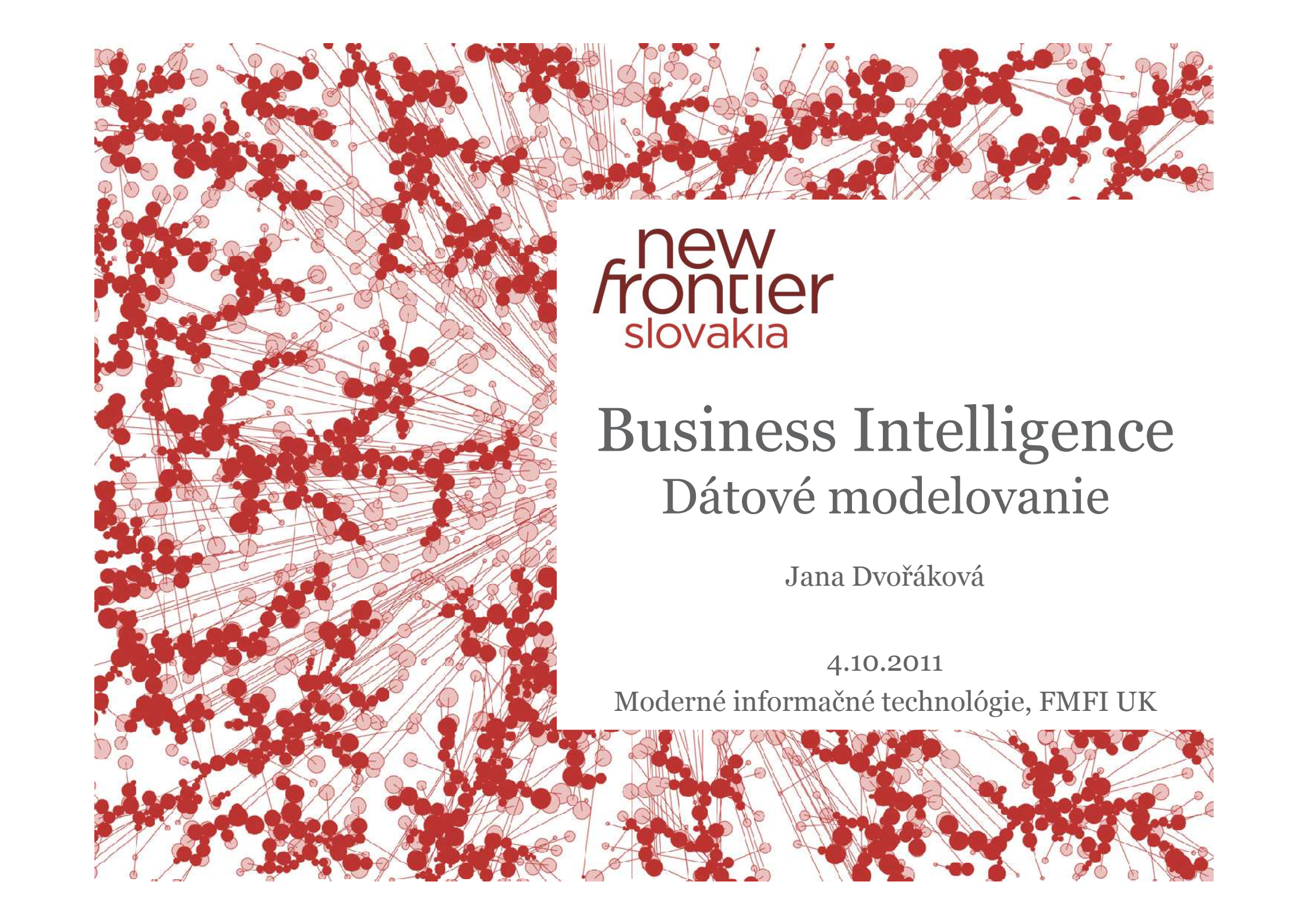


The background of the slide is a complex network of red circles of various sizes connected by thin red lines, creating a dense, web-like pattern. This pattern is visible on the left side and at the top and bottom of the slide, framing the central text area.

