

Business intelligence v praxi

Mgr. Peter Nagy

Business intelligence specialist

Intelligent Technologies s. r. o.

Obsah

- Predstavenie Intelligent Technologies
- Čo je business intelligence (BI)
- Architektúra BI riešení
- Dátový sklad (DWH), princípy loadu dát (ETL)
- Dátové modely



ŠKODA



Business Intelligence - BI

- Súhrn technológií, aplikácií a procesov
 - Zber dát
 - Normalizácia dát
 - Analýza dát
 - Prezentácia a interpretácia dát („Čo sa stalo a prečo?“)

Význam BI riešení pre firmy

- Integrácia a centralizácia firemných dát (DWH)
- Jednoduchý a rýchly prístup k dátam
- Kontrola dát
- Získavanie cenných informácií z vlastných dát (konkurenčná výhoda)

Používaný software



Služby:



Programy:



- Reporty
- Dátové modelovanie



- Reporty
- ETL projekty
- Dátové modelovanie



SQL Server
Management Studio

- Správa SQL databáz

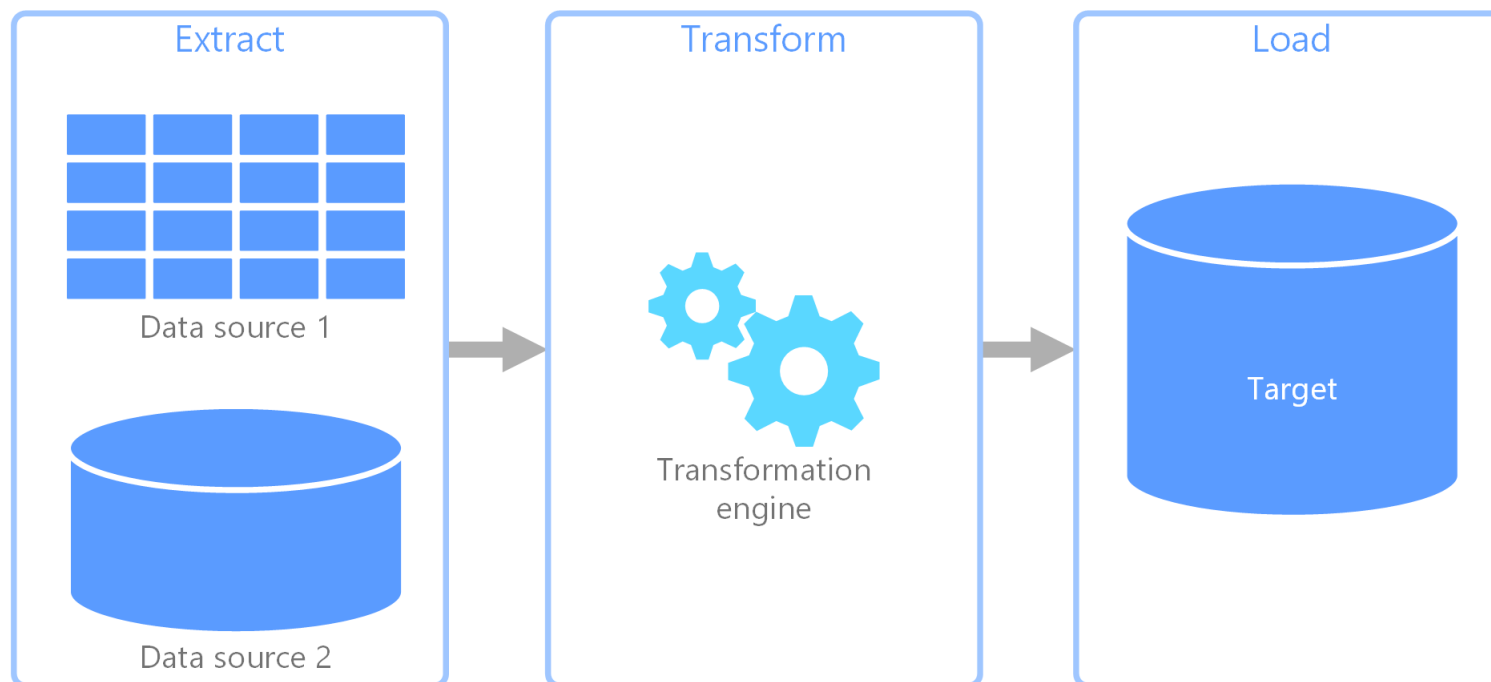
Najpoužívanéjšie DB objekty

- Tables
- Views
- Procedures
- Functions
- Indexes

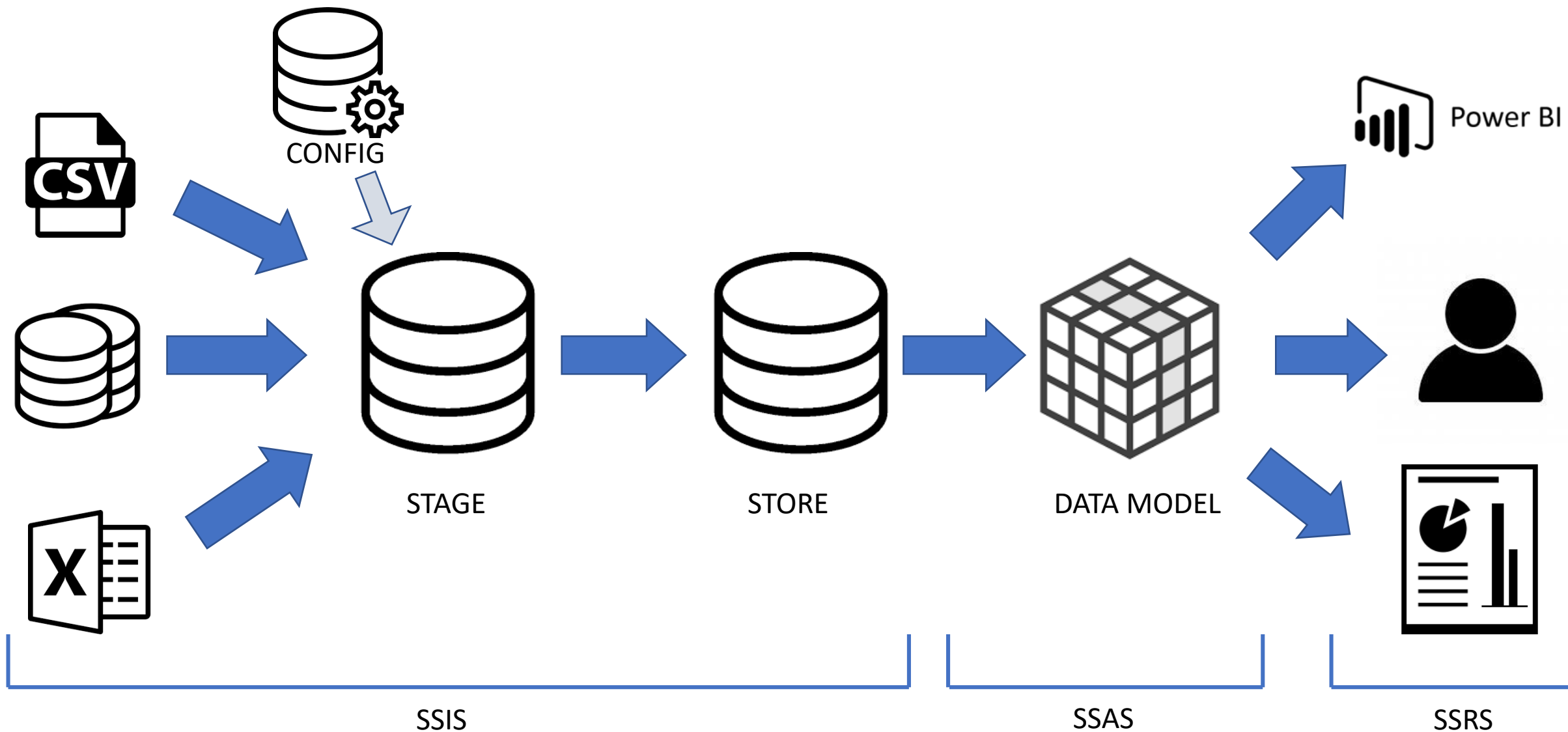
Základná architektúra BI riešení

1. Dátová vrstva (DWH, priamy zdroj dát)
2. Analytická vrstva (OLAP cube, TABULAR model)
3. Prezentačná vrstva (Stránkové reporty, Power BI)

