Cursus Databasedesign

Inhoudsopgave

Inleiding

In deze cursus leer je hoe je een informatiesysteem kunt ontwerpen en bouwen waarmee je gegevens kunt opslaan en beheren op een dusdanige manier dat je er snel en zonder fouten informatie uit kunt halen.

Hoofdstuk 1 Databases

Termen

Data (gegevens)

Informatie

Kennis

Database

SQL databases

NoSQL databases

Relationele database

RDBMS

Records

Collectie

Document

JSON

API

Horizontaal schaalbaar

Verticaal schaalbaar

Grid computing

Cloud computing

Data versus informatie

Data is een ander woord voor gegevens. Gegevens zijn een weergave van feiten. Zij hoeven niet perse nut te hebben. Informatie zijn gegevens of een combinatie van gegevens die wel nut hebben.

Voorbeeld 1 Verkeersbord

Met bovenstaand verkeersbord wil men de verkeersdeelnemer erop wijzen dat deze geen voorrang heeft. Het verkeersbord is dus de weergave van een feit. In de foto hierboven is dit gegeven ook meteen informatie. Staat dit bord echter bij iemand in zijn achtertuin dan is het geen informatie meer maar blijft het wel een gegeven. Hetzelfde geldt als dit bord bijvoorbeeld achter een boom verscholen staat.

Dat dit bord informatie bevat komt ook doordat alle verkeersdeelnemers dezelfde betekenis toekennen aan dit bord. Bij onderstaand bord is dit niet het geval als het in Nederland zou staan.

De manier waarop een gegeven wordt vastgelegd, is ook belangrijk. Het verkeersbord is gemaakt van stevig en duurzaam materiaal. Wanneer het bord gemaakt zou zijn van karton zou het na één flinke regenbui al niet meer leesbaar zijn.

Voorbeeld 2

14	15	16	17	18

In de tabel hierboven is het onduidelijk wat er met de cijfers bedoeld wordt. Er is dus wel data maar geen informatie.

Weekdag van week 17	Ma	Di	Wo	Do	Vr
Temperatuur in ºC	14	15	16	17	18

In de tabel hierboven is dezelfde data nu verheven tot informatie.

Data kan informatie zijn maar informatie kan niet altijd data zijn.

De temperatuur op dit moment is data maar kan ook nuttig zijn en dan is het ook informatie. De gemiddelde temperatuur in een maand is informatie maar is geen data omdat dit getal berekend wordt uit de temperaturen per dag. Tenzij de maandtemperaturen ook zijn opgeslagen. Dan is het ook data.

Kennis

Kennis is wat je nodig hebt om iets te kunnen doen met informatie.

Opgaven

Opgave 1
Leg uit wat het verschil is tussen gegevens en informatie.

Opgave 2
Gaat het bij de volgende voorbeelden om data, om informatie of kan het allebei:

a) De temperaturen op iedere dag in de laatste honderd jaar.
b) Het aantal bezoekers van een website op een bepaalde dag.
c) De hoeveelheid verkeer op de Nederlandse wegen neemt toe.

Opgave 3
Een aantal jaren geleden raakte de Britse belastingdienst een aantal CD's kwijt met daarop de belastingaangifte van 25 miljoen mensen. Op de belastingaangifte staat bijvoorbeeld iemands geboortedatum. Leg uit of hier sprake is van dataverlies en/of informatieverlies.

Opgave 4 Kantine universiteit

Datum	Verkoop aan studenten	Verkoop aan staf	Hamburger/ Taco bar	Pizza Bar	Soep/ Salade Bar
15-02-11	497	23	335	122	63
15-03-11	440	19	285	126	48
15-04-11	447	30	301	126	50
15-05-11	442	27	325	107	37
15-06-11	330	12	229	83	30

a) Waar gaat dit rapport over?
b) Welke gegevens(data) zijn verzameld?
c) Welke informatie kun je uit de tabel halen?
d) Waarvoor denk je dat deze gegevens gebruikt worden?
e) Verzin minstens twee conclusies op grond van de gegevens hierboven.
f) Verzin minstens twee vragen die je zou willen vragen over de gegevens hierboven.

Opgave 5
Bekijk het filmpje op
https://www.youtube.com/watch?time_continue=262&v=sdzUfHwNCVQ
om te leren wat het verschil is tussen data, informatie en kennis. In het filmpje wordt een verhaal verteld over een fabriek.

a) Wat is in dit verhaal de data?
b) Leg uit waarom het data is.
c) Wat is in dit verhaal de informatie?
d) Leg uit waarom het informatie is.
e) Wat is in dit verhaal de kennis?
f) Leg uit waarom het kennis is.
g) Noem vijf manieren om kennis te vergaren in het geval van de fabriek.

Databases

Een database moet je zien als een soort kast met vakjes waarbij in ieder vakje bepaalde data ligt opgeslagen.

Bij het ontwerp van een database zijn de volgende twee eisen erg belangrijk:

1. De informatie moet er snel uitkomen.
2. Er moet zoveel mogelijk informatie uit te halen zijn.

We behandelen hier twee soorten van databases:

1. SQL Databases waarvan de relationele databases de belangrijkste zijn.
2. NoSQL Databases.

Relationele databases

Tabel: KLANT

klantnummer	klantnaam	plaats	adres	rekeningnummer
100345	P. Knobbel	Breda	Kerkstraat 4	NL07INGB0005261370
100783	K. Slijm	Breda	Zwaanstraat 56	NL06ABNB000985234
110432	R. Zwabber	Roosendaal	Smidstraat 18	NL03ASNB000975438

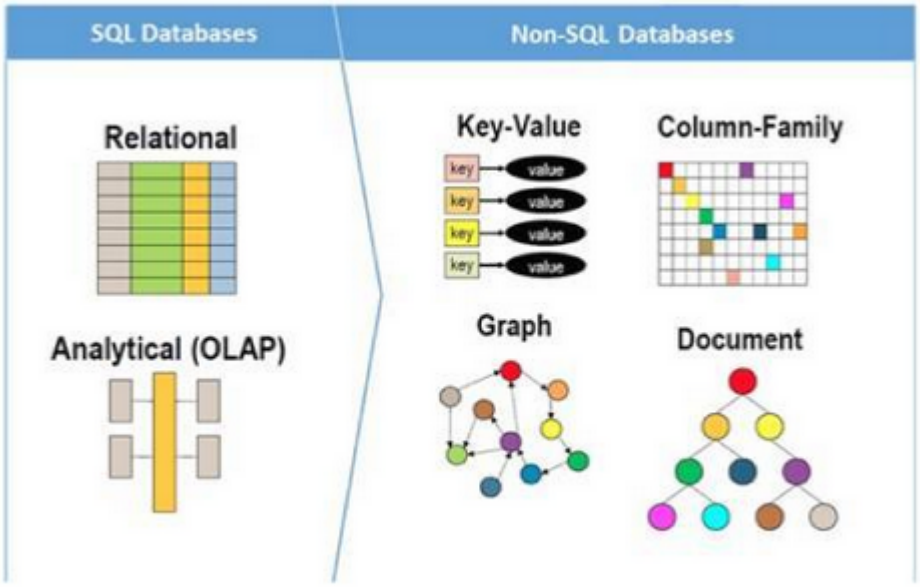
1. Een relationele database is een verzameling tabellen.
2. Een tabel is een verzameling rijen.
3. Een rij is een verzameling kolomwaarden.
4. Een record is een verzameling kolomwaarden met bijbehorende veldnaam.