new
frontier
slovakia

Business Intelligence

Dátové modelovanie

Jana Dvořáková

4.10.2011

Moderné informačné technológie, FMFI UK

Dátové modely – použitie v DWH

Dimenzionálny model – typy tabuliek

Ukážka modelu dátového skladu

Nástroje

Zhrnutie a diskusia



3-NF vs. dimensionálny model

o Relačný dátový model v 3-NF

- Odstránenie duplicitných dát – zmenšenie počtu záznamov
- Zvýšenie počtu tabuliek
- Prepojenie cez cudzie kľúče a tabuľky relácii
- Efektívny insert/update, menej efektívne dotazovanie

o Dimensionálny dátový model

- Adaptácia relačného modelu
- Faktové a dimensionálne tabuľky
- Denormalizovaný, duplicitné dáta
- Menší počet tabuliek
- Efektívne dotazovanie

Dimensionálny model
typicky nie je v 3-NF
(aj keď definícia to
nevyklučuje)

DWH – relačný model v 3-NF

0 3-NF



Jednoduché ETL

- Prenos dát zo zdrojových systémov a ich integrácia



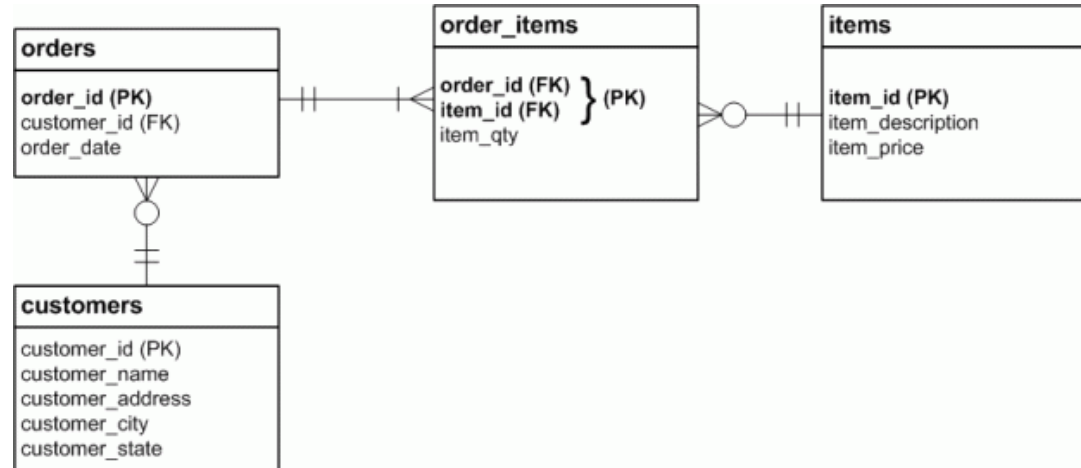
