

# Data Analysis & SQL

# Inleiding & basisbegrippen



**DE HOGESCHOOL  
MET HET NETWERK**

Elfde-Liniestraat 24, 3500 Hasselt, [www.pxl.be](http://www.pxl.be)





Algemene begrippen  
Soorten gegevens  
Sleutels





Algemene begrippen

Soorten gegevens

Sleutels

# Algemene begrippen

- Wat?

Een database, gegevensbank of databank is een (digitaal) opgeslagen **archief**, ingericht met het oog op **flexibele raadpleging** en gebruik.
- Verschillende begrippen die worden gebruikt voor databank
  - de opgeslagen gegevens (**data**) als zodanig (naam, tel.,...)
  - de wijze waarop de gegevens zijn opgeslagen (**datamodel**)
  - de software waarmee databases kunnen worden aangemaakt en benaderd (database management systeem of **DBMS**)

# Algemene begrippen

- Minimale voorwaarden (om te spreken van een DB)
  - Gegevens moeten eenvoudig kunnen worden opgeslagen.
  - Gegevens moeten eenvoudig kunnen worden opgezocht en doorzocht.
  - Gegevens moeten gewijzigd kunnen worden.
  - Gegevens moeten verwijderd kunnen worden zonder dat de werking van dat systeem nadelig beïnvloedt.
- Index (onontbeerlijk voor DB)

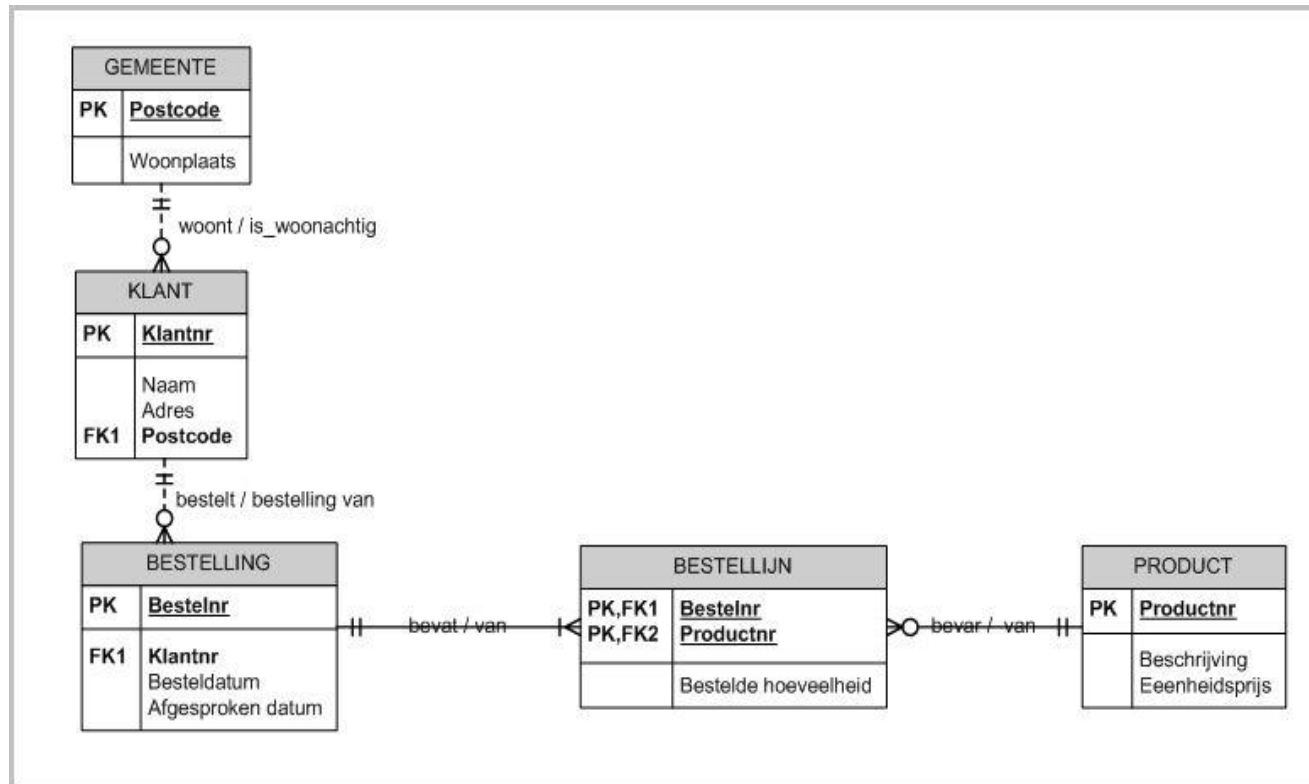
Gegevens in een database zijn zodanig opgeslagen dat deze gegevens optimaal doorzoekbaar zijn. Deze toevoeging staat bekend onder de naam index.