Hieronder staat een voorbeeld van een record.

klantnummer	klantnaam	plaats	adres	rekeningnummer
110432	R. Zwabber	Roosendaal	Smidstraat 18	NL03ASNB000975438

In dit type database is SQL de standaardtaal voor het relationeel database managementsysteem RDBMS. Hiermee kan de data gestructureerd worden ingevoerd.

NoSQL Databases



De definitie van NoSQL (Not only SQL) geeft aan dat NoSQL databases naast SQL ook met andere talen kunnen omgaan die de database kunnen bevragen en aanpassen. Dit brengt grote voordelen met zich mee voor data die niet op de SQL manier beschreven kan worden. Hierdoor vormen NoSQL Databases een aanvulling op relationele databases. Er zijn verschillende types NoSQL databases.

Bij een NoSQL database wordt de data niet opgeslagen in tabellen maar in verschillende structuren zoals bijvoorbeeld documenten en collecties.

1. Een collectie is een verzameling documenten.
2. Een document is een gestructureerd tekstbestand. De bekendste formaten hiervoor zijn XML en JSON.

In een JSON bestand wordt de data opgeslagen aan de hand van veldnamen en waardes. Zo kan een veldnaam heel snel aangepast worden en hoeft, zoals bij een relationele database, niet de tabelstructuur worden aangepast. De data zit dus in één document en is niet langer verdeeld over verschillende tabellen, wat bij een RDBMS (relationeel database managementsysteem) het geval is.

Hieronder staat een voorbeeld van JSON tekstbestand van een lijst met twee elementen:

```
[ {
    "Naam": "JSON",
    "Type": "Gegevensuitwisselingsformaat",
    "isProgrammeertaal": false,
    "Zie ook": [ "XML", "ASN.1" ]
  },
  {
    "Naam": "JavaScript",
    "Type": "Programmeertaal",
    "isProgrammeertaal": true,
```

```
    "Jaar": 1995
  }
]
```

Hieronder zie je de een voorbeeld van een key-value opslag in een document waarbij “id” de key is.

```
{
        _id: ObjectID('4bd9e8e17cefd644108961bb'),
        name:'Vivek',
        class : '12th',
        subjects: [ 'physics', 'chemistry', 'math', 'english', 'computer'],
        address: {
                                        house_no: '12B',
                                        block: 'B',
                                        sector: 12,
                                        city : 'noida',
                                        },
        grade: [
                                        {
                                        exam: 'unit test 1',
                                        score: '60%'
                                        },
                                        {
                                        exam: 'unit test 2',
                                        score: '70%'
                                        }

                                ]
}
```

Voordelen NoSQL:

- Veelzijdig want stelt weinig eisen aan de organisatie en structuur van de data.
- Eenvoudig in gebruik want via API aan te sturen.
- Goedkoper
- Horizontaal schaalbaar. D.w.z. als er meer moet worden opgeslagen, dan voeg je meer computers toe. Dit kan onmiddellijk als het nodig is. Een relationele database is verticaal schaalbaar. Dat betekent dat deze alleen maar uitgebreid kan worden op dezelfde computer. Je hebt dan een computer

nodig met meer power (CPU, RAM) wat duurder is en ook niet één, twee, drie geregeld is.

Nadeel:
- Onnauwkeurigheid.

Wanneer relationele database en wanneer NoSQL database
Als een applicatie bijvoorbeeld verkooporders verwerkt en absoluut zeker moet zijn van de transacties, dan kun je het beste een relationele database met ondersteuning voor ACID-transacties gebruiken. Maar als het gaat om miljoenen gebeurtenissen in korte tijd, zoals bijvoorbeeld de analyse van clickstreams op een site, om een online sales catalogus te optimaliseren, dan kun je prima kiezen voor een NoSQL-database. Het missen van een paar van deze gebeurtenissen is namelijk niet kritisch en stelt je in staat een schaalbare applicatie met gedeelde data te maken, wat uiteindelijk flink in de kosten kan schelen en een snelle responstijd oplevert.

Opgaven

Opgave 6
Zoek op wat ACID transacties zijn. Niet alleen het volledige woord per letter maar leg ook uit wat ermee wordt bedoeld. Gebruik hiervoor meerdere websites ter vergelijking; ook Engelstalige.

Opgave 7
Zoek op en leg uit wat een API is.

Opgave 8
Welk type database is meer geschikt voor Twitter; een relationele database of een NoSQL database.

Opgave 9
Welk type database is meer geschikt de betalingen bij de ING bank; een relationele database of een NoSQL database.

Geschiedenis van de database

1960
Computers komen in gebruik bij bedrijven. Opslag van veel gegevens wordt mogelijk.

1970
Databases op mainframes met domme terminals.

1980
Eerste commerciële relationele databases.

1995
Het internet wordt gebruikt om data op servers op te slaan. Clients kunnen gegevens opvragen.

2000
Eerste NoSQL (Not only SQL) databases. Voor databases wordt het mogelijk om gebruik te maken van een grid-structuur. Bij Grid computing kunnen alle computers en servers binnen een netwerk gezamenlijk voor één taak worden gebruikt. Het internet is in feite één grote GRID.

2010
Opkomst van Big Data en de data analyse hiervan. Steeds meer wordt gebruik gemaakt van Cloud computing. Hierbij is niet alleen de data verspreid maar bevindt zich ook de gebruikte applicatie vaak elders op het internet.

Opgaven

Opgave 10
Rond welk jaar ontstonden de eerste commerciële relationele databases?
A 1970 B 1980 C 1995 D 2000

Opgave 11
Rond welk jaar ontstonden de eerste NoSQL databases?
A 1970 B 1980 C 1995 D 2000

Opgave 12
Met SETI@home kunnen vrijwilligers over de hele wereld een programma op hun computer installeren waarmee ze mee kunnen helpen met de zoektocht naar intelligent buitenaards leven. SETI gebruik een programma waarbij de analyse van een enorme hoeveelheid data verspreid kan worden over meerdere computers.
Zoek op of hierbij sprake is van:
A Grid computing
B Cloud computing
C Cluster computing
D Distributed computing

Hoofdstuk 2 Database ontwerpen

Termen

Consulting

Informatie analist

Proces analist

Logisch ontwerp

Fysiek ontwerp

Mapping

Entiteit

Attribuut

Waarde

Null

Datatype

Vluchtig (Engels: volatile)

Verplicht (Engels: mandatory)

Optioneel

primary UID (unieke identifier)

secondary UID

PK (primairy key)

Inleiding

Bij het ontwikkelen van een informatiesysteem kan men vier fasen onderscheiden:

1. Analyse
2. Logisch ontwerp
3. Fysiek ontwerp
4. Bouw

Analyse

Consulting
Om een informatiesysteem te kunnen ontwerpen moet je gesprekken voeren met de klant waarin deze zijn/haar wensen kenbaar kan maken. Dat is lastiger dan het lijkt

want de klant heeft er vaak geen idee van hoe een database werkt. Twee dingen zijn hierbij belangrijk:

- Stel de juiste vragen aan de klant.
- Stel ook de vragen waar de klant niet aan denkt.

Het beroep wat hier bij hoort heet consultant. Zij moeten de vertaalslag maken van wat de klant wil, naar degenen die het uiteindelijk moeten gaan maken. Mensen die dit goed kunnen zijn schaars en worden zeer goed betaald.

Bij de analyse zijn twee soorten consultanten nodig:

- De informatie-analist: Deze houdt zich vooral bezig met het ontwerpen van de database.
- De proces-analist Deze houdt zich vooral bezig houdt met welke processen er in het bedrijf spelen en wat er op de verschillende gebruikersschermen zichtbaar moet worden.

Bij kleine projecten is er één consultant die beide doet.

