegevoegde waarde leveren aan een efficiëntere en wendbare gegevensvoorziening.

Dit boek bestaat uit vier delen, in ieder deel wordt vanuit een ander gezichtsveld gekeken naar data en architectuur. Ieder deel bestaat uit een aantal inhoudelijk hoofdstukken en wordt afgesloten met een aantal praktische tips.

In het eerste deel wordt gekeken naar de omgeving van de data-architect. Vanuit de rol van de architect worden stakeholders en concerns behandeld. Verder wordt ingegaan op hoe de omgeving gemodelleerd kan worden om inzicht te krijgen in de onderlinge verbanden binnen de stakeholders en concerns.

In het tweede deel ligt de focus op de wijze waarop de data-architect kaders kan stellen aan de data zoals aanwezig in organisatie. Hierbij wordt uitgegaan van het sturen van verandering binnen projecten. Er wordt uitgebreid stilgestaan bij data-kwaliteiten en hoe deze gebruikt kunnen worden om meetbare kaders te stellen. Daarnaast wordt ingegaan op andere kadervormen zoals patronen, risico's en principes.

Modelleren van data-architectuur is het onderwerp van het derde deel. Op basis van verschillende modelleertechnieken zoals ER-diagrammen, UML en ArchiMate wordt een beeld gegeven van data-architectuurmodellering.

Het vierde deel gaat in op belangrijke producten van de data-architectuur namelijk patronen en bouwblokken. Hierbij is de keuze gemaakt om een beperkt voorbeelden van patronen en bouwblokken in detail te beschrijven.

Het boek is een herziene uitgave van de publicatie “Data-architectuur in de praktijk” verschenen in 2014. In elf jaar is er veel veranderd in het werkveld van de data-architect.

Het boek is beschikbaar in een papieren- en een elektronische versie. Daarnaast is er het ondersteunende website waarop hulpmiddelen, checklists en een register met data-architectuur voorbeelden te vinden zijn.

Zonder hulp van een aantal mensen had dit boek niet de huidige vorm gekregen. Drie personen wil ik met name noemen. Allereerst Huup de Jong voor zijn suggesties over conceptuele data modellering. Ten tweede Yorrick Tillemans voor zijn bijdragen aan de data-architectuur patronen. Als laatste mijn partner Anneke Hubert voor haar steun. Rest mij iedereen veel inzichten over data-architectuur toe te wensen bij het lezen van dit boek.

Culemborg, Augustus 2025.

Inhoud

Voorwoord	5
Inhoud	7
Data-architectuur producten.....	11
Deel 1: Data-architectuur en omgeving	12
Wat is data?	13
Definitie van data	13
Aspecten van data	15
Data, informatie en kennis	16
Data, informatie kennis pyramide	16
Data als productiemiddel.....	17
Data toepassingen	17
Productiemiddel	19
Data Management	21
Data Management Raamwerk	21
DaMa Body of Knowledge	22
Context van data-architectuur	25
Inleiding	25
Definities	25
Gerelateerde werkvelden	27
Waarde van de data-architect	30
Inleiding	30
Activiteiten van de traditionele architect	30
Nieuwe rolinvulling voor de architect	33
De rol van de data-architect	42
Inleiding	42
Veranderen	42
Dimensies van data-architectuur	43
Inventariseren, modelleren, begeleiden	47
Beïnvloedingsstrategieën	49
Projecten	50
Roluitvoering	51

Data-architectuur producten	54
Inleiding	54
Overzicht data-architectuur Producten	55
Vaardigheden van een data-architect	61
Inleiding	61
Competenties data-architect	62
Stakeholders van de data architect.....	74
Inleiding	74
ISO 42010 model	74
Stakeholders bij data-architectuur	76
Overzicht stakeholders data-architectuur	77
“Onzichtbare stakeholders”	83
Stakeholder management	84
Concerns	85
Inleiding	85
Inventarisatie	85
Concerns modelleren	87
Concerns bij de data-architectuur	88
Tips over de waarde van data-architect	92
Inleiding	92
Overzicht van tips	92
Tot slot	95
Deel 2 Data-architectuur kaders	96
Data kwaliteiten	97
Inleiding	97
Dimensies	97
Data kwaliteiten en maatregelen	101
Data Kwaliteit maatregelen werkprocessen	103
Data Kwaliteit maatregelen RDBMS	105
Data Kwaliteit maatregelen software	107
Kwaliteit Maatregelen data integratie	109
Data Kwaliteit Maatregelen user interface	110
Datakwaliteiten en Maatregelen per dimensie	112