# Algemene begrippen

- Structuur op 3 niveaus beschreven:
- Het **intern** niveau: hoe gegevens worden vastgelegd. Bv. op welke schijf moeten de gegevens komen, ruimte op schijf, indexeringstechnieken, ...
- Het **conceptuele** niveau: (zie ERD volgende dia)
  - Welke gegevens in de database?
  - Welke gegevens vormen samen tabellen (kolommen/rijen)?
  - Welke regels gelden (waarde bereik, relaties tussen tabellen,...).
- Het **extern** niveau:
  - Hoe krijgt eindgebruiker de gegevens? (vb. welke applicatie)
  - Welke gegevens mag eindgebruiker raadplegen? (beveiliging van gegevens)

# Algemene begrippen

- De grafische vastlegging van het conceptuele datamodel gebeurt veelal in een Entity Relationship Diagram (ERD).







Algemene begrippen

Soorten gegevens

Sleutels



# Soorten gegevens

## Samengestelde gegevens vs elementaire gegevens

- Neiging om gegevens te groeperen (Vb. Dhr. Joachim Vandeput)
- Onduidelijk (geen consistentie of afspraken)
- In DB enkel gebruik maken van **elementaire** gegevens: aanspreektitel, voornaam, naam
- Geboortedatum (dd-mm-yyyy) beschouwen we als een enkelvoudig (of atomair) attribuut, maar kan altijd uitgesplitst worden.

# Soorten gegevens

## Procesgegevens

- WAT?  
Attributen die uit andere attributen door **berekening of afleiding** bekomen worden.
- REGELS  
Regel1: Wordt **niet bewaard**, want er vindt een herbewerking plaats bij wijzigen van basisgegevens.  
Regel 2: Is **functioneel afhankelijk** van andere gegevens en wordt dus niet in de gegevensstructuur opgenomen.
- VOORBEELD  
Leeftijd: wordt berekend o.b.v. de geboortedatum



Algemene begrippen  
Soorten gegevens  
Sleutels



# Sleutels

- **Primaire sleutel** (uniek identificeren van gegevens)
  - Uniek, minimaal, numeriek
  - Vb. Studentnummer
  - Synoniem: Primary Key
- **Surrogaatsleutels** (kunstmatige kolom die fungeert als PK = **kortere PK**)
  - Er wordt een surrogaatsleutel gebruikt als de primaire sleutel te groot of onhandig is.
- Vb. VERHUUR\_ITEM (Straat, Postcode, Gemeente, VerhuurTarief)  
VERHUUR\_ITEM (VerhuurID, Straat, Postcode, Gemeente, VerhuurTarief)

Door de surrogaatsleutel *VerhuurID* toe te voegen, wordt het aanmaken en opzoeken in een tabel efficiënter.



001	Briers	...
002	Vandeput	...
003	De Cooman	...
004	Reyskens	...

# Sleutels

## Stelling:

- tabel **Personen**  
→ voornaam, achternaam, geboortedatum, geboorteadres, geboorteplaats, nationaliteit  
*Mogelijke primaire sleutel is onderlijnd*
- tabel **Huwelijken**  
→ voornaam1, achternaam1, geboortedatum1, geboorteadres1, voornaam2, achternaam2, geboortedatum2, geboorteadres2, trouwdatum

## Oplossing:

- tabel **Personen**  
→ Persoonld, voornaam, achternaam, geboortedatum, geboorteadres, geboorteplaats, nationaliteit
- tabel **Huwelijken**  
→ Persoonld1, Persoonld2, trouwdatum

# Sleutels

**Voor het bepalen van de primaire sleutel onderzoeken we:**

- Een supersleutel is de combinatie van aanwezige attributen, ten minste één en ten hoogste gelijk aan het totale aantal attributen van de tabel.
- Een kandidaatsleutel is
  - een minimale reeks attributen die nodig is om een rij te identificeren
  - die overigens uniek is
  - en bovendien niet leeg mag zijn
- Dit wordt ook wel een minimale superkey genoemd en vormt uiteindelijk de primaire sleutel.