Logisch ontwerp

Vanaf dit punt gaan we uit van het ontwerp van een relationele database.

Een logisch model:

- modelleert functionele en informatieve behoeftes
- is gebaseerd op huidige behoeftes en houdt alvast rekening met toekomstige behoeftes
- gaat enkel over business needs, heeft dus niets van doen met de implementatie
- noemen we ook wel een Entity relationship Model (ERM), zie hoofdstuk 3.
- wordt getoond met een Entity Relationship Diagram (ERD), zie hoofdstuk 3.
- nodigt uit tot discussie
- voorkomt fouten en misvattingen
- vormt meteen een ideaal documentatiesysteem
- geeft een goede basis voor een fysiek database design
- documenteert de business rules (de processen van een organisatie)
- houdt rekening met regels en wetten
- is niet gericht op een bepaald soort techniek (DBMS, platvorm)
- gebruikt voor het bedrijf begrijpelijke benamingen

Fysiek ontwerp

Op basis van het logisch ontwerp wordt het fysiek ontwerp gebouwd. In tegenstelling tot het logische model is het fysieke model wel gericht op één van de belangrijkste DBMS-typen relationeel, hiërarchisch, netwerk of object-georiënteerd.

Een fysiek model:

- richt zich op een bepaald type DBMS
- bevat bij een relationele database de structuur van de tabellen
- bevat alle kenmerken van de gegevens
- richt zich ook op de details van de implementatie

Entiteiten, attributen en waarden

Een logisch ontwerp bestaat uit entiteiten en attributen. In een fysiek ontwerp komt dit overeen met de tabellen en de kolomnamen.

Een entiteit is het object waarvan informatie wordt opgeslagen.

Een attribuut is een eigenschap.

Logisch ontwerp

Uiteindelijke tabel

Leerlingen			
voornaam	achternaam	leeftijd	telefoonnr
Piet	Konijn	13	0611211211
Jan	Haas	15	
Sylvia	Slak	14	0600700700

In het voorbeeld hierboven is leerling een entiteit en leeftijd een attribuut. Wanneer je de database gaat bouwen, vul je de rijen van de tabellen met waarden (Engels: values). Konijn is dus een waarde.

Een leeg vakje vakje betekent dat er geen waarde is. Je noemt de waarde dan null.

Null betekent dus geen waarde. Dat is wat anders als het getal 0 want dat is wel een waarde.

Eén regel is een record.

Wanneer je een entiteit tekent moet je aan bepaalde regels voldoen:

1. Gebruik rechthoeken met afgeronde hoeken.
2. De entiteitnaam is in hoofdletters en enkelvoud
3. De attribuutnamen zijn in kleine letters

Fysieke ontwerp van tabel leerlingen			
Kolomnaam	**Datatype**	**Lengte**	**Standaardwaarde**
v_naam	varchar	20	
a_naam	varchar	30	
leeftijd	tinyint		
tel_nr	char	10	

Hierboven zie een voorbeeld van een fysiek model van de tabel leerlingen. Er zijn meer kolommen maar deze volgen verderop in het verhaal.

De kolomnamen zijn wat afgekort om de tabellen overzichtelijk te houden. In het fysieke model wordt aan iedere kolomnaam ook een datatype (format) toegevoegd.

Een datatype is het soort data dat wordt opgeslagen. Dit kan zijn een getal (integer, kommagetal), string (varchar, char), datum (date), plaatje, geluid enzovoort.

Met de lengte bedoelen we bij een string uit hoeveel tekens deze maximaal mag bestaan.

Met de standaardwaarde bedoelen we dat je standaard al een waarde hebt ingevuld. Wanneer je een kolomnaam “nationaliteit” hebt en bijna alle klanten hebben de Nederlandse nationaliteit dan zou je hier “NL” kunnen invullen. Slechts enkele klanten hoeven dit veld dan aan te passen in hun formulier.

Vluchtige attributen

Niet alle eigenschappen zijn even geschikt om te gebruiken in een database. In bovenstaand voorbeeld is bijvoorbeeld de eigenschap “leeftijd” gebruikt. Dat is niet handig want dan moet je dit na iedere verjaardag bijwerken. Zulke eigenschappen heten “vluchtige” eigenschappen. De oplossing is om “leeftijd” te vervangen door de niet vluchtige eigenschap “geboortedatum”. De leeftijd kan hiermee berekend worden.

Verplicht of optioneel

Sommige attributen moeten altijd worden ingevuld. In het voorbeeld hierboven is dat bijvoorbeeld “achternaam”. Dit zijn verplichte attributen. Andere zijn optioneel en dus niet verplicht. In het voorbeeld hierboven is dat bijvoorbeeld telefoonnummer. Niet iedereen heeft een telefoonnummer of dit nummer is geen noodzakelijk gegeven.

UID Unieke identifier

In een database wil je perse dat ieder record uniek is. Geen enkele regel mag exact hetzelfde zijn. Om te garanderen dat dit het geval is maakt men gebruik van unieke identifiers. Dit kan een attribuut zijn of combinatie van attributen en zelfs een relatie kan deel uitmaken van een UID. In het voorbeeld hierboven kun je denken aan “achternaam”. Op een school zitten echter verschillende leerlingen met dezelfde achternaam. Een betere UID zou zijn de combinatie van voor en achternaam maar het komt regelmatig voor dat leerlingen dezelfde voor en achternaam dragen. Een goede UID zou zijn de combinatie van voornaam, achternaam en geboortedatum. De kans dat twee verschillende leerlingen dit hetzelfde hebben is zeer klein. Om alle risico’s uit te sluiten voegt men echter vaak een attribuut toe. In dit geval “leerlingnummer”. Door ieder jaar andere nummers te gebruiken is persoonsverwisseling uitgesloten.

In het fysieke model heet de UID de primairy key (PK).

De entiteit LEERLING zou er als volgt uit kunnen zien:

= primary UID, dit is de UID die uiteindelijk gebruikt gaat worden.

(#) = secondary UID, deze UID heeft men als reserve en wordt gebruikt als controlemiddel.

* = verplicht

o = optioneel

Opgaven

Opgave 1

Zet de volgende vier fasen van het ontwikkelen van een informatiesysteem op volgorde van begin naar eind:

A Fysiek model B Bouw C Analyse D Logisch model

Opgave 2

Waar of niet waar: Een logisch model houdt rekening met het type DBMS waarmee de database gebouwd gaat worden.

Opgave 3

Welke vier DBMS-typen kennen we?

Antwoord:

Relationeel, hiërarchisch, netwerk of object-georiënteerd.

Opgave 4

Voetbalclub	Stad	Aantal gespeelde wedstrijden	Aantal punten
FC Emmen	Emmen	3	0

Waar of niet waar:

De waarde van het aantal punten in de record hierboven is null.

Opgave 5

Hieronder staat een verhaaltje. Met dit verhaaltje wil je een database ontwerpen. Geef aan welke woorden uit dit verhaal je als entiteit zou gebruiken, welke als eigenschap en welke als waarde? Noem ook de eigenschappen die niet in de tekst genoemd staan maar waarvan wel waarden zijn genoemd.

Bij autohandel “Krakkemik” staan er auto's van verschillende merken op het terrein. Zo staat er een rode Opel Astra uit 1998 voor 1500 euro, een groene Ford Escort uit 2002 voor 2000 euro en een Peugeot waarvan de prijs 4000 euro is.

Opgave 6
Teken de entiteit die bij de opgave uit de vorige vraag hoort. Bedenk zelf de attributen die niet in de tekst voorkomen. Geef aan welke attributen tot de UID behoren, welke verplicht zijn en welke optioneel.