Zložité reportovanie

- Veľké množstvo JOIN operácií
- Ťažšie pochopiteľný bussiness používateľmi

o Vhodné pre niektoré typy databáz

- TERADATA

o Model pre centrálnu úložisťu dát – podľa B. Inmona



DWH – dimenzionálny dátový model

o Štandardne odporúčaný pre DWH



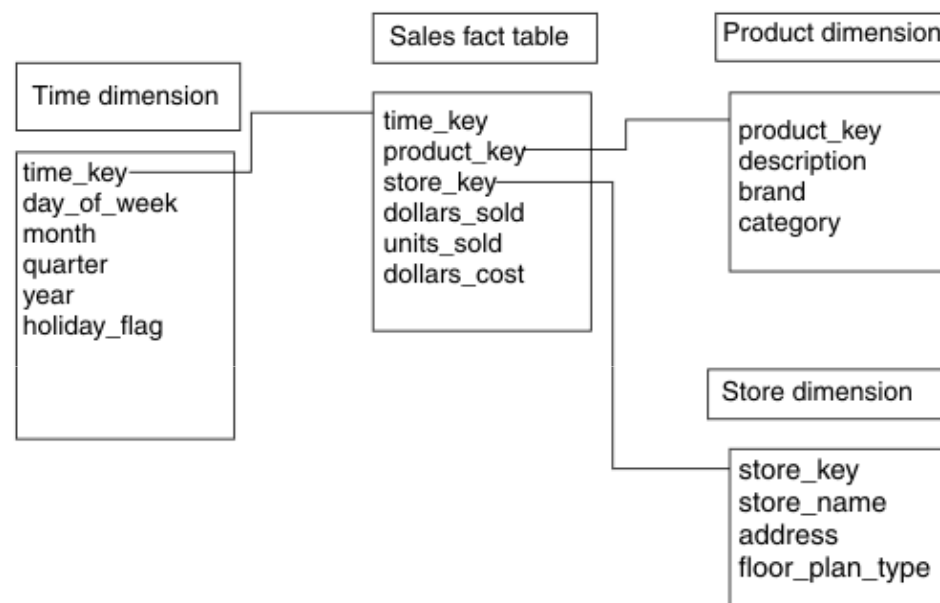
Zložité ETL

- Transformácie dát
- Integrácia, ...



Jednoduché reportovanie

- Ľahšie pochopiteľný
business používateľmi



- o Vhodný (aj) pre relačné databázy
 - Oracle, MSSQL, ...
- o Model pre datamarty – podľa B. Inmona aj R. Kimballa

Agenda

Dátové modely – použitie v DWH

Dimenzionálny model – typy tabuliek

Ukážka modelu dátového skladu

Nástroje

Zhrnutie a diskusia



Typy tabuliek DWH

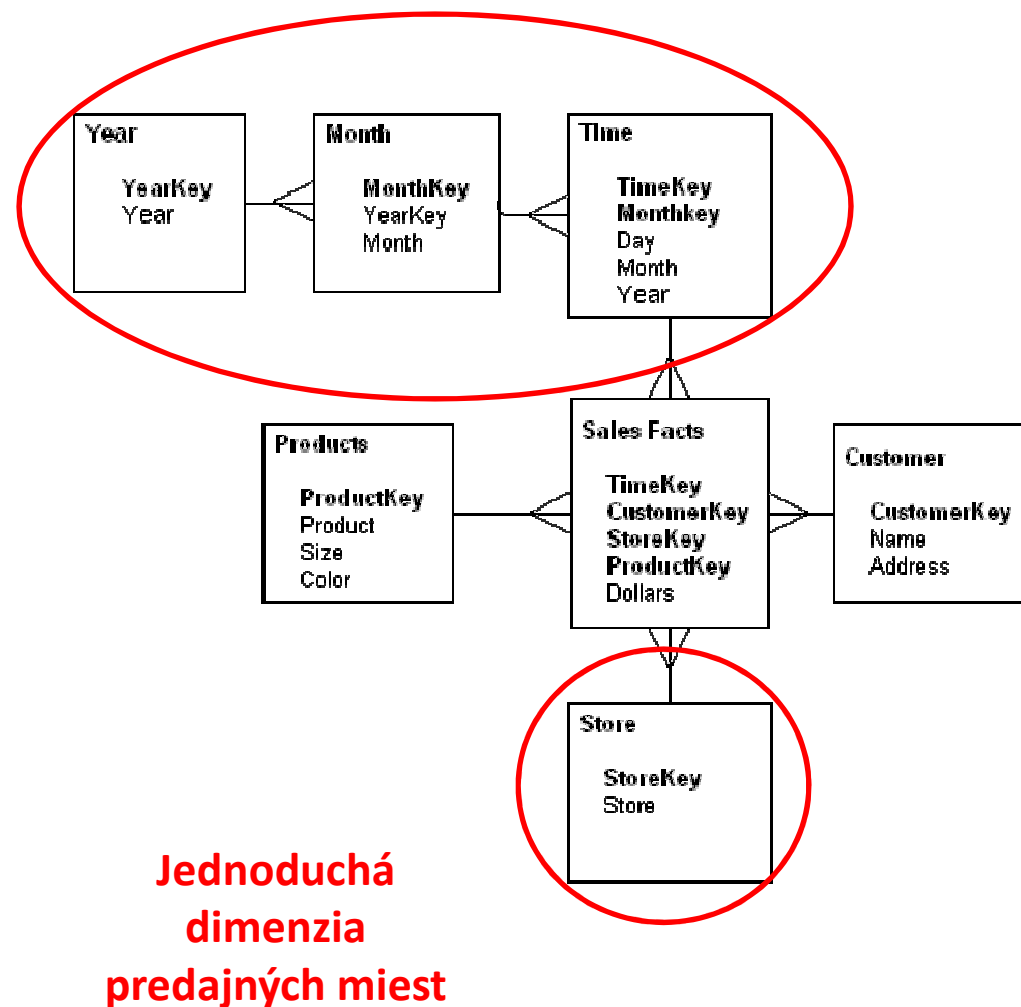
- o Dimenzionálne tabuľky – dimenzie
- o Faktové tabuľky – fakty
- o Agregáčné tabuľky
- o Ostatné tabuľky – LNK, TMP, ...

