



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Název projektu	Rozvoj vzdělávání na Slezské univerzitě v Opavě
Registrační číslo projektu	CZ.02.2.69/0.0./0.0/16_015/0002400

Prezentace předmětu: **Business Intelligence**

Vyučující:
doc. Mgr. Petr Suchánek, Ph.D.



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Business Intelligence

Přednáška 3



**SLEZSKÁ
UNIVERZITA**

**OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ**

doc. Mgr. Petr Suchánek, Ph.D.

Komponenty BI – multidimenzionální databáze



- Základním předpokladem pro kvalitní výstupy BI jsou vstupní data;
 - Z existujících dat (např. o chování zákazníků, objemech prodeje v různých regionech) je třeba vytěžit maximum informací, jež mohou při správném využití poskytnout výraznou konkurenční výhodu;
 - Primární požadavky na data:
 - přesná a správná;
 - konzistentní;
 - kompletní;
 - neredundantní.
 - Data jsou prostřednictvím komponent BI transformována na informace
 - operativní;
 - analytické.
-

- Operativní informace
 - slouží pro realizaci obchodních a dalších transakcí v podniku. Jsou uloženy většinou v relačních databázích, zobrazují aktuální stav podniku a v průběhu jednoho dne se mohou i několikrát měnit. Příkladem může být např. účetnictví, data v dokumentech obchodních případů apod. Tyto data jsou chápána jako primární.
 - Analytické informace
 - zdrojem (vstupem) jsou informace vytvořené pomocí OLTP (On-Line Transaction Processing) a zpracovávají je systémy OLAP. Data jsou uložena v multidimenzionálních databázích obsahujících různé úrovně agregace dat. Jedná se o data a informace, které jsou zpracovávány a vyhodnocovány za delší časové intervaly s cílem získání časových řad, predikcí, srovnání za různá časová období apod.
-

Komponenty BI – multidimenzionální databáze



- Transakční databáze
 - ukládání operativních údajů;
 - vysoká strukturovanost tabulek;
 - problémem je decentralizace.
 - Nevýhody transakčních systémů pro analýzy
 - neuchovávají historické údaje;
 - nehomogenní struktura údajů;
 - dlouhý čas přípravy údajů;
 - obtížné hledání příčin a závislostí jednotlivých údajů;
 - analýza přímo v operačním prostředí snižuje výkon důležitých systémů.
 - příliš rozsáhlé výstupy.
-

- OLTP
 - technologie uložení dat v databázi, která umožňuje jejich co nejsnadnější a nejbezpečnější modifikaci v mnohauživatelském prostředí. Jedná se o přístup používaný v současné době v převážné většině databázových aplikací (historicky ještě poměrně nedávno dokonce ve všech databázových aplikacích).
 - OLAP
 - technologie uložení dat v databázi, která umožňuje uspořádat velké objemy dat tak, aby byla data přístupná a srozumitelná uživatelům zabývajícím se analýzou obchodních trendů a výsledků BI. Způsob uložení dat se svým zaměřením liší od běžněji užívaného OLTP (Online Transaction Processing), kde je důraz kladen především na snadné a bezpečné ukládání změn v datech v konkurenčním (víceuživatelském) prostředí.
-



- OLTP
 - detailní data;
 - význam ve chvíli zpracování;
 - častá změna dat;
 - transakční orientace;
 - výkonnost je důležitá;
 - vysoká dostupnost je důležitá;
 - redundance dat je nežádoucí;
 - slouží technicko-hospodářským pracovníkům;
 - předem známy požadavky na zpracování.