Opgave 7
Geef bij ieder attribuut uit de vorige opgave aan welk datatype en lengte je zou gebruiken in het fysieke model. Op https://www.techonthenet.com/mariadb/datatypes.php vind je een lijst met datatypes waaruit je kunt kiezen. Als deze pagina niet meer werkt zoek je op “datatypes mysql”.

Opgave 8
Leg uit waarom in iemands paspoort niet diens haarlengte staat opgegeven.

Hoofdstuk 3 Relaties

Termen

Relatie

Kardinaliteit

Erdish

Kraaienpootnotatie

ERD

ERM

Constraint

Overdraagbaarheid (transferability)

Niet aanpasbaar (not updatable)

Bedrijfsapplicatie

Module

CRUD analyse

Relaties

Wanneer er sprake is van twee of meer entiteiten dan kunnen deze entiteiten een relatie hebben.

Relaties:

- tonen iets belangrijks
- geven aan hoe entiteiten zich tot elkaar verhouden
- komen in paren; van A naar B en van B naar A
- hebben een mate van kardinaliteit

Kardinaliteit wil zeggen: hoeveel?

Erdish

De relaties schrijven we op een bepaalde manier op zodat we ze later makkelijk in het ERD kunnen tekenen.

Voorbeelden:

Iedere werknemer heeft één of meer banen.

Iedere baan wordt door precies één werknemer vervuld.

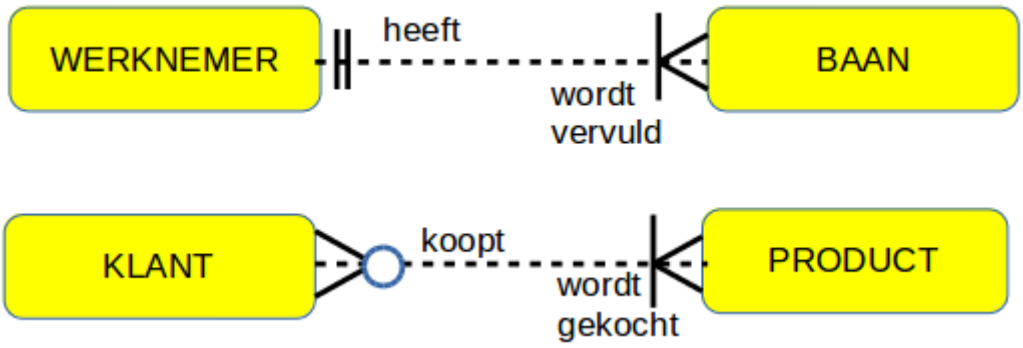
Iedere klant koopt één of meer producten.

Ieder product wordt gekocht door nul of meer klanten.

Toelichting:

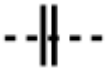
- Iedere zin begint met “iedere” of “elke”.
- De zinnen komen in paren.
- Er zit een werkwoord in de zin.
- Eén of meer, precies één en nul of meer geeft de kardinaliteit aan.
- Vaak moet met de opdrachtgever besproken worden hoe het precies zit. Is het bijvoorbeeld toegestaan dat een werknemer meerdere banen heeft.

Tekenafspraken



Er bestaan meerdere notaties maar wij gebruiken de kraaienpoot notatie omdat die het meest intuïtief is.

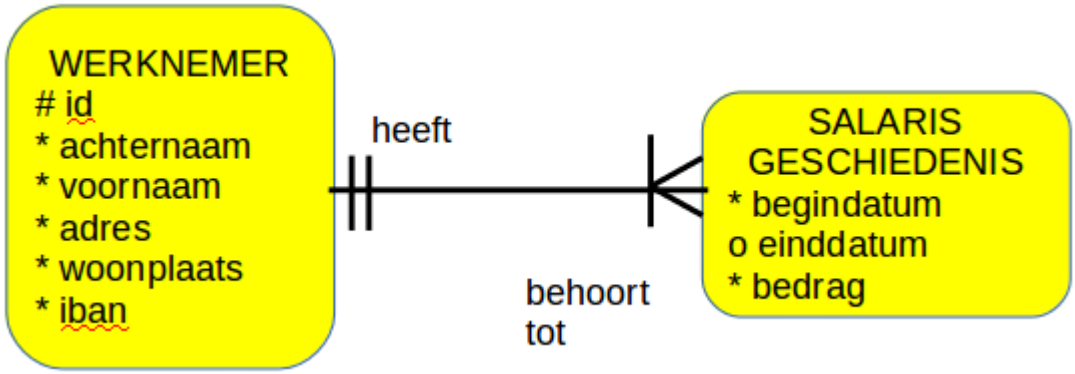
 Nul of één

 Precies één

Nul of meer

 Eén of meer

Hierbij erft de kind entiteit de UID van de ouder entiteit; hieronder staat een voorbeeld. Een nadeel van deze notatie is dat je niet in één oogopslag kunt zien wat de ouder entiteit is en wat de kind entiteit.



- Hierboven geeft de doorgetrokken lijn aan dat de UID van de ouder entiteit WERKNEMER wordt overgedragen op de kind entiteit SALARISGESCHIEDENIS.
- De UID van SALARISGESCHIEDENIS is dus werknemer_id.
- Merk op dat de UID van SALARISGESCHIEDENIS hier dus uit een relatie bestaat.
- In het fysieke model komt hiervoor een extra kolom werknemer_id in de tabel SALARISGESCHIEDENIS.

ERD

Wat je hierboven getekend ziet is een eenvoudig ERD, een Entity Relatitonship Diagram. Een ERD geeft de entiteiten weer die van belang zijn en de relaties die tussen deze entiteiten bestaan.

- Het doel van een ERD is om een voorstel te documenteren waarover discussie kan plaatsvinden.
- Informatie mag slechts één keer worden getoond in een ERD,

- Informatie die van andere informatie kan worden afgeleid moet je niet in het model stoppen. Wanneer bijvoorbeeld geboortedatum genoemd wordt hoef je niet nog eens leeftijd te noemen want die kan uit geboortedatum berekend worden.

ERM

Een ERM Entity Relationship Model is een logisch model voor relationele databases.

- Een ERM bevat meestal een ERD maar niet altijd.
- Bevat ook informatie die niet in de ERD opgenomen kunnen worden zoals datatypen en bepaalde constraints (beperkingen).
- Is onafhankelijk van de implementatie van hardware en software.

Constraints

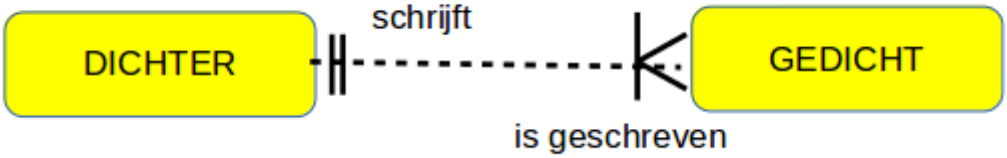
Constraints zijn beperkingen of voorwaarden die voortvloeien uit de bedrijfsregels. Stel dat een bedrijf wil dat het informatiesysteem een seintje geeft als een werknemer 25 jaar bij het bedrijf werkt, zodat er iets georganiseerd kan worden, dan moet dit na de bouw van de database erbij worden geprogrammeerd. Dit staat dus niet in de ERD maar wordt wel gedocumenteerd in het ERM.

Andere voorbeelden zijn: Er mogen geen geboortedata mogen worden ingevoerd die in de toekomst liggen. De begindatum moet voor de einddatum liggen. Een afdeling moet uit minimaal drie werknemers bestaan. Dit kun je voor elkaar krijgen door wat extra programmeerwerk.

