


Data Vault in het Business Intelligence onderwijs

Dineke Romeijn
HAN, ICA



Voorstellen

- Dineke Romeijn
 - Docent
 - Minor Business Intelligence (teamtrekker)
 - Master
 - Onderzoeker lectoraat M-BIS
 - Studieloopbaanbegeleider DT I, BIM, TI
- Kennischeck



Inhoud

- Data Vault
- BI in het curriculum van HAN
- Data Vault in Minor BI -Ervaringen
- Bedrijfscontacten



Inhoud

- **Data Vault**
- BI in het curriculum van HAN
- Data Vault in Minor BI-Ervaringen
- Bedrijfscontacten



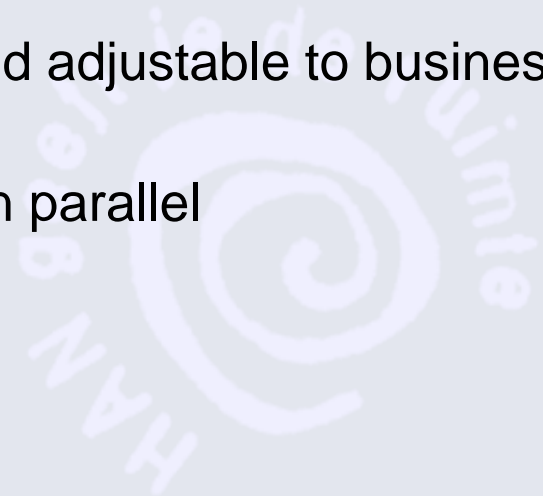
Wat is Data Vault?

- Een gegevensmodel voor een DWH
- Een architectuur



Data Vault: het model

- Ontworpen door Dan Linstedt vanaf 2002
- Volgens Linstedt:
 - A collection of detail-oriented, history-tracing, and uniquely linked collection of normalized tables
 - A hybrid approach combining the best of 3NF and dimensional modelling
 - Flexible, scalable, consistent and adjustable to business needs
 - Handle high-speed data loads in parallel



Data Vault elementen

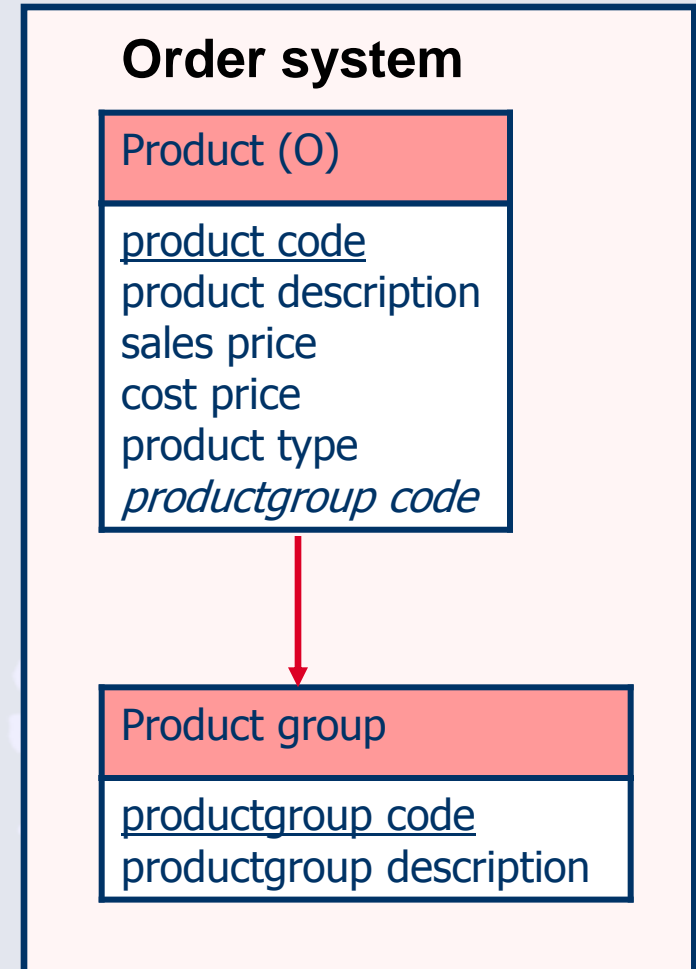
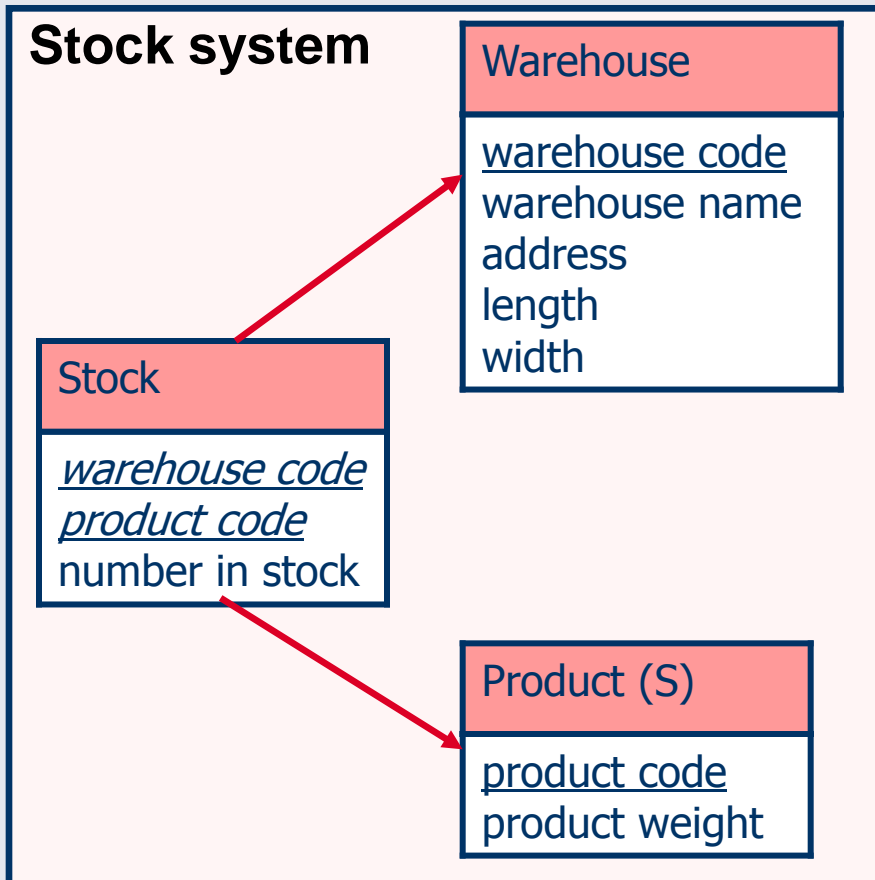
Drie componenten:

- Hubs: bevatten de bedrijfs sleutels van entiteiten
 - Links: relaties tussen entiteiten, afhankelijke entiteiten
 - Satellieten: alle andere relevante attributen van een entiteit of relatie

 - Surrogaatsleutels voor hubs, eventueel ook voor links
 - Voor alle elementen een verwijzing naar de bron en een timestamp voor de laaddatum
 - Voor satellieten is de timestamp deel van de sleutel (vergelijkbaar met SCD van Kimball)
-

Data Vault: een voorbeeld

- Twee bronsystemen



Source: Stock system

Sat_Warehouse

warehouse id
load datetime stamp
end datetime stamp
record source
warehouse name
address
length
width

Sat_Stock

warehouse id
product id
load datetime stamp
end datetime stamp
record source
number in stock

Sat_Product_Stock

product id
load datetime stamp
end datetime stamp
record source
product weight

Hub_Warehouse

warehouse id
warehouse code
load datetime stamp
record source

Link_Stock

warehouse id
product id
load datetime stamp
record source

Hub_Product

product id
product code
load datetime stamp
record source

Sat_Product_Order

product id
load datetime stamp
end datetime stamp
record source
product description
sales price
cost price
product type

Link_ProductGroup

product id
productgroup id
load datetime stamp
record source

Hub_ProductGroup

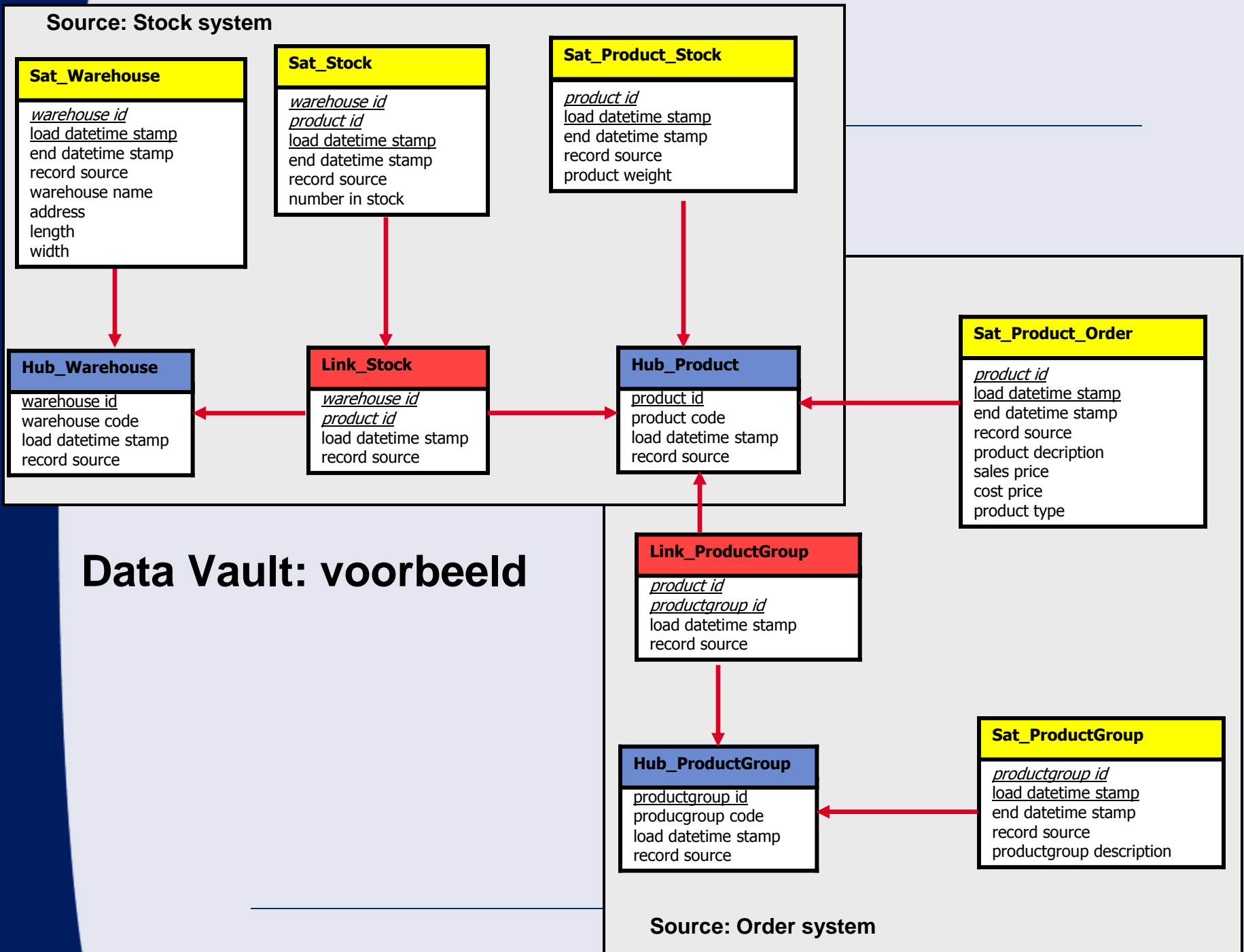
productgroup id
productgroup code
load datetime stamp
record source