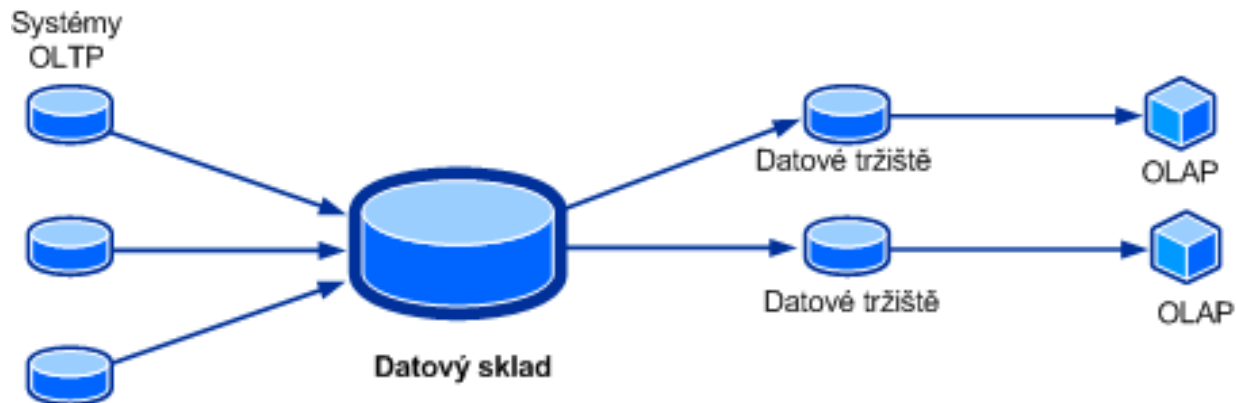


- OLAP
 - agregovaná data;
 - zpracování za období;
 - data téměř neměnná;
 - orientace na analýzu;
 - výkonnost není tak důležitá;
 - na vysoké dostupnosti příliš nezáleží;
 - redundance dat je běžná;
 - slouží především analytikům a manažerům;
 - většina požadavků není předem známa.

Komponenty BI – multidimenzionální databáze



- OLAP



Komponenty BI – multidimenzionální databáze

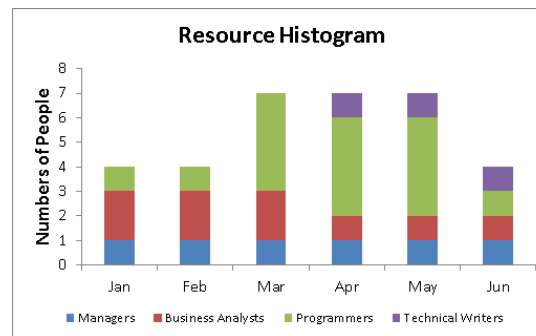
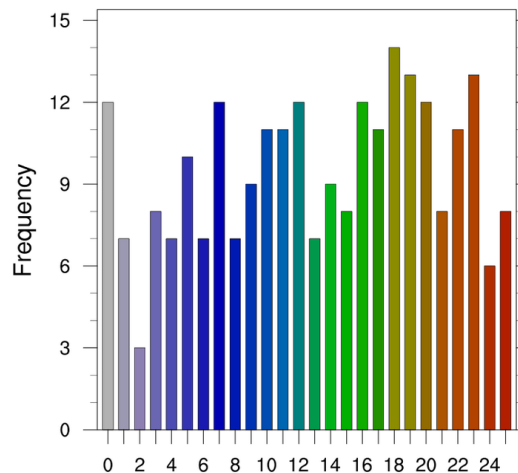
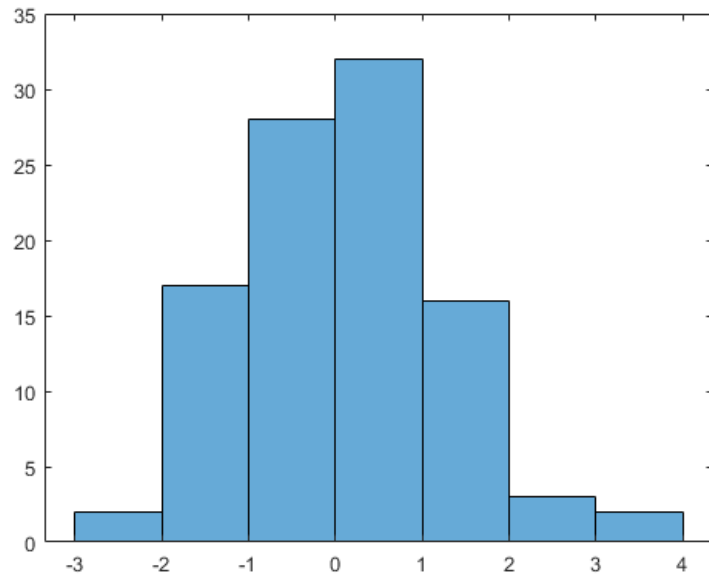


- Datová analýza
 - formulace dotazu.
- Získání relevantních dat z databáze
 - získání výsledků;
 - získání agregovaných hodnot.
- Vizualizace výsledků
 - zobrazení v 2D a 3D objektech;
 - co nejvíce závislostí najednou.
- Analyzování výsledků a formulování nového dotazu.

Komponenty BI – multidimenzionální databáze



- Prezentace dat - histogram



Komponenty BI – multidimenzionální databáze



- Prezentace dat – křížová tabulka

Year ▾ Month ▾ + ▾

Country	City	2015			2016		
		January	February	March	January	February	March
France	Lyon	22	35	21	15	29	22
	Marseille	13	41	11	21	40	11
	Paris	9	54	8	10	60	8
	Subtotal	44	130	40	46	129	41
Germany	Berlin	12	45	9	15	60	9
	Hamburg	21	23	19	24	44	19
	Munich	29	38	32	30	33	35
	Subtotal	62	106	60	69	137	63

Sum(Sales) ▾ + ▾