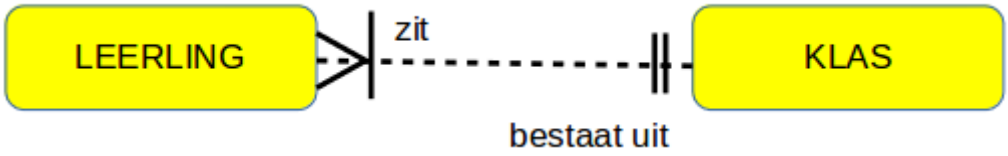
Andere constraints zoals welke attributen tot de UID behoren kun je wel in de ERD opnemen.

Overdraagbaarheid

Sommige relaties zijn niet overdraagbaar (transferable). Het komt alleen voor aan de “meer” kant van één meer relaties. Dit moet in de het ERM gedocumenteerd worden. In het fysieke model maak je vervolgens de relatie aan die kant niet aanpasbaar (not updatable). Bekijk het voorbeeld hieronder.



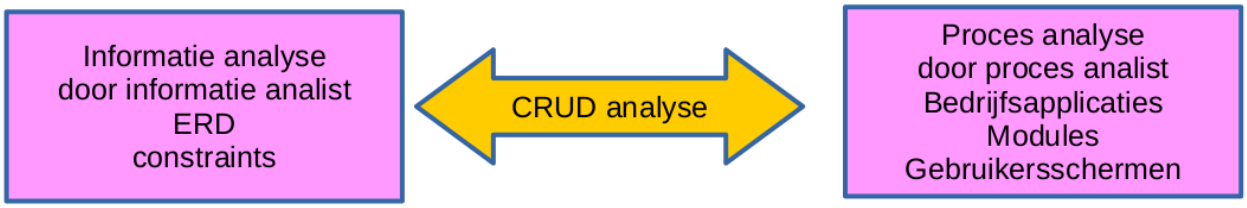
Een gedicht dat eenmaal door een bepaalde dichter is geschreven kan later niet door een andere dichter zijn geschreven. Er moeten maatregelen genomen worden zodat dit niet aangepast kan worden. De relatie van gedicht naar dichter is dus niet overdraagbaar. De relatie van dichter naar gedicht is wel overdraagbaar omdat de dichter meerdere gedichten heeft geschreven en hij dus het ene gedicht kan uitwisselen met het andere.



De relaties hierboven zijn beiden overdraagbaar want een leerling kan achteraf naar een andere klas worden overgeplaatst.

CRUD analyse

Een ERD zal uiteindelijk leiden tot een verzameling tabellen waarin we gegevens als rijen gaan opslaan (de database). Een database vormt de basis van een bedrijfsapplicatie. Een bedrijfsapplicatie bestaat uit verschillende modules. Bijvoorbeeld een module “Klant gegevens onderhouden”.



Een ander onderdeel van het ERM is de CRUD analyse per module van de bedrijfsapplicatie. Genoteerd moet worden welke data in die module gemaakt, opgehaald, aangepast en verwijderd moet kunnen worden.

Create	Welke data wordt aangemaakt?
Retrieve	Welke data wordt opgehaald?
Update	Welke data wordt ververst?

Delete	Welke data wordt gewist?

Opgaven

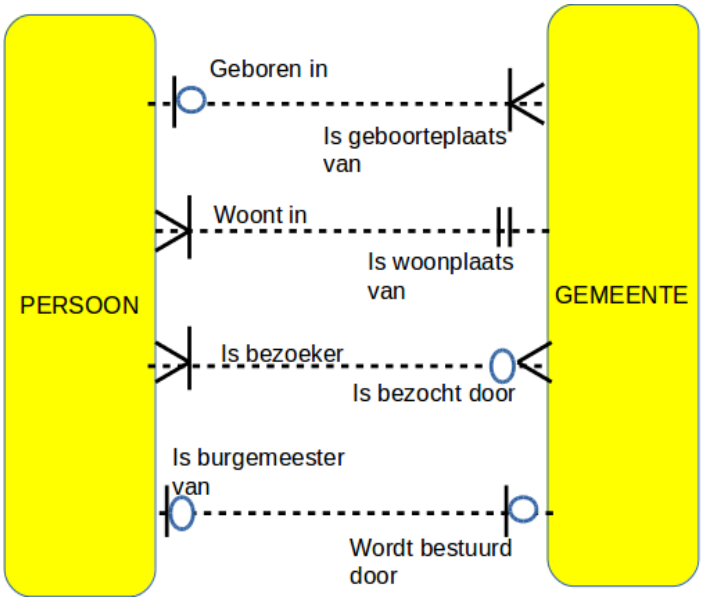
Opgave 1 Erdish zinnen maken

Maak steeds twee Erdish zinnen over de relaties tussen de twee entiteiten.

a) Leerling Klas
b) Leerling Stoel
c) Fototoestel Foto
d) Gedicht Dichter

Opgave 2 ERD lezen

a) Schrijf de Erdish zinnen op bij onderstaande relaties in het ERD.
b) Er staan een paar relaties tussen die in werkelijkheid niet kloppen. Schrijf achter iedere zin of de relatie kan of niet kan is.



Opgave 3 Tekenafspraken

Teken ERD's bij de volgende relaties.

a) Iedere woonplaats is de geboorteplaats van nul of meer personen. Iedere persoon is geboren in precies één woonplaats.
b) Iedere kamer herbergt nul of meer gasten. Iedere gast logeert in precies één kamer.
c) Iedere werknemer werkt op precies één afdeling. Iedere afdeling heeft één of meer werknemers.
d) Iedere e-mail is gericht aan één of meer personen. Ieder persoon is de geadresseerde van nul of meer e-mails.
e) Ieder stuk gereedschap heeft precies één prijs. Iedere prijs hoort bij één of meer stukken gereedschap.
f) Ieder kind heeft precies één moeder. Iedere moeder heeft één of meer kinderen.
g) Iedere leerling heeft les van één of meer docenten. Iedere docent geeft les aan één of meer kinderen.

h) Iedere vingerafdruk behoort tot precies één persoon. Ieder persoon heeft één of meer vingerafdrukken.

Opgave 4 Overdraagbaarheid
Geef aan welke van de relaties uit de vorige opgave niet overdraagbaar zijn.

Opgave 5 Crud analyse: Create, Retrieve, Update, Delete
In computerprogrammaś komen allerlei handelingen voor die allemaal te herleiden zijn tot één van de vier CRUD acties. Zet het juiste CRUD letter achter iedere term.

a) Alter
b) Bring up
c) Change
d) Discard
e) Enter
f) Find
g) Import
h) Input
i) Load
j) Look up
k) Modify
l) Print
m) Purge
n) Read
o) Record
p) Remove
q) Report
r) Trash
s) View

Hoofdstuk 4 De weg naar een goede ERD

Termen

Matrix

Normalisatie

Eerste normaalvorm

Tweede normaalvorm

Derde normaalvorm

Meer meer relaties (Engels: many to many relationship)