Sat_ProductGroup

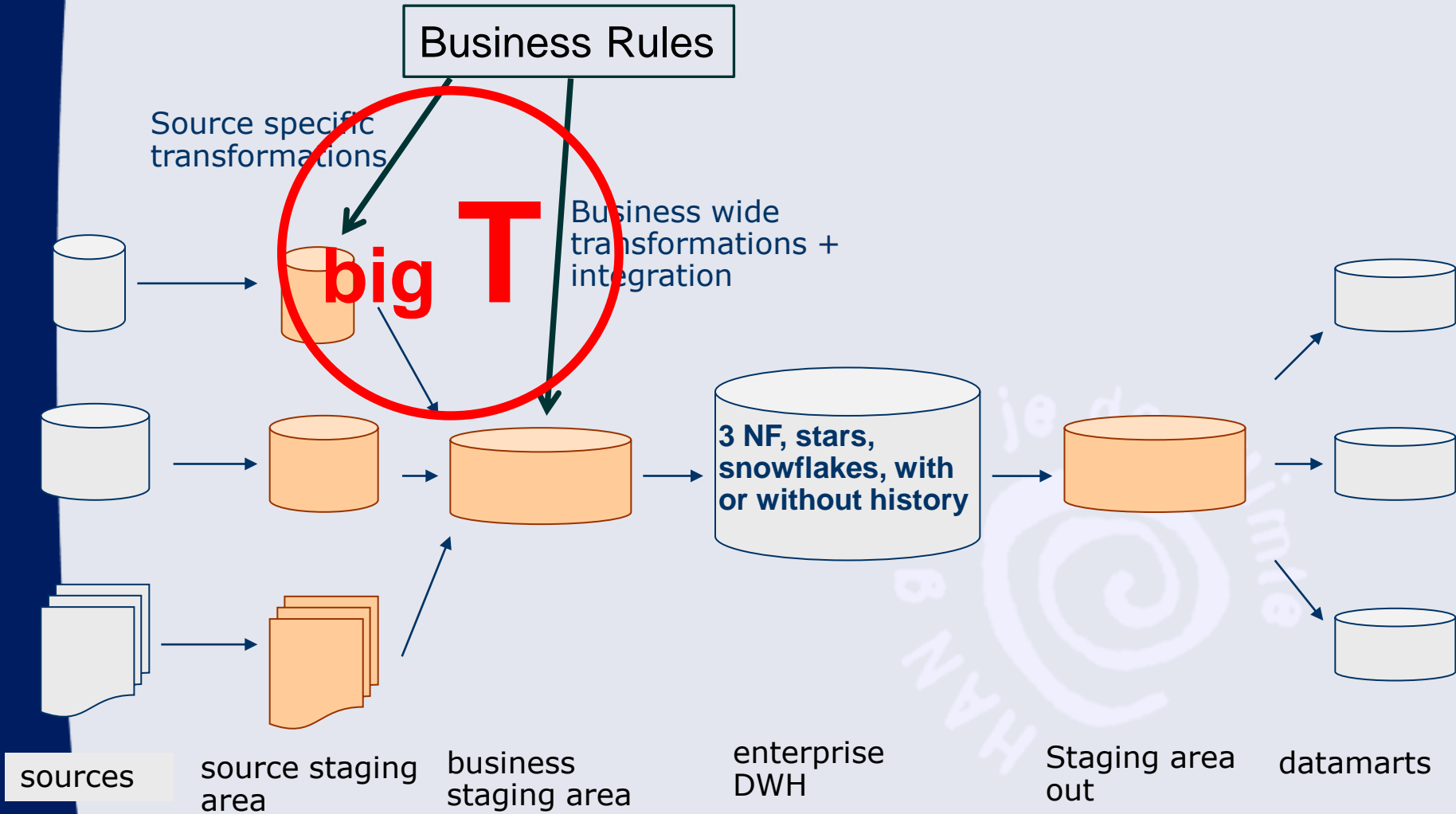
productgroup id
load datetime stamp
end datetime stamp
record source
productgroup description