# Sleutels

## Voorbeeld supersleutel + kandidaatsleutel:

- Supersleutels:
  - Gsm-nummer
  - Gsm-nummer + naam
  - Gsm-nummer + voornaam
  - Gsm-nummer + naam + voornaam
  - Gsm-nummer + geboortedatum
  - Gsm-nummer + geboortedatum + naam
  - Gsm-nummer + geboortedatum + voornaam
- Geen supersleutels (wegens niet uniek):
  - voornaam
  - naam
  - voornaam + naam
  - voornaam + geboortedatum
  - naam + geboortedatum
- Kandidaatsleutel: Gsm-nummer  
**lechte sleutelkeuze, want iedereen MOET een GSM-nummer hebben**

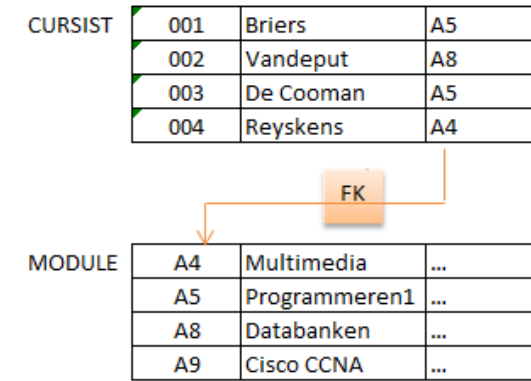
# Sleutels

- **Secundaire sleutel**

- Niet uniek, gebruikt bij opzoeken en sorteren
- Vb. geboortedatum
- Datadictionary houdt plaats van sleutels bij en kan dus sneller werken. Teveel sleutels werkt traag.

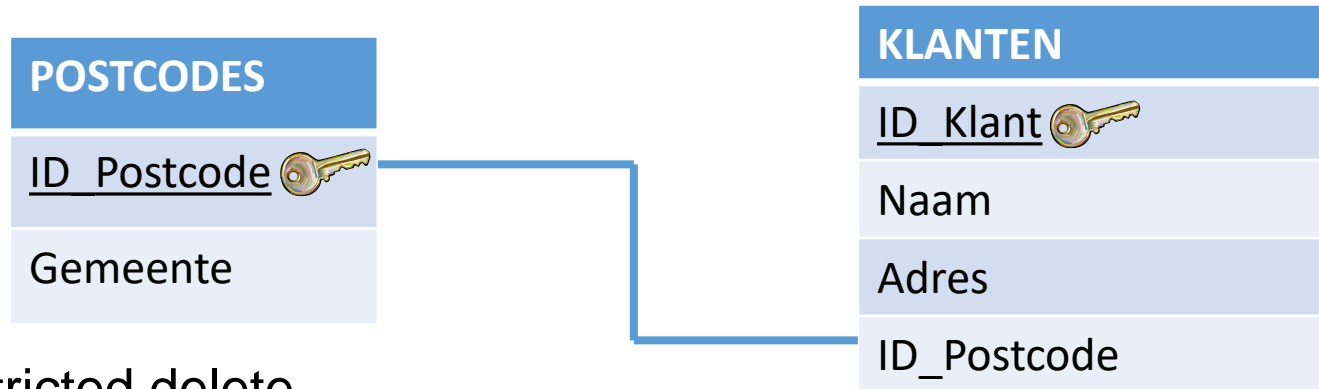
- **Refererende sleutel** (vreemde sleutel of foreign key)

- Associaties leggen  
VB. attribuut Vak in tabel CURSIST stamt uit tabel MODULE
- Eisen: referentiële integriteit: iedere waarde voor dit attribuut moet voorkomen als waarde (PK) bij andere tabel
- Bij schending: restrict, cascade



# Referentiële integriteit

- Consistentie tussen verschillende tabellen waarborgen in een database



- Restricted delete  
postcode 3500 verwijderen niet mogelijk indien er klanten uit Hasselt zijn
- Cascaded delete  
postcode 3500 wordt verwijderd, maar ook de klant(en) uit Hasselt,  
maar ook de facturen van die klanten uit Hasselt en de betalingen van die facturen en ...



# Redundantie

- In een goed ontworpen relationele databank bewaren we gegevens op **één plaats** in een tabel. Geven we informatie toch meermaals weer dan spreken we over **redundantie**.
- Indien we ook in de tabel Cursist de naam van de module opnemen zou dat tot volgende problemen kunnen leiden.

Vb. Naam van A5 wordt Web Essentials

→ Op 2 plaatsen wijzigen?

→ Op 1 plaats vergeten? Wat dan?...