Intersection entiteit

Matrix

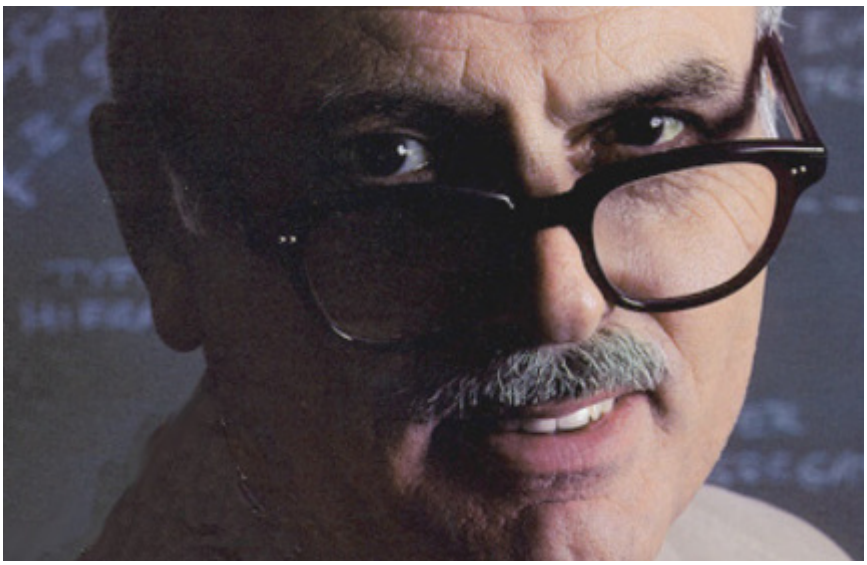
Het komt vaak voor dat een ERD uit meer dan twee entiteiten zal bestaan. In de analyse komen er allerlei kandidaat entiteiten naar voren. Alleen entiteiten waartussen een relatie bestaat worden in de ERD opgenomen. Een goed hulpmiddel hierbij is een matrix. Hieronder staat een voorbeeld.

	REIZIGER	LAND	BEZIENSWAARDIGHEID	BETAALMIDDEL
REIZIGER	x	bezoekt	wordt bezocht	betaalt met
LAND	wordt bezocht	x	bezit	x
BEZIENSWAARDIGHEID	wordt bezocht	ligt in	x	wordt betaald met
BETAALMIDDEL	wordt gebruikt	x	wordt gebruikt	x

Een matrix helpt bij het vinden van alle mogelijke relaties tussen entiteiten.

Normalisatie

Bij het bepalen van de juiste entiteiten en attributen kan normaliseren een belangrijk hulpmiddel zijn. Hiervoor zijn door de Amerikaan Ted Codd regels opgesteld waarvan hier de drie belangrijkste worden toegelicht. Stapsgewijs helpen ze om belangrijke fouten in het ontwerp te ontdekken.

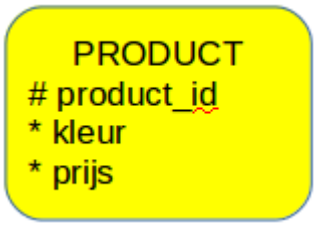


Ted Codd

Eerste normaalvorm

Een entiteit is in de eerste normaalvorm als de attributen niet meer dan één waarde kunnen hebben en de attributen zelf ook maar één keer voorkomen.

Voorbeeld



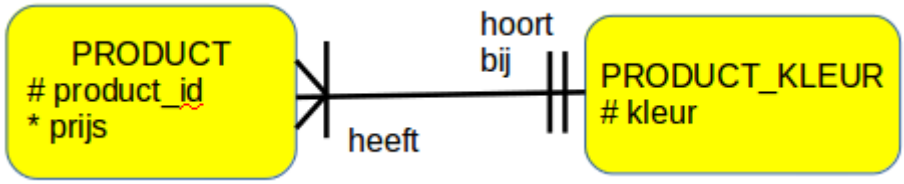
Hier lijkt niets mee aan de hand te zijn totdat je in de analyse een proeftabel laat invullen door de opdrachtgever. Deze vult dan bijvoorbeeld onderstaande tabel in:

PRODUCT

Product_id	Kleur	Prijs
1	rood of groen of blauw	7,99
2	geel	12,99
3	groen	5,49
4	rood of blauw	24,99
5	rood	14,99

Blijkbaar kan een product met hetzelfde product_id meerdere kleuren hebben. Dit is op meerdere manieren op te lossen: je kunt het product met een andere kleur een

andere product_id geven of je maakt een aparte entiteit PRODUCT_KLEUR. De laatste oplossing zie je hieronder. De UID van product is nu een combinatie van product_id en kleur. De doorgetrokken lijn geeft dit aan.



Hieronder staat een voorbeeld van de tabellen die hier uit voortvloeien:

PRODUCT

Product_id	Prijs	Kleur
1	7,99	rood
2	7,99	groen
3	7,99	blauw
4	12,99	geel
5	5,49	groen
6	24,99	rood
7	24,99	blauw
8	14,99	rood

PRODUCT_KLEUR

Kleur
rood
geel
groen
blauw

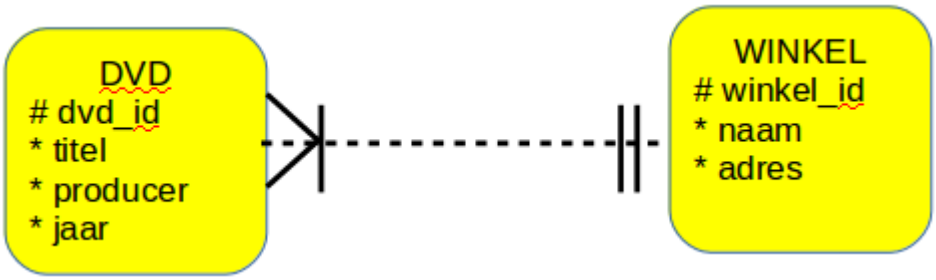
Je ziet dat bij de bouw van de database de relatie wordt omgezet in een extra kolom in de tabel PRODUCT. Kleur kun je hier kiezen uit een rijtje dat je in de tabel van PRODUCT_KLEUR hebt ingevoerd.

Tweede en derde normaalvorm

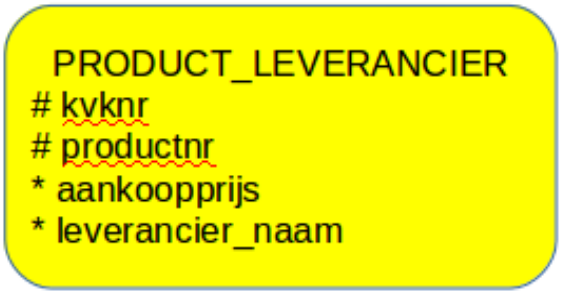
De tweede en derde normaalvorm houden samen in dat iedere attribuut afhankelijk moet zijn van de gehele UID en alleen van de UID. Er mogen dus geen onderlinge afhankelijkheden zijn van andere attributen.

Voorbeeld 1

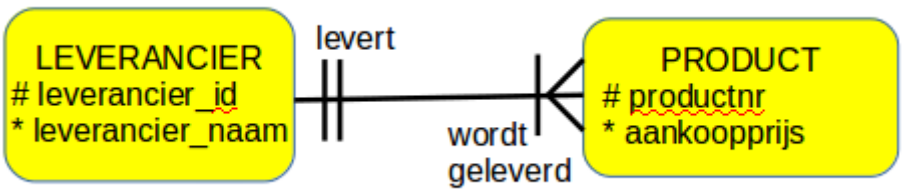
In deze entiteit zijn winkelnaam en winkeladres onderling afhankelijk terwijl zij niet tot de UID behoren. Dat mag niet. Verder zou dezelfde winkelnaam op heel veel records voorkomen. Als de winkelnaam zou veranderen zou je dat op heel veel plekken aan moeten passen. De oplossing is een aparte entiteit WINKEL.