ETL – „dátová pumpa“



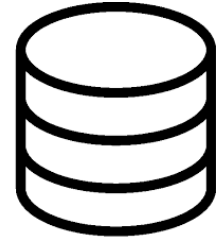
Architektúra BI riešení





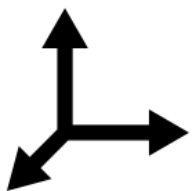
Stage DB

- Kópia dát zo zdroja
- Všetky atribúty
- Rovnaké názvy atribútov
- Tabuľky sa denne premazávajú



Store DB

- Transformované dáta
- Vybrané atribúty
- Zmena názvov atribútov
- Tabuľky obsahujú históriu
- Príprava dát pre dátový model



Dimenzné tabuľky (Dimension tables)

- Atribúty slúžiace na filtrovanie (text)
- Menej riadkov, viac stĺpcov
- Nížší objem rastu dát v čase
- Primary keys
- V loade ako prvé
- Napr.: Predajne, produkty, zákaznicke karty, dátum,...

dimStore
StoreDwId
Store Name
Store Address
Store Manager
Store Type

1:N



Faktové tabuľky (Fact tables, transactions)

- Atribúty obsahujú metriky (čísla)
- Viac riadkov, menej stĺpcov
- Vysoký rast objemu dát v čase
- Foreign keys
- Load až po dimeziách
- Napr.: Predaje, pohyby v sklade, vernostné body

factSales
StoreDwId
Amount

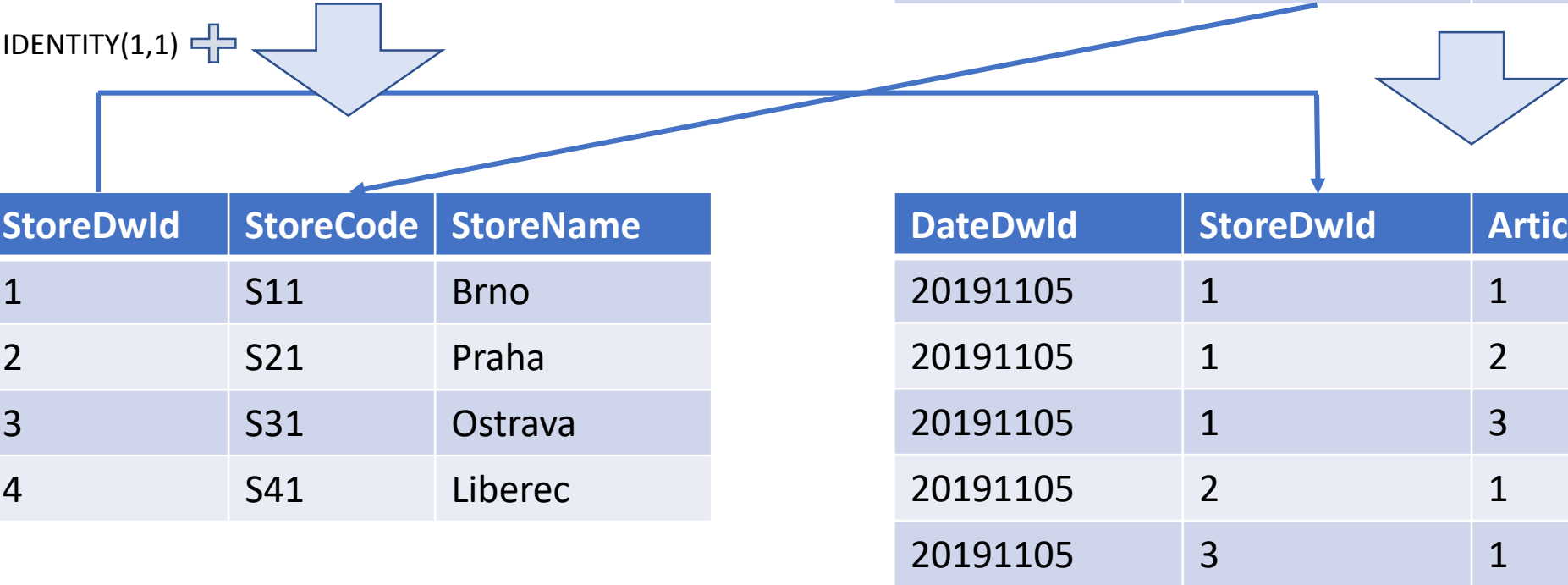
Princíp Dwld (Datawarehouse ID)