Source: Order system

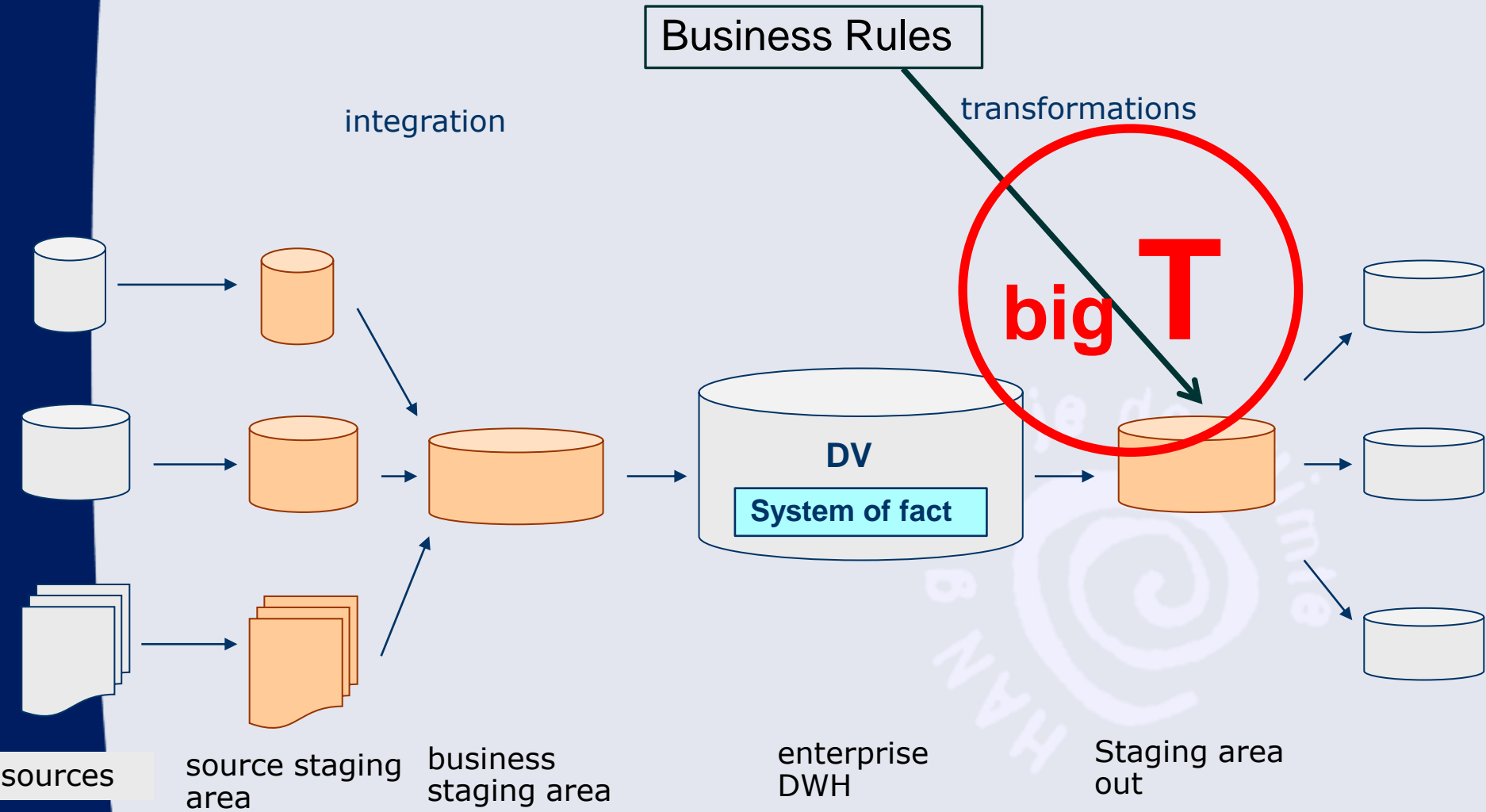
Data Vault: voorbeeld



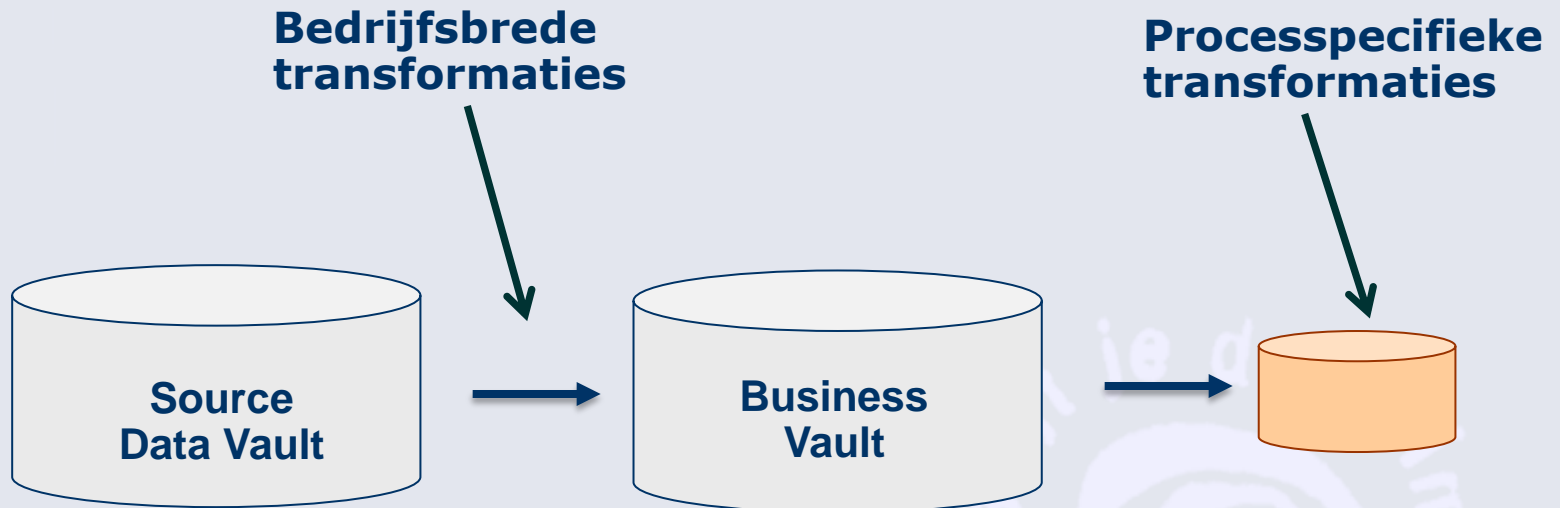
Klassieke architectuur



Data Vault Architectuur



Data Vault Architectuur (uitgebreid)



Voordelen beroepspraktijk

- Opslag complete historie mogelijk
 - Compliance + auditibility

 - 3 simpele elementen
 - Schaalbaar, flexibel

 - System of fact: feiten zoals bekend in de bronnen
 - Volledig bronafhankelijk niet regelafhankelijk
 - Scheiding verantwoordelijkheden IT en business
-

Model driven generation

- Model-gedreven generatie mogelijk
- Parallel laden mogelijk, ETL snel en genereerbaar
- Lagere ontwikkel- en onderhoudskosten
- Ware explosie van tooling op dit vlak
 - Topic in BI Dutch 15+ tools in 2011
 - http://www.linkedin.com/groups/Is-datawarehouse-generatie-de-nieuwe-46641.S.49985459?trk=group_search_item_list-0-b-ttl&goback=%2Egna_46641

Nadelen 1

Gebrek aan standaardisatie:

- Geen heldere formele beschrijving
- ‘Gesloten’ bedenker daardoor
 - veel discussiepunten en
 - variaties
- Discussie over best practices



Nadelen 2

Technisch:

- Kan leiden tot extra (evt. virtuele) lagen