DIM tabuľky - dimenzie

Hierarchická
časová
dimenzia

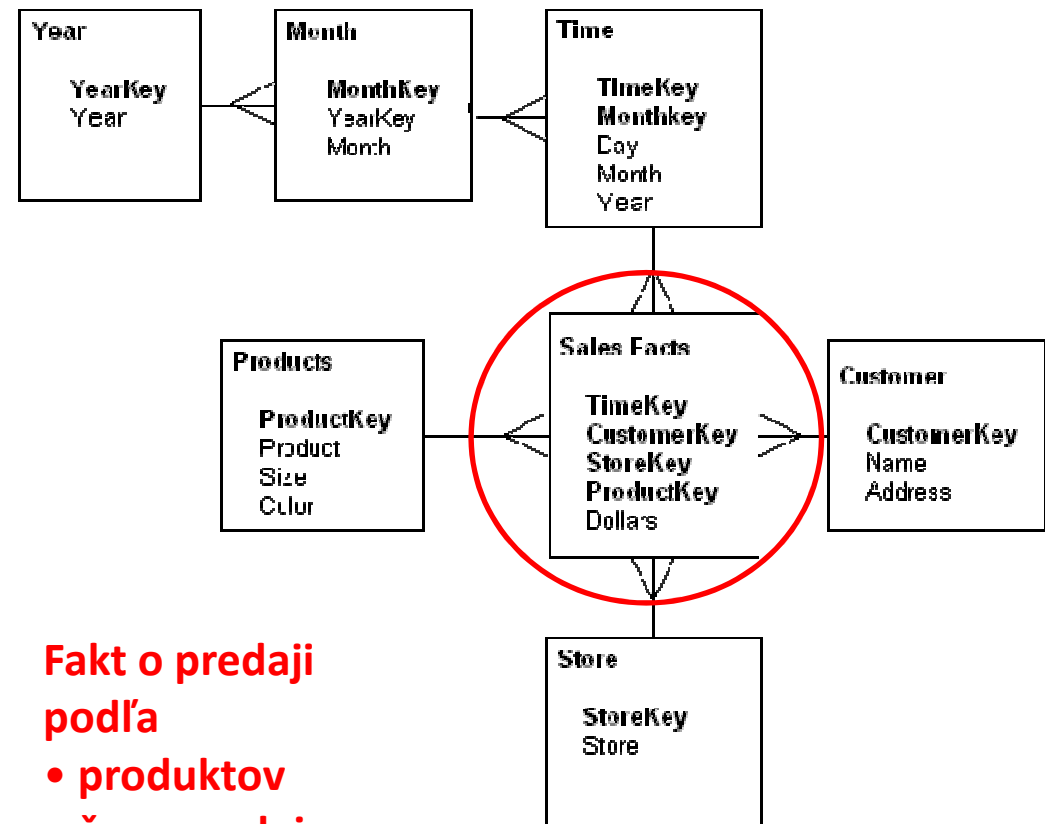
new
frontier
slovakia

- o Obsahujú **atribúty**
 - Napr. zákazník, produkt, dátum
- o Majú prirodzený primárny kľúč
- o Poskytujú opisné informácie k faktom
 - „Prístupové cesty“ k faktom
- o Môžu obsahovať hierarchie
- o Conformed dimensions
 - Zdieľané viacerými FCT tabuľkami