Organisatievolwassenheid	147
Data-principes.....	148
Inleiding	148
Doelen en stakeholders en principes	148
Principes als kaders bij veranderingen	149
Opbouw principe	150
Voorbeeld principes	152
Principes en projecten	152
Principe register	158
Andere kadervormen	159
Acceptatiecriteria	159
Risico's	160
Architectuurbesluit	161
Tips voor architectuorkaders	162
Inleiding	162
Overzicht van tips	163
Tot slot	165
Deel 3 Beschrijvende architectuur	166
Beschrijvende data-architectuur	167
Inleiding	167
Data-architectuur modelleren raamwerk	167
Modelleertalen	169
Stakeholders	170
Requirements	171
Verschijningsvormen van data-architectuur modellen	173
Opzet data-architectuur modelleerwijzen	178
Beschrijvende architectuur	181
Kaderstellende architectuur	215
Data management	231
Data platform en toepassingen	245
Tooling & aanvullende producten	262
Evaluatie	262
data-architectuur metamodel	263

Tips voor data-architectuur modelleren	264
Inleiding	264
Overzicht van modelleertips	264
Tot slot	272
Deel 4 Data-architectuur bouwblokken	273
Data-architectuur patronen.....	274
Wat is de IKEA metafoor	274
Architectuur patronen en -bouwblokken	277
Definitie patronen	277
Definitie bouwblokken	278
Patroon en bouwblok register	279
Voorbeelden van patronen	280
Data gedreven werken	280
Data gedreven grondplaat	280
Data integratie	286
Logische datamodellen	295
Fysieke datamodellen	301
Metadata	302
Master en Referentie Data	307
Business Intelligence, DWH en data analyse	325
Tips voor patronen en bouwblokken.....	335
Inleiding	335
Overzicht van data patronen	335
Referenties	338

Deel 1: Data-architectuur en omgeving

In dit eerste deel wordt ingegaan op de omgeving van de data-architect.

Deze omgeving wordt vanuit meerdere gezichtsvelden beschouwd.

Allereerst wordt ingegaan op de verschillende werkvelden gerelateerd aan data-architectuur. Ook de stakeholders van de data-architect worden beschreven. De stakeholders worden gecategoriseerd en er wordt ingegaan op zichtbare en onzichtbare stakeholders. Verder wordt gekeken naar stakeholdermanagement een bijzondere manier van het categoriseren van stakeholders.

Via stakeholders wordt gekeken naar welke concerns of belangen de verschillende stakeholders hebben. Door concerns in kaart te brengen ontstaan toepasbare eisen voor de gewenste situatie binnen de architectuur.

De rol van de architect in relatie tot de omgeving en de belangen die daar bestaan wordt beschreven in een hoofdstuk in dit deel. De manier waarop een architect zijn rol invult heeft gevolgen voor de successen die gehaald worden. Ook is de rol afhankelijk van de volwassenheid en de cultuur van de organisatie. Er dient dan ook een match te zijn tussen rolinvulling en inrichting en cultuur van de organisatie.

Wat is data?

Veel organisaties onderkennen de waarde van data en willen deze data inzetten om meerwaarde te realiseren. Vaak wordt hiervoor de term “data gedreven werken” gebruikt.

We beginnen met de definitie van data en kijken naar de aspecten van data die helpen te ontdekken waar knelpunten kunnen ontstaan als we data op effectieve wijze willen inzetten. Vervolgens kijken we naar hoe we vanuit data via informatie kennis kunnen ontwikkelen die ons wijsheid brengt waarmee we uiteindelijk meer gewogen en betere beslissingen kunnen nemen.

Is data een waardevol productiemiddel voor organisaties, dan zullen we data willen managen. Willen we data managen binnen de verandering van een organisatie dan zal data-architectuur essentieel zijn.

Definitie van data

Data is overal aanwezig en kent daarbij vele verschijningsvormen. Echter wat is data precies? Zoek je op internet dan kom je bijna 20 definities tegen. Ook op Wikipedia wordt een definitie gegeven. Deze definitie staat hieronder en deze gaan we uitdiepen.