- Kan leiden tot extra modellen
- ETL complexer bij handmatige ontwikkeling

Geen nadeel maar ook geen voordeel:

Aan de 'big T' wil niemand de vingers branden, ook DV niet....

Inhoud

- Data Vault
- **BI in het curriculum van HAN**
- Data Vault in Minor BI-Ervaringen
- Bedrijfscontacten



Waarom Data Vault in curriculum?

- Waarom BI en DWH überhaupt?
 - Toenemende vraag naar analyse van gegevens
 - Data explosie
 - Schatting sommige consultants:
 - 80 % van DWH projecten, nieuw of herziening bestaande oplossing, maakt gebruik van Data Vault
 - Volgens anderen wat optimistisch, maar toch..
 - Hoewel er anderssoortige oplossingen bij komen, blijft klassiek datawarehousing in de praktijk nog steeds zeer belangrijk.
-

Historie BI onderwijs HAN

Vanaf 1998 tot 2002:

Keuzemodule Datawarehousing / Business Intelligence
voor de Hogere Informatica opleiding

Vanaf 2002 tot 2008:

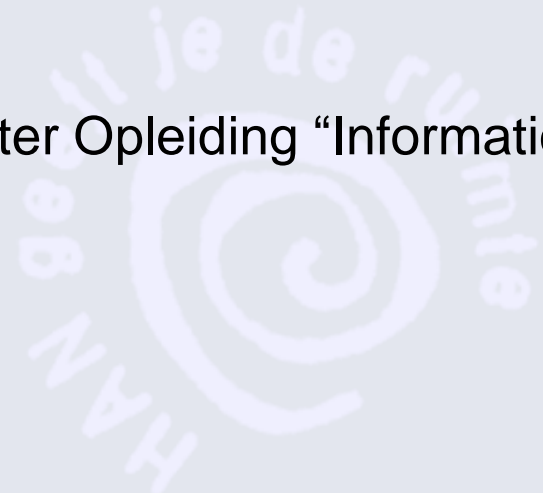
Business Intelligence in het oude curriculum voor BI, I
en TI opleidingen

Vanaf 2003-2012:

Business Intelligence in de Master Opleiding “Information
Systems Development”