Dimenzia - Store:

StoreCode	StoreName
S11	Brno
S21	Praha
S31	Ostrava
S41	Liberec

Faktové záznamy - Sales:

Date	StoreCode	ArticleCode	Amount
2019-11-05	S11	12345	25,50
2019-11-05	S11	54321	80,50
2019-11-05	S11	67890	12,40
2019-11-05	S21	12345	51,00
2019-11-05	S31	12345	76,50



Extrakcia a Load Dát

- **Full load**

- Load celej tabuľky
- SCD = Slowly changing dimension
- Historical vs. Changing attributes

- **Incremental load**

- Load tabuľky so zarážkou (dáta za posledných X dní)
- Optimalizácia loadu objemných dát
- Schopnosť zachytiť zmeny v histórii za posledných X dní

Incremental Load

STAGE

Date	Data
2019-11-04	...
⋮	...
2019-10-04	...

1.



STORE

Date	Data
2019-11-03	...
⋮	...
2019-10-04	...
⋮	...
2010-01-01	...

2.



STORE

Date	Data
2019-11-04	...
⋮	...
2019-10-04	...
⋮	...
2010-01-01	...

1. **DELETE** STORE.Table
WHERE Date >= (SELECT MIN(Date)
FROM STAGE.Table)

2. **INSERT INTO** STORE.Table
SELECT *
FROM STAGE.Table

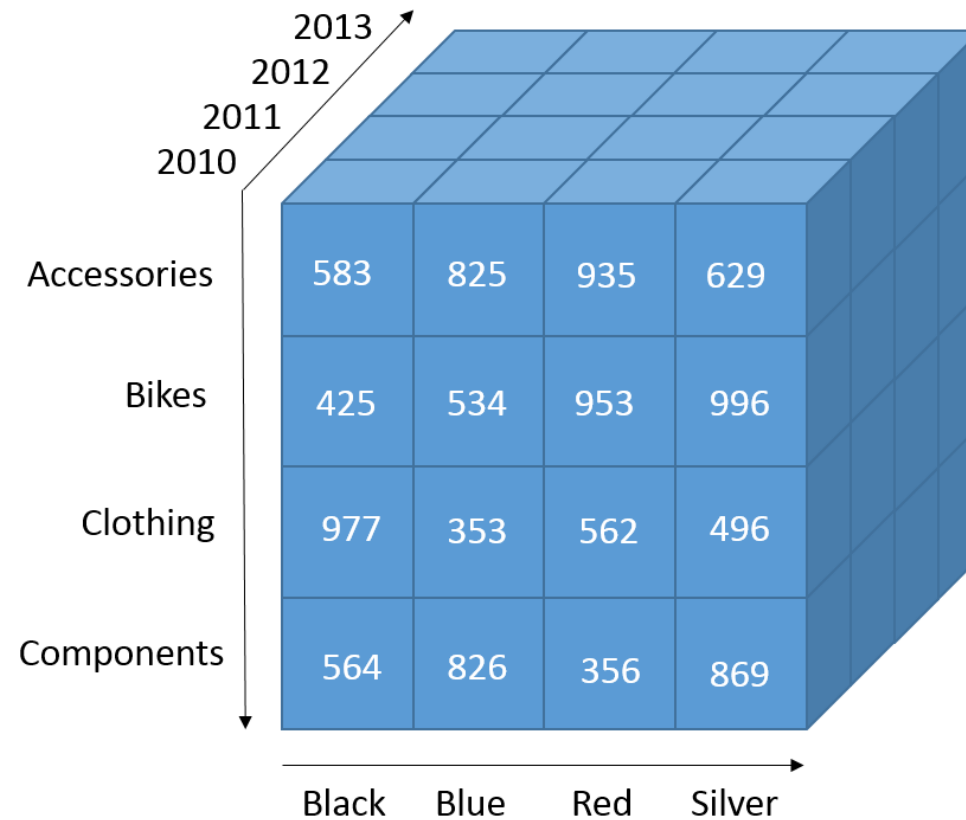
DWH/SQL best practices

- Úprava kódu
- Komentáre (why?)
- Myslieť na výkonosť v budúcnosti (Full Load vs. Incremental Load)
- Nepoužívať SELECT * v produkčnom kóde
- Indexy
- Používanie [] v MS SQL
- Pro tip – klávesové skratky ;-)