FCT tabuľky - fakty

- o Obsahujú **merania, metriky** alebo **fakty k business procesom**
 - Napr.obrat obchodu, množstvo tovaru na sklade
- o Primárny kľúč - dimenzie
- o Na základe odkazov do dimenzií poskytujú súbor hodnôt
- o Conformed FACTs – Zdieľané údaje medzi viacerými business oblasťami

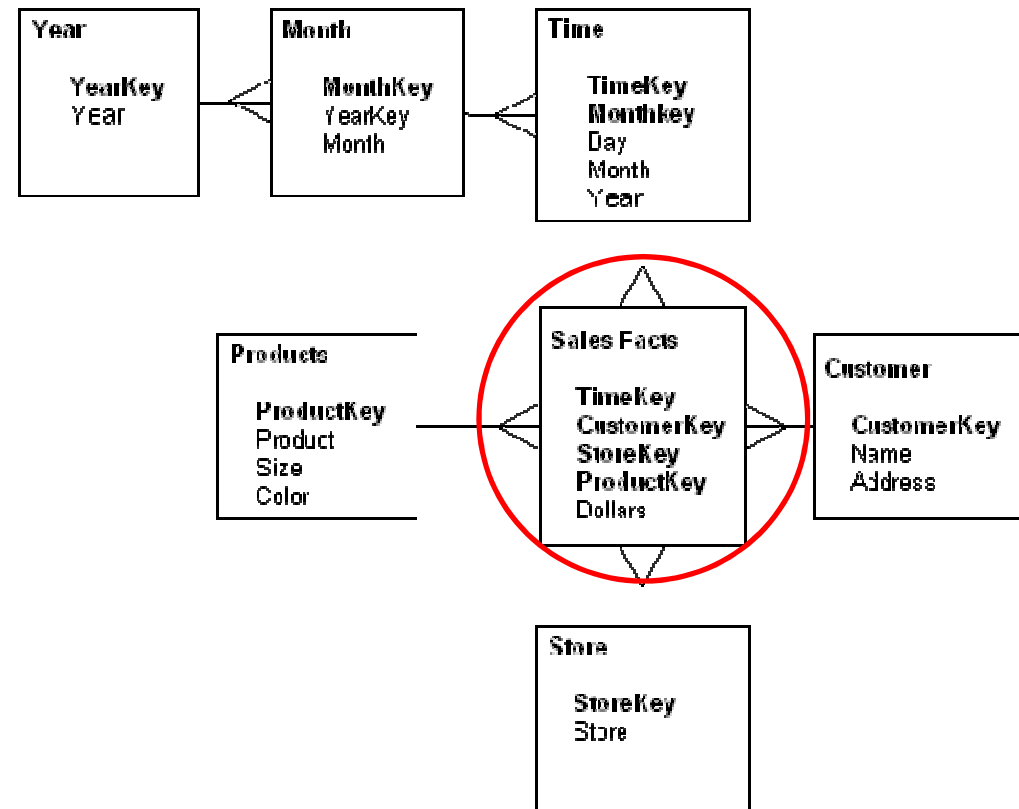


Fakt o predaji podľa

- produktov
- času predaja
- zákazníka
- predajného miesta

AGG tabuľky - agregácie

- o Obsahujú **agregované/sumárne** dáta
- o Špecifický typ FCT tabuliek
- o Dôvod existencie – performance, drill down
- o Analytici DWH často prístupujú
 - Najskôr k sumárnym dátam – rýchly prístup
 - Neskôr k detailným dátam – len pre určitú časť dát



Nad faktovou tabuľkou je možné vytvárať agregácie napr. podľa vyšších úrovní hierarchie časovej dimenzie

Granularita dát

Granularita = najjemnejšia úroveň detailu dát
(faktov aj dimenzií)

Nízka granularita

- (-) obrovské množstvo dát
- (+) väčšia množina realizovateľných dotazov - 15-30% dotazov je postavených na detailných dátach