Vanaf febr. 2009:

Minor Business Intelligence



BI in huidig curriculum HAN

- Voorbereiding in semesters: DIS, ABI
- Basis in semester DECO
- Zwaartepunt in minor Business Intelligence
- Afstudeeropdrachten



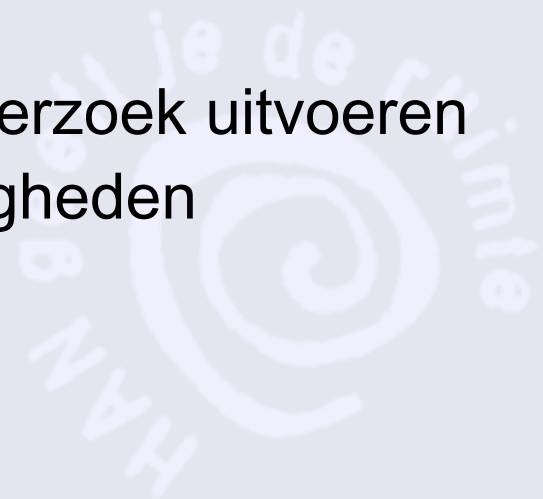
Minor BI

- Onderzoeksminor
 - Verdiepende minor voor BIM, I
 - Voorkennis studenten wisselend, jaar 2, 3 of 4
 - Verbredend voor studenten FEM en Technische bedrijfskunde (FT), alleen na voorbereidingstraject
-



Minor BI - onderdelen

- Course 1 (Basiskennis): 7,5 ec
 - Voortraject BI (VB)
 - *Prototypetraject BI (PB)*
- Course 2 (BI thema's): 7,5 ec
- Project: 15 ec
 - Een in overleg gekozen onderzoek uitvoeren
 - Start met onderzoeksvaardigheden



Course 1: Voortraject

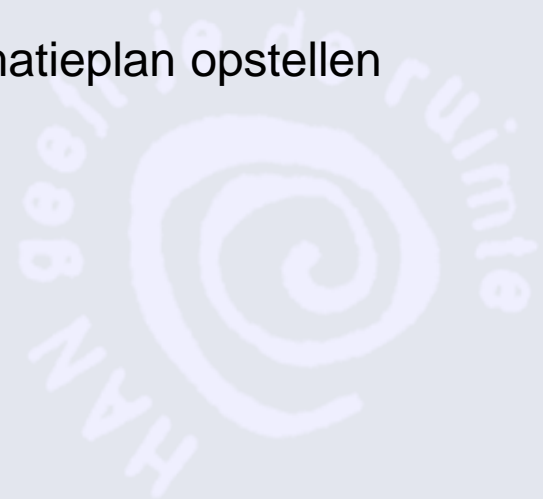
Onderwerpen:

- De Business Intelligence (BI) processen
 - Doelstellingen voor bedrijfsprocessen
 - Strategische doelstellingen
 - (Meet-)gegevens, een meetplan en een actieplan opstellen voor een doelstelling
 - Presenteren en verspreiden van informatie
 - BI-projecten, beheer en exploitatie, outsourcing
 - Business Intelligence implementeren
 - BI promoten
-

Course 1: Voortraject

Opdrachten:

- Een BI advies voor een bedrijfsprocessen
 - Doelstellingen, (meet-)gegevens, meetplan, actieplan, business case en implementatieplan opstellen
- Een Strategisch BI advies
 - Doelstellingen en strategisch informatieplan opstellen



Course 1: Prototypetraject

Centrale beroepstaken:

- Hoe ontwerp je een systeem dat snel, accuraat en flexibel de gewenste informatie levert?
- Hoe realiseer je vervolgens een prototype van dit systeem?



Course 1: Prototypetraject

Onderwerpen:

- transactional processing vs analytical processing
 - datawarehouses, datamarts
 - BI architecturen
 - **Data Vault**
 - sterren en dimensies
 - historie van gegevens
 - het ontwikkelproces en ondersteunende methoden
 - ETL
-

Course 1: Prototypetraject

Casus:

- Integreert voortraject en prototypetraject
- Ontwerp en realiseer een BI omgeving



Course 1: Prototypetraject

Leermiddelen

Literatuur:

- “Sterren en dimensies”, H. van der Lek, F. Habers, M. Schmitz
- “Business Intelligence en datawarehousing”, Tweede editie, K. Verhagen
- Diverse uit te reiken artikelen: Scholar

Software:

- Microsoft Excel 2003/2007/2010
 - SQL Server 2008 R2, eventueel 2005 (incl service pack 3)
 - SQL Server Business Intelligence Studio
 - Power Designer
 - Front-end tool naar keuze
-

Inhoud

- Data Vault
- BI in het curriculum van HAN
- **Data Vault in Minor BI –ervaringen**
- Bedrijfscontacten



Data Vault in minor BI

Historie:

- 2008 eenmalige, op zichzelf staande les

Via:

- Losse workshop halve dag en
- Aanvullende opdracht bij casus

Naar:

- Volledig geïntegreerd in casus en lessen
 - Centrale methode voor modelleren DWH
-



Data Vault in de minor BI

Drie aspecten:

- Data Vault model
- Data Vault architectuur
- Model gedreven generatie

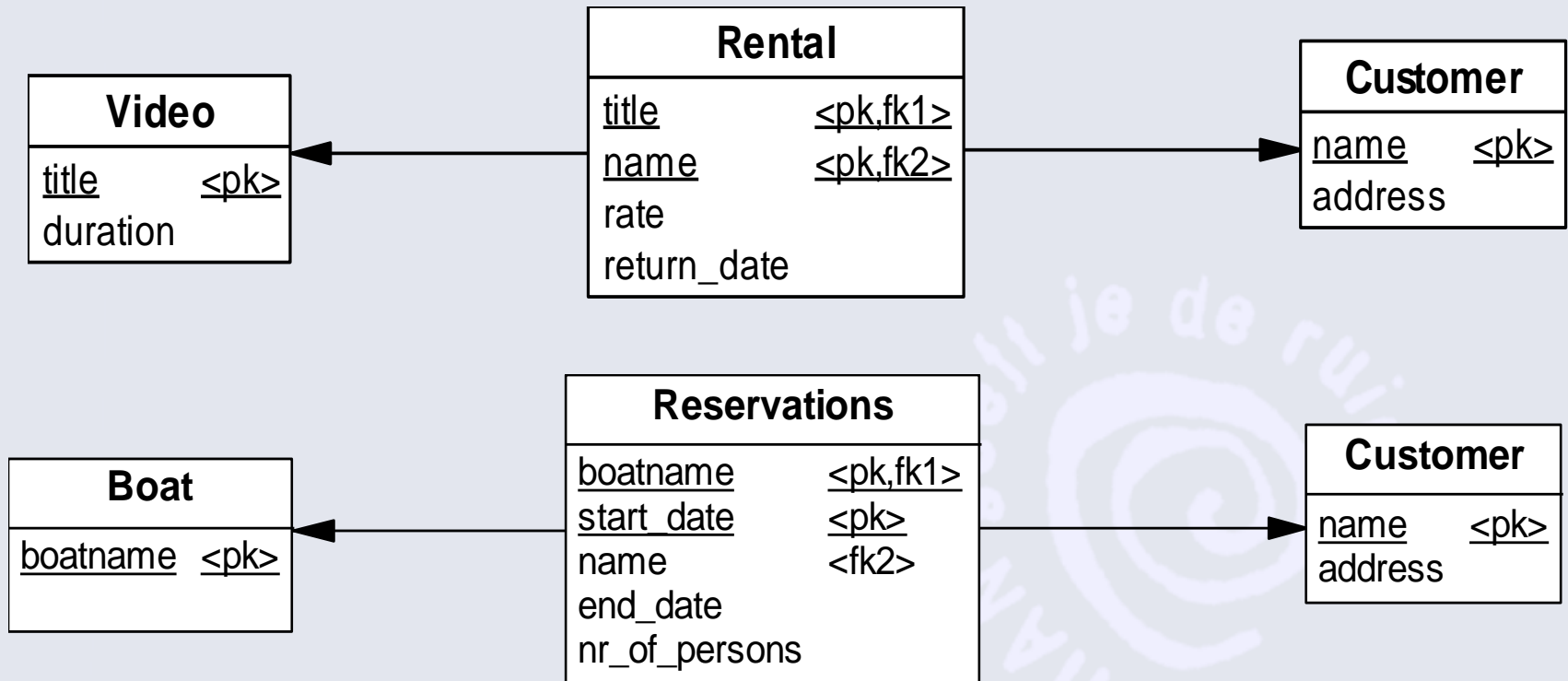


Data Vault model

- Het gegevensmodel voor een DWH
 - 2 dagdelen aan begin van prototypetraject
 - Inleidende presentaties
 - Gericht oefenen met kleine voorbeelden
 - Via vaardigheidstraining goede basis leggen
 - Integraal onderdeel casus
 - Complexere modelleerkwesties zitten in casus verwerkt
 - Met goede basis ontdekken de meeste studenten dit zelf
 - Nadere uitleg a.d.h.v. vragen
 - Bespreken alternatieven
-