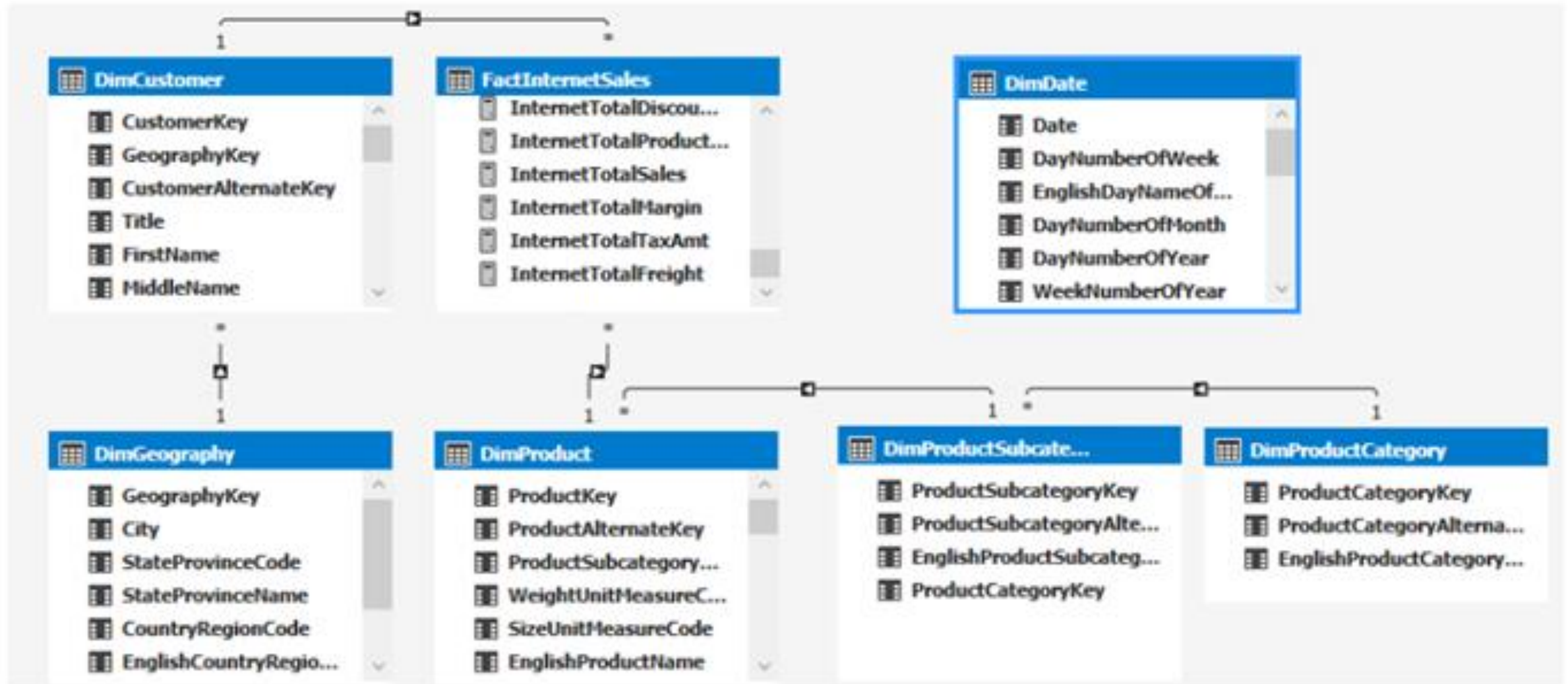
Dátový model

- Rozdielny význam od matematického modelu
- Obsahuje fyzicky dáta
- Optimalizácia výkonu
- Rozsiahle analýzy (browse - kontingenčné tabulky v Exceli)
- Podklad pre reportovacie nástroje

OLAP cube – multidimensional model



TABULAR model



Power BI

- Self-service analytics
- Interaktívne reporty
- Množstvo dátových zdrojov
- Široká ponuka vizualizácií
- Moderný prístup k dátovej analytike
- Mobilná aplikácia
- Real-time dáta

Ďakujem za pozornosť