Vysoká granularita

- (+) menšie množstvo dát
- (-) niektoré informácie sú už nedostupné
- (-) niektoré reporty nie je možné realizovať



Denormalizácia

= redundancia dát

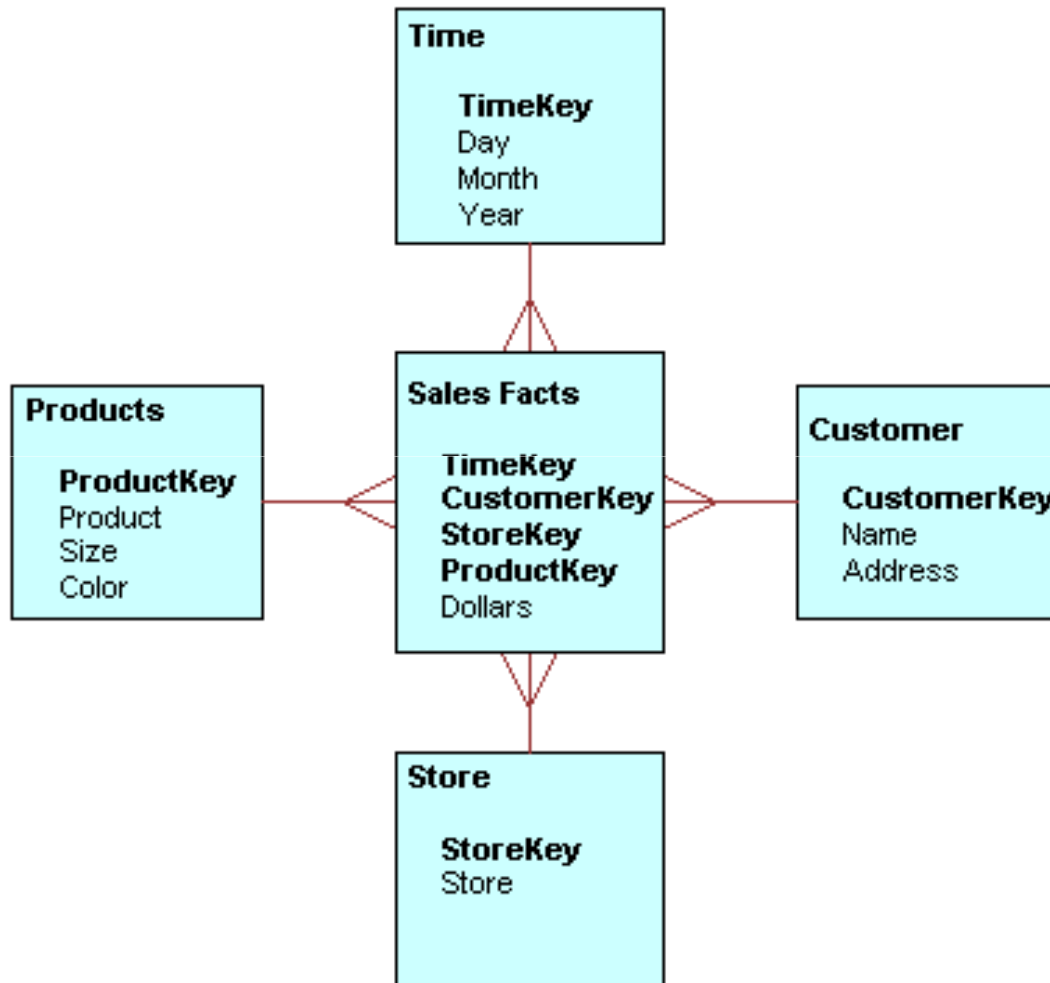
Cieľ

- o Redukcia JOIN operácií → zvýšenie výkonnosti

Spôsoby

- o Indikátory z dimenzií uložené aj pri faktoch
 - Nie sú nutné JOIN operácie faktov s dimenziami
- o Dopredu vypočítané hodnoty
 - Rýchlejšie dotazy
- o Predpočítané výstupné reporty
 - Rýchlejšie dostupné reporty