Voorbeeld oefeningen

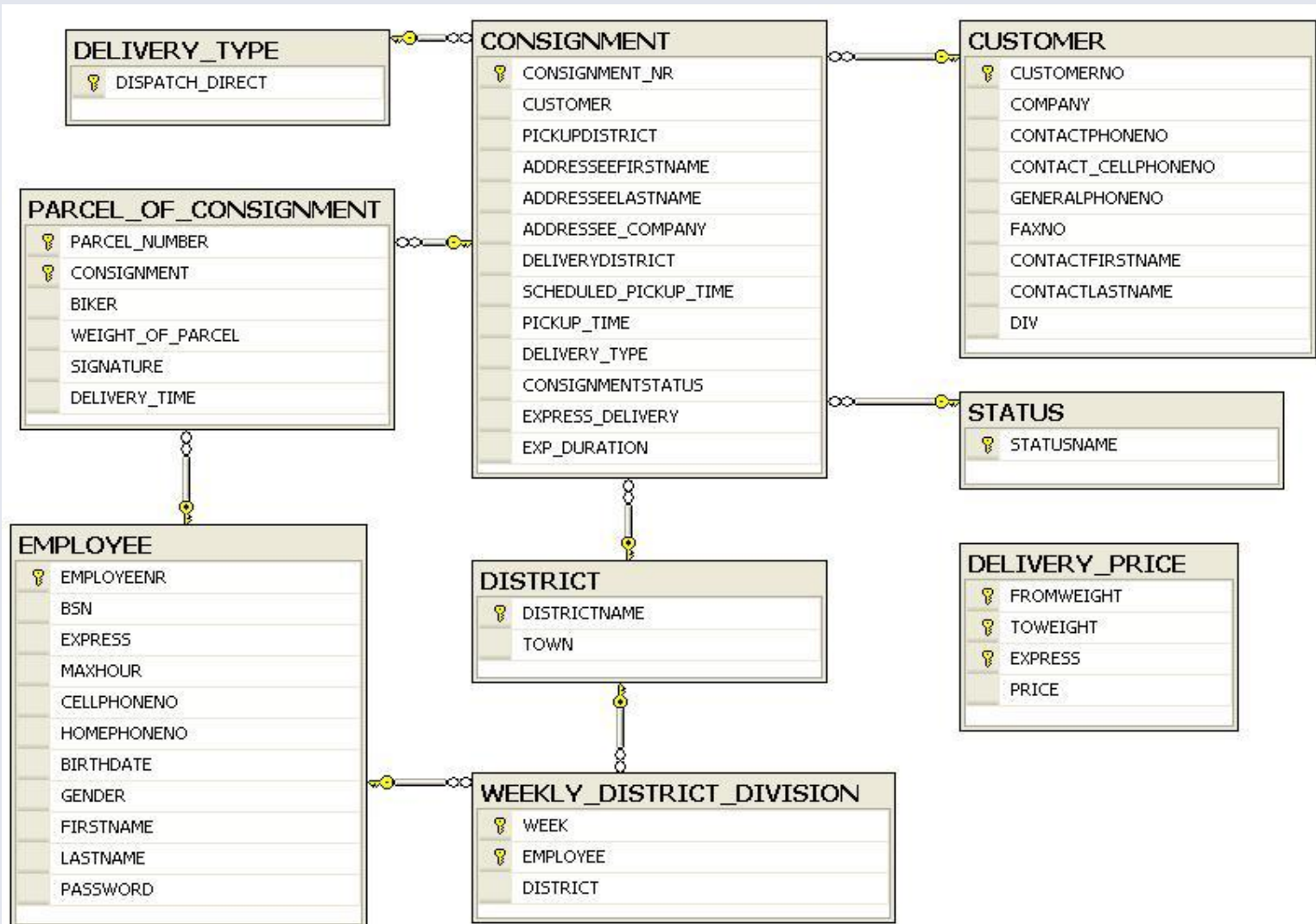
Design a Data Vault for the following 3NF logical relational schemas:



Casus

- Uitgangspunt: KPI's uit het voortraject
 - Ontwerp en realiseer een DWH (Data Vault) uitgaande van de gegeven brondatabase
 - Ontwerp en realiseer een DM (ster) uitgaande van de gegeven brondatabase en de KPI's
 - Ontwerp en realiseer het ETL proces
 - Ontwerp en genereer een dimensionele kubus
 - Analyseer de gegevens in de kubus met
 - Excel
 - andere front-end tool
 - Controleer en vergelijk de resultaten en werkwijzen.
-

De casus: De groei van Arnhem Biker Couriers



Knelpunten

- Voor BIM studenten ETL lastig:
 - Goede voorbeelden
 - Gebruik tooling in SSIS
 - Benadrukken belang goede kennis SQL
 - Voorkennis studenten verschillend, niet allemaal bekend met ERM of FOM
 - Één van de eerste onderdelen van de casus:
 - Blijven hangen
 - Overschatten belang
 - Er is niet één standaard, alternatieven maakt zaak complexer:
 - Leuk voor goede studenten
 - Verwarrend voor anderen
-

Data Vault architectuur

- Vanaf week 1 steeds terugkerend thema
 - Oplopend in complexiteit
 - Land pas echt goed na diepgaand inzicht in BI, meestal aan eind van de minor
 - Deels vraaggestuurd, n.a.v. de stof komen veel studenten met vragen om nadere uitleg
 - Dan ook alternatieven bespreken
 - Inzicht bieden in architectuur blijkt veel lastiger dan model
-

Model gedreven generatie

Gastpresentatie en workshop:

- Genereren sterschema's en Data Vault via tooling
 - praktijkvoorbeeld uit bedrijfsleven
 - QOSQO – Lulzim Bilali
 - Quipu – open souce tool
- Presentatie vanuit lectoraat:
 - Transformatie vanuit een conceptueel model naar DV
- Knelpunt wisselende voorkennis studenten



Inhoud

- Data Vault
- BI in het curriculum van HAN
- Data Vault in Minor BI-Ervaringen
- **Bedrijfscontacten**



Contacten bedrijfsleven

- NLDVGG – Martijn Evers
- Organisatie studiedagen informatiemodellering voor deskundigen bedrijfsleven en medewerkers M-BIS
- Via bedrijfsbezoeken en gastpresentaties
- Organisatie van
 - Cursussen met en voor bedrijfsleven – MATTER
 - Congressen, bv Data Modeling Zone Europe 2013