Voorbeeld 2:



De entiteit PRODUCT LEVERANCIER heeft de combinatie van kvknr (kamer van koophandel nummer) en productnr als UID. Het probleem is dat leverancier_naam alleen van kvknr afhangt dus niet van de hele UID. Stel dat een leverancier vijf verschillende producten verkoopt en dat op een gegeven moment de leverancier van naam verandert. Dan moet op vijf verschillende plaatsen de naam van de leverancier aangepast worden. Dat is overbodig werk en foutgevoelig. Zorg er daarom voor dat dezelfde naam nooit meer dan één keer genoteerd hoeft te worden in een database. Oplossing: Maak twee entiteiten: LEVERANCIER en PRODUCT.

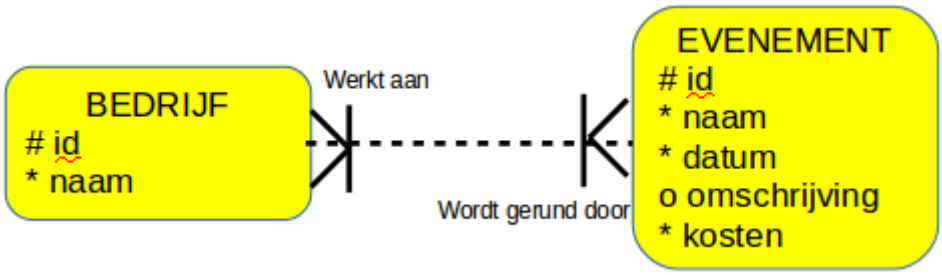


Merk op dat hier een doorgetrokken lijn staat. De UID van PRODUCT is de een combinatie van leverancier_id en product_nr. De eigenschap aankoopprijs is van beide afhankelijk.

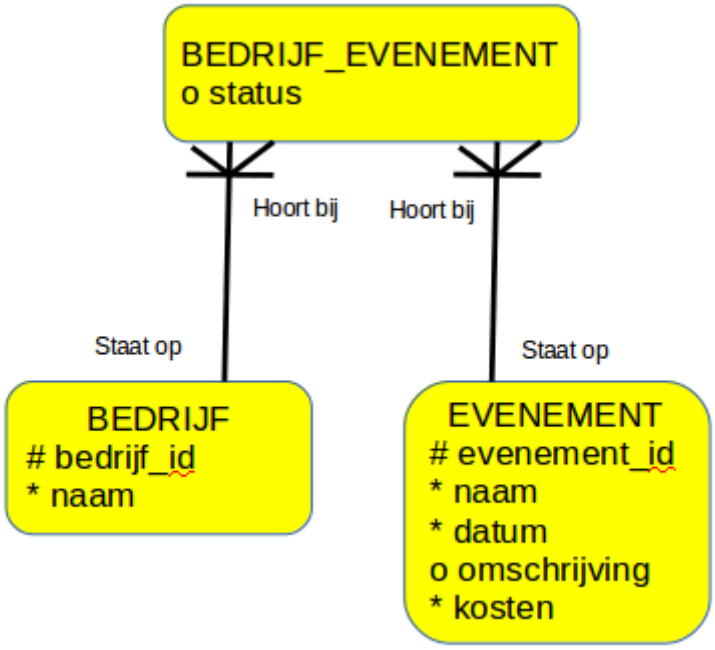
Meer meer relaties

Wanneer je een ERD maakt kom je vaak meer meer relaties tegen. Relationele databases kunnen hier niet mee omgaan. De oplossing is het maken van een intersection entiteit; een tussenentiteit.

Voorbeeld



Oplossing



Je ziet dat de meer meer relatie nu is verdwenen. De lijnen zijn doorgetrokken. Dat betekent dat de UID van BEDRIJF_EVENEMENT een combinatie is van bedrijf_id en evenement_id. De intersection entiteit moet je zien als een soort kaartje. Op ieder kaartje staat een combinatie van een evenement en een bedrijf. Soms, zoals hier, kun je ook nog andere attributen aan de intersection entiteit toevoegen.

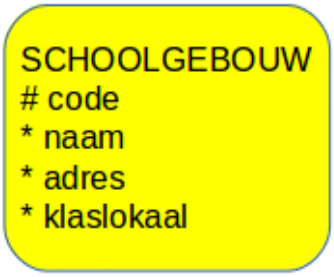
Opgaven

Opgave 1 Garage

Maak een matrix bij de volgende entiteiten: garagebedrijf, auto, persoon.

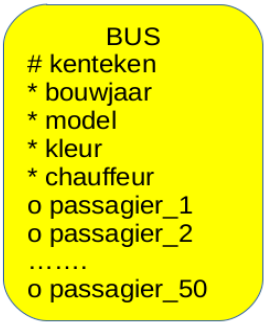
Opgave 2 Schoolgebouw

Normaliseer het volgende ERD:



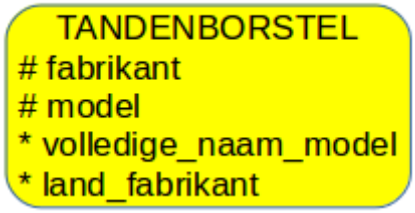
Opgave 3 Bus

Normaliseer de volgende ERD die gaat over een busmaatschappij waarbij de passagiers vooraf via internet een bepaalde lange afstands bus moeten boeken.



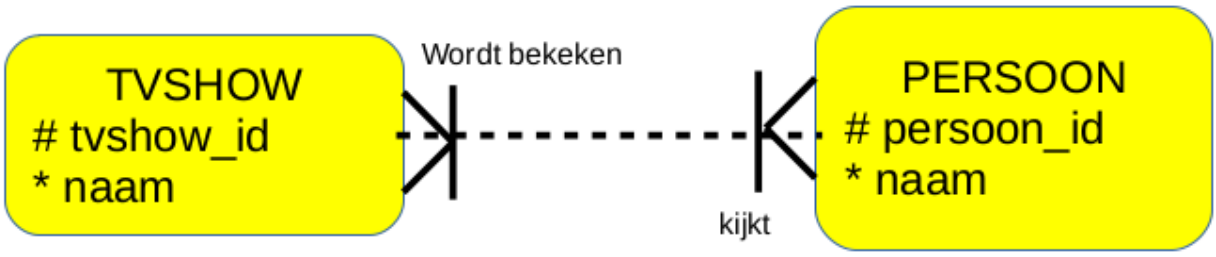
Opgave 4

Normaliseer de volgende ERD:



Opgave 5
Normaliseer de volgende ERD:

Opgave 6
Los de volgende meer meer relatie op.



Hoofdstuk 5 Geschiedenis en supertype subtype

Termen

Geschiedenis

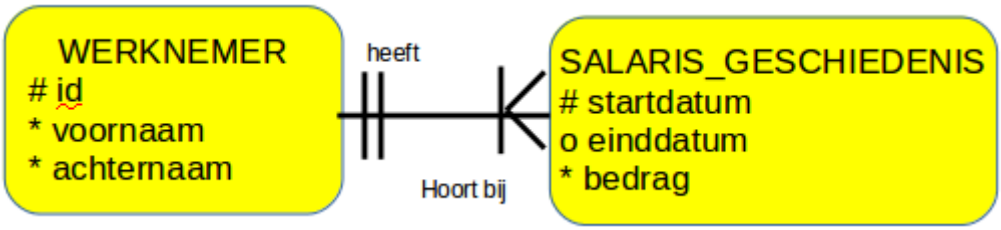
Supertype

Subtype

Geschiedenis

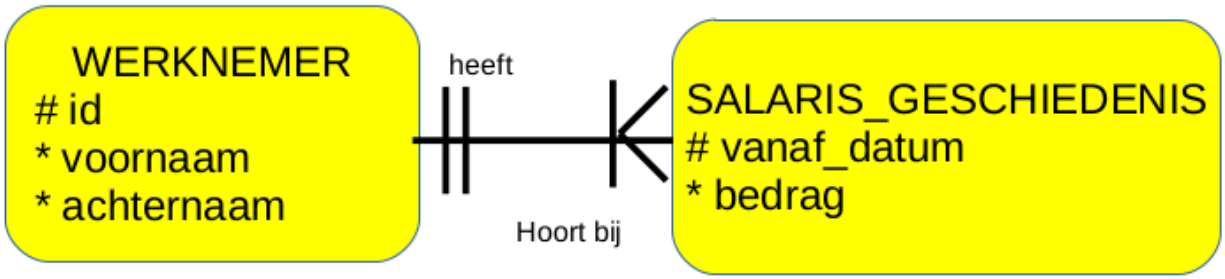
Er zijn veel situaties waarin je iets met tijd moet doen. Hieronder staan twee voorbeelden.

Voorbeeld 1 Salarisgeschiedenis



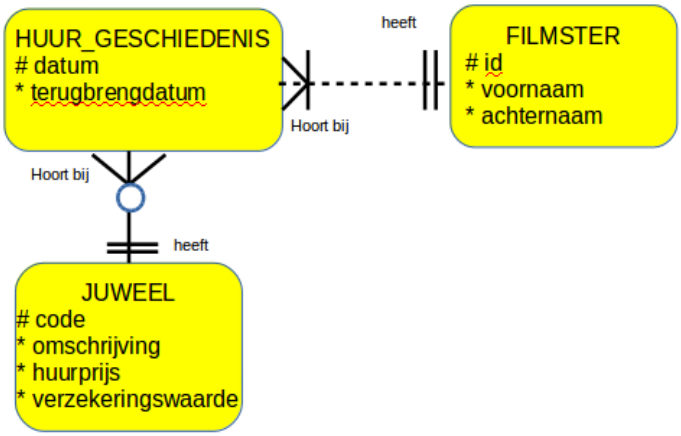
Met behulp van een extra entiteit SALARIS_GESCHIEDENIS kun je het salaris terugzien wat een een werknemer in het verleden heeft verdiend en vanaf welke datum tot welke datum. De einddatum is optioneel omdat bij het verhogen van het salaris nog niet vaststaat tot welke datum de werknemer dit salaris zal krijgen. Merk op dat werknemer_id tot de UID van SALARIS_GESCHIEDENIS behoort want het is een doorgetrokken lijn.

In de documentatie moet verder een constraint worden opgenomen dat de einddatum niet voor de begindatum kan liggen. Er is een ander ontwerp mogelijk waarbij dit niet nodig is:



Hier werk je niet met een einddatum. Verlaat de werknemer de firma dan zet je het salaris op 0,00.

Voorbeeld 2 Juwelenhuur

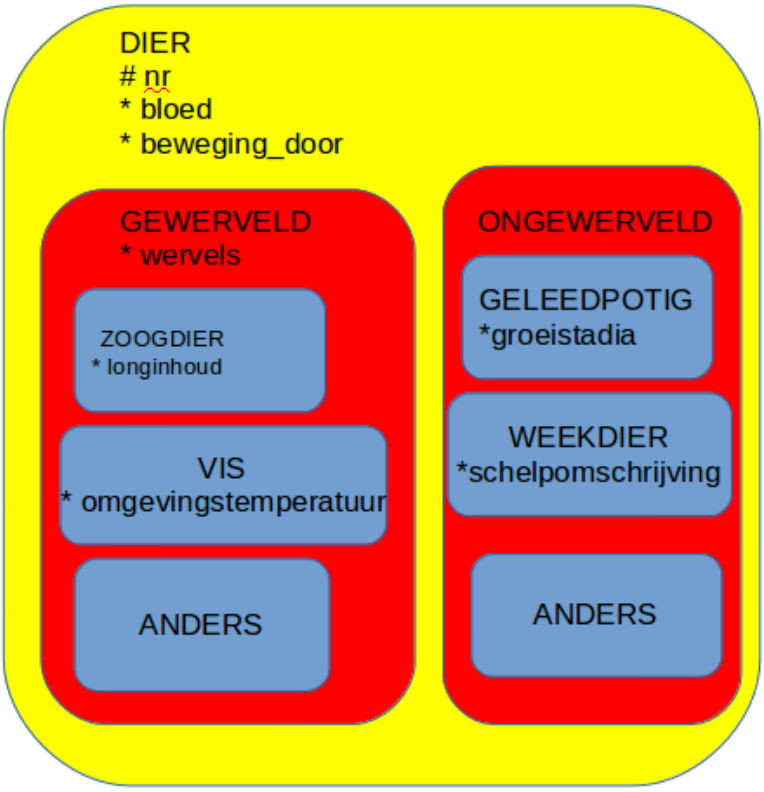


Filmsterren treden vaak op en moeten bij elk optreden andere juwelen dragen. Die juwelen worden niet allemaal gekocht maar vaak gehuurd. Je wilt bijhouden welke filmster welke juwelen heeft gehuurd. De UID van HUUR_GESCHIEDENIS bestaat hier uit datum en juweel_code. Filmster_id behoort niet tot de UID van HUUR_GESCHIEDENIS (stippellijn) omdat, als je dit zou doen, je zou toestaan dat een juweel op dezelfde datum aan twee filmsterren uitgeleend kan worden. Dat gebeurt echter nooit.

Als HUUR_GESCHIEDENIS nog een één-meer relatie zou hebben met een derde entiteit is het handiger om HUUR_GESCHIEDENIS een eigen UID te geven van slechts één attribuut.

Supertype en subtype

Wanneer entiteiten veel dezelfde eigenschappen hebben en een paar verschillende maken we gebruik van de supertype subtype structuur. Hieronder zie je een voorbeeld.



In dit voorbeeld is DIER het supertype en zijn er twee subtypes GEWERVELD en ONGEWERVELD. Deze subtypes bestaan zelf ook weer uit subtypes. GEWERVELD is dus het supertype van VIS. De subtype ANDERS is aangemaakt omdat er nog steeds nieuwe diersoorten worden ontdekt.

Bij supertype subtype gelden een aantal kenmerken:

De subtypes erven alle attributen van het supertype.

In een subtype kunnen opnieuw andere subtypes worden aangemaakt.

Alle exemplaren van het supertype zijn ook exemplaren van één van de subtyes.

Een supertype moet minstens twee subtypes hebben.

Een subtype kan een relatie hebben die een supertype niet heeft.

Opgaven

Opgave 1

Stel dat er een meerdaagse markt op school gehouden wordt en dat je bij wilt houden wie wanneer welke kraam gaat bemannen. Een kraam wordt maar door één vrijwilliger tegelijk bemand. Sommige vrijwilligers kunnen langer werken dan anderen. Het schema moet van tevoren worden gemaakt, zodat bepaald kan worden wanneer de kraam nog niet bemand is. Maak een ERD bestaande uit drie entiteiten voor deze situatie.

Opgave 2

Noem minimaal drie constraints die apart geprogrammeerd moeten worden bij het informatiesysteem uit de vorige opgave.

Opgave 3 Kledingzaak

Onze zaak verkoopt verschillende soorten vrouwenkleren: jurken, shirts en blouses.
le der product heeft een naam, omschrijving en een prijs. Alle producten hebben ook een taillemaat. Jurken en shirts hebben een lengte maar blouses niet. Jurken en en blouses hebben een bustemaat maar shirts niet.

a) Welke entiteiten zitten in dit verhaal?
b) Welke entiteit is het supertype?
c) Welke attributen behoren tot het supertype?
d) Welke UID heeft het supertype?
e) Noteer bij ieder subtype de attributen.
f) Teken de ERD.