Een **gegeven**, ook wel **datum** [wikipedia]:

- Is een constructieve bijzonderheid: het is de vastgelegde uitdrukking van een feit. De woorden worden praktisch altijd in het meervoud geschreven; gegevens en data.
- Gegevens zijn de objectief waarneembare neerslag of registratie van feiten of op een bepaald medium, zodanig dat deze gegevens uitgewisseld en voor langere tijd bewaard kunnen worden. Data zijn dan ook steeds het resultaat van codering.
- Data zijn in die zin abstract, dat je ze steeds kan hercoderen in een ander medium of drager.
- Met deze data wordt er een model (een selectief deel dus) van de werkelijkheid vastgelegd in de tijd. Ofschoon de werkelijkheid nooit stilstaat, kan deze door het vastleggen van de gegevens toch worden bevroren.

Het eerste deel van de definitie gaat in op het aspect dat het een uitdrukking is van een feit en dat data bijna altijd in meervoud voorkomt. Het tweede deel van de definitie is vanuit het perspectief van de waarde van data interessanter. Een waarneembare neerslag of registratie van feiten. Hier zit een bijzonder interessant aspect in dat er blijkbaar een

parallele wereld bestaat binnen de datawereld die op enigerlei wijze een weerspiegeling is van feiten die in de werkelijkheid plaatsvinden. Dit is met name interessant omdat hiermee de data ontstaat rond feiten uit de werkelijkheid en door die data te verzamelen, te bewerken en te gebruiken kan we waarde uit data ontstaan.

Ik wil dat met een aardige metafoor inzichtelijk maken rond data. Stel ik ga met het openbaar vervoer naar de kapper om mijn haar te laten knippen. Ik heb al jarenlang dezelfde kapper en zij weet precies hoe ik het kapsel wil hebben.

Het feit uit de werkelijkheid is daarbij dat ik met het openbaar vervoer reis. In de parallele wereld betekent dit dat ik bij het instappen bij de bus in- en bij aankomst uitcheck. Door die gebeurtenissen in de werkelijkheid ontstaat er data die een weerspiegeling is van het feit in de werkelijk in de data.

Namelijk er vindt registratie plaats waar ik ben in- en uitgecheckt.

Het derde deel gaat in op het feit dat er data verzameld is over mijn reis namelijk twee data concepten rond in en uitchecken. Echter om de data te kunnen gebruiken, bijvoorbeeld over welke kosten gemoeid zijn bij deze reis, moet de data getransformeerd of gehercodeerd worden naar een model op basis waarvan de vervoerder kan berekenen wat deze reis kost.

Echter wil de vervoerder weten hoeveel reiziger er per maand op een bepaald traject zijn op welke dag en op welk tijdstip dan zal er wederom een hercodering plaats moeten vinden om dat inzicht te verkrijgen.

Het laatste deel van de definitie bevat een aantal zeer interessante karakteristieken rond data. Namelijk in de parallele wereld van data zal iedereen die de data gaat gebruiken een datamodel, dus een vereenvoudiging van de werkelijk, maken. Echter dat datamodel zal voor iedere organisatie vanuit een eigen perspectief naar de data kijken. De openbaar vervoerder wil weten waar ik in en uitgecheckt ben op welk moment van de dag en welke leeftijd ik heb in verband met mogelijke kortingen. Echter de kapper wil weten hoe laat ik kom, of mijn haar eerst gewassen moet worden. Welk kapsel ik heb, misschien hoe lang geleden het is dat ik voor het laatst geweest ben. Ook hierbij bestaat mogelijk een model vanuit het perspectief bijvoorbeeld als een kapper een klantregistratie heeft waarover men een aantal kenmerken bijhoudt. Ook dit is wederom een datamodel maar vanwege het andere perspectief een compleet ander model.

Aspecten van data

In de definitie in Wikipedia wordt nog iets dieper ingegaan op data, namelijk op de aspecten van data. Hieronder de vier aspecten van data:

- **Pragmatisch aspect** Het pragmatische aspect van gegevens (de gewenste uitwerking van gegevens) heeft betrekking op de uitwerking die een bepaald signaal heeft op gedragingen van de ontvanger.
- **Semantisch aspect** Het semantische aspect van gegevens heeft betrekking op de betekenis van de gegevens. Bij het uitwisselen van gegevens moeten zender en ontvanger dezelfde betekenis toekennen aan deze tekens.
- **Syntactisch aspect** De overdracht van gegevens gebeurt door middel van tekens of combinaties daarvan. De syntaxis legt vast volgens welke regels men tekens of combinaties van tekens mag of moet gebruiken.
- **Technisch aspect** Onder het technische aspect van gegevens wordt de wijze verstaan waarop gegevens worden vastgelegd, verwerkt of getransporteerd. Het technische aspect handelt over het fysieke medium, betrokken bij dergelijke gegevensmanipulaties.

