

Opgave til undervisning i Data Warehousing

Der skal oprettes en data mart baseret på en **dimensionaliseret** model (et stjerneschema i dette tilfælde). Data marten skal tage udgangspunkt i at gemme data om salg ned til de enkelte linier på kundeordrer. Disse data dimensioneres på kunder, tid, medarbejder og produkt.

Nedenunder er der oplyst, hvilke dimensioner (med niveauer, hierarkier og attributter) samt hvilke fakta (med attributter), der skal oprettes. Datatyper og længde skal I selv bestemme – det anbefales dog kraftigt at se hvor store felter der er afsat til data i SAP Business One. Bemærk at der også er nævnt nogle regler i parenteser som der skal tages hensyn til ved opbygningen af transformationer til overførslen af data.

Det anbefales at starte med at lave et Excel-ark til at danne et overblik over de dimensioner, niveauer, attributter og fakta der skal anvendes i stjerneschemaet samt hvilke kildetabeller og kolonner fra SAP, hver enkelt data kommer fra. Heri bør I også notere hvilke datatyper og længder der skal anvendes, samt hvordan kildetabeller joins. Anvend f.eks. disse kolonner:

- Dimension/Fakta
- Dimensionsniveau
- Dimensions-/Faktaattribut
- Kolonnenavn Typisk sammensat af dimensionsniveau og dimensionsattribut
- Datatype Der skal anvendes i Data Warehouse
- Længde Der skal anvendes i Data Warehouse
- Not Null Ja eller Nej til hvorvidt om feltet er obligatorisk
- Kildesystem
- Kildetabel(ler)
- Kildekolonne(r) Det kan være flere kildekolonner til en dimensionsattribut
- Datatype Anvendt i kildesystemet
- Længde Anvendt i kildesystemet

Bemærk at mht. datalængde i en Microsoft SQL-database er datatypen bigint = op til 19 tal, int = op til 10 tal, smallint = op til 5 tal samt tinyint = op til 3 tal.

Når dette er klarlagt kan I lave E/R-diagrammet med de korrekte datatyper & længder i Designer. E/R-diagrammet skal efterfølgende danne grundlag for databasedesign og den faktiske data warehouse-database.

Datakilden er den danske demodatabase (SBODemo_dk) i SAP Business One 2005:

brugernavn: **sa**

password: **manager**

database: **SBODemo_DK**

ODBC-forbindelse: **sbodemo_dk**

Den logiske model skal designes i Oracle Designer:

brugernavn: **<initialer>designer**

password: **<initialer>designer**

database: **orcl**

Data Warehouse (stjerneschemaet) skal dannes i en ny bruger (for at adskille jeres data):

brugernavn: **<initialer>dwh**

password: **<initialer>dwh**

database: **orcl**

Dimensioner

Tid (denne dimensions kolonnenavne og datatyper skal anvendes præcist som angivet)

Niveauer og attributter/kolonnenavne:

dag

- date_tk NUMBER(6)
- date_field DATE
- day_of_year NUMBER(3)
- day_of_month NUMBER(2)
- day_of_week NUMBER(1)
- day_of_week_desc VARCHAR2(30)
- day_of_week_short_desc VARCHAR2(3)

uge

- week_of_year NUMBER(2)

maaned

- month NUMBER(2)
- month_desc VARCHAR2(30)
- month_short_desc VARCHAR2(3)

kvartal

- quarter NUMBER(1)

aar

- year NUMBER(4)

Hierarki:

år → kvartal → måned → uge → dag

Produkt (kun salgsartikler)

Niveauer og attributter:

produkt

- syntetisk_id
- artikelnr
- navn

gruppe (artikelgruppe)

- kode
- navn

salgspotentiale

(Data til dette niveau kommer fra et Excel-ark)

- kode VARCHAR2(1)
- navn VARCHAR2(30)

Hierarkier:

gruppe → produkt

salgspotentiale → produkt

Medarbejder

Niveauer og attributter:

medarbejder

- syntetisk_id
- nummer
- navn (fulde navn i et felt)

stilling

- titel

koen

- navn (skrevet ud som mand eller kvinde)

Hierarkier:

stilling → medarbejder

køn → medarbejder

Kunde (skal kun indeholde kunder)

Niveauer og attributter:

kunde

- syntetisk_id
- kode
- navn
- adresse

by

- postnr
- navn

land

- kode
- navn

kategori

- kode
- navn

Hierarkier:

land → by → kunde

kategori → kunde

Faktatabel

Salg (baseret på kundeordre)

Fakta:

ordrenummer

status

maengde

salgspris

Den naturlige nøgle i hver dimension skal også findes i de anvendte kildetabeller til fakta. Ift. tidsdimensionen anvendes Ordredato (bogføringsdato) som naturlig nøgle til fakta, mens den ansvarlige medarbejder på ordren er den naturlige nøgle ift. medarbejderdimensionen.