Star schéma

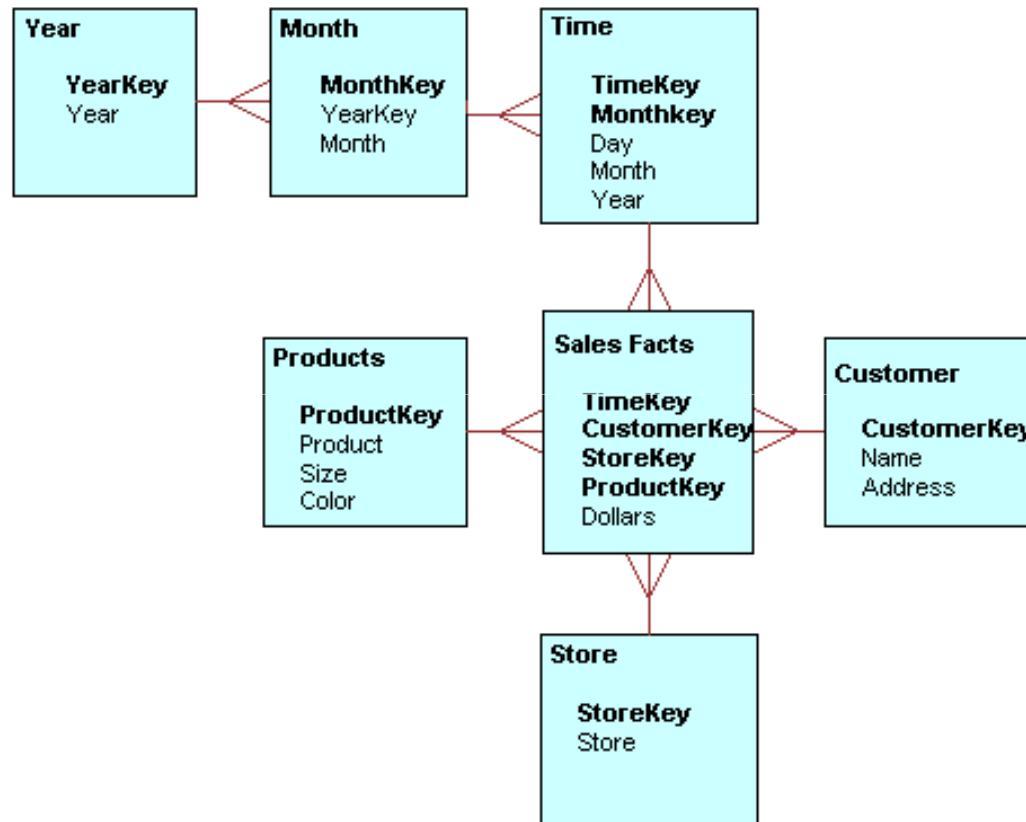


- o Business dimenzia obsiahnutá v jednej fyzickej tabuľke (vedie k denormalizácii dimenzií)

Constellation schéma

= star schéma s viacerými FCT tabuľkami (zdieľajúcimi dimenzie)

Snowflake schéma



- o Business dimenzia obsiahnutá aj vo viacerých fyzických tabuľkách (často vedie k zachovaniu normalizovaných dimenzií)

Snowstorm schéma

- = snowflake schéma s viacerými FCT tabuľkami (zdieľajúcimi dimenzie)

Agenda

Dátové modely – použitie v DWH

Dimenzionálny model – typy tabuliek

Ukážka modelu dátového skladu

Nástroje

Zhrnutie a diskusia



Agenda

Dátové modely – použitie v DWH

Dimenzionálny model – typy tabuliek

Ukážka modelu dátového skladu

Nástroje

Zhrnutie a diskusia



Nástroje

- o Sybase PowerDesigner
- o Rational Rose (IBM)
- o Enterprise Architect
- o Microsoft Visio

- o Eclipse UML2 Tools

SYBASE®

Rational® software

ENTERPRISE
ARCHITECT

Microsoft®
Office Visio 2007

Diskusia

new
frontier
slovakia