In deze aspecten van data wordt gewerkt vanuit de informatie theorie van Shannon uit 1948. Namelijk er is een verzender van data naar een ontvanger. Deze zender heeft hierbij een bepaald doel namelijk dat de ontvanger van de data deze gebruikt bijvoorbeeld door begrip te vormen en eventueel door actie te ondernemen. Tussen de verzender en ontvanger kan onduidelijkheid ontstaan doordat er aan beide zijden transformatie plaatsvindt en er ruis kan ontstaan bij de overdracht.

In de vier aspecten van data wordt dit opgesplitst in het gezichtspunt vanuit techniek, syntaxis, semantiek en pragmatiek. En als er in één of meerdere aspecten onduidelijkheid of verwarring is over de data tussen de verzender en ontvanger dan zullen er misverstanden of erger ontstaan.

Vanuit de waarde van dat is dit een belangrijk gegeven. Ook tussen de verzender en de ontvanger van data zal bij onduidelijkheid de waarde van data dalen. Er is tenslotte onduidelijkheid en het is maar de vraag of de ontvanger het juiste begrip krijgt en vervolgens de juiste actie onderneemt.

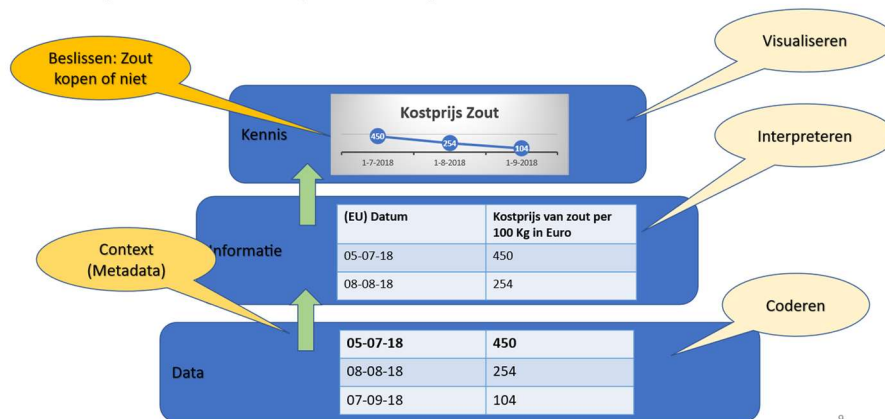
Data, informatie en kennis

Rond data zijn reeds een viertal aspecten behandeld vanuit de techniek uiteindelijk naar het pragmatisch aspect van de data en de invloed van de data op de ontvanger. We kunnen dit aspect nog iets verder uitdiepen vanuit het perspectief van de waarde van data. Namelijk met de data, informatie en kennis pyramide.

Data, informatie kennis pyramide

De data, informatie en kennis pyramide kan het best worden verklaard met een afbeelding die we van onder naar boven in de afbeelding wordt toegelicht.

Data, informatie, kennis, inzicht en beslissen



In de afbeelding zien we een eenvoudig willekeurig voorbeeld. In de onderste laag zien we data. Echter deze data bestaat uit een tabelstructuur met waarschijnlijk een kalenderdatum en een reeks getallen. Dit is gecodeerd op enigerlei wijze. Echter alleen met de data kunnen we aan de ontvangerszijde van de datastroom niets. Er mist namelijk iets essentieels bij deze data.

Er mist context. Context is metadata die je aan de data kunt toevoegen waardoor de metadata iets verteld over deze data. Het blijkt inderdaad over een kalenderdatum te gaan en over een kostprijs. Deze metadata wordt in dit voorbeeld weergegeven als een kolomkop in de tabel. Hiermee is data in combinatie met metadata informatie geworden.

Op basis van deze informatie kunnen we dit gaan interpreteren. De kalenderdatums en de getallen vertellen ons iets en op basis van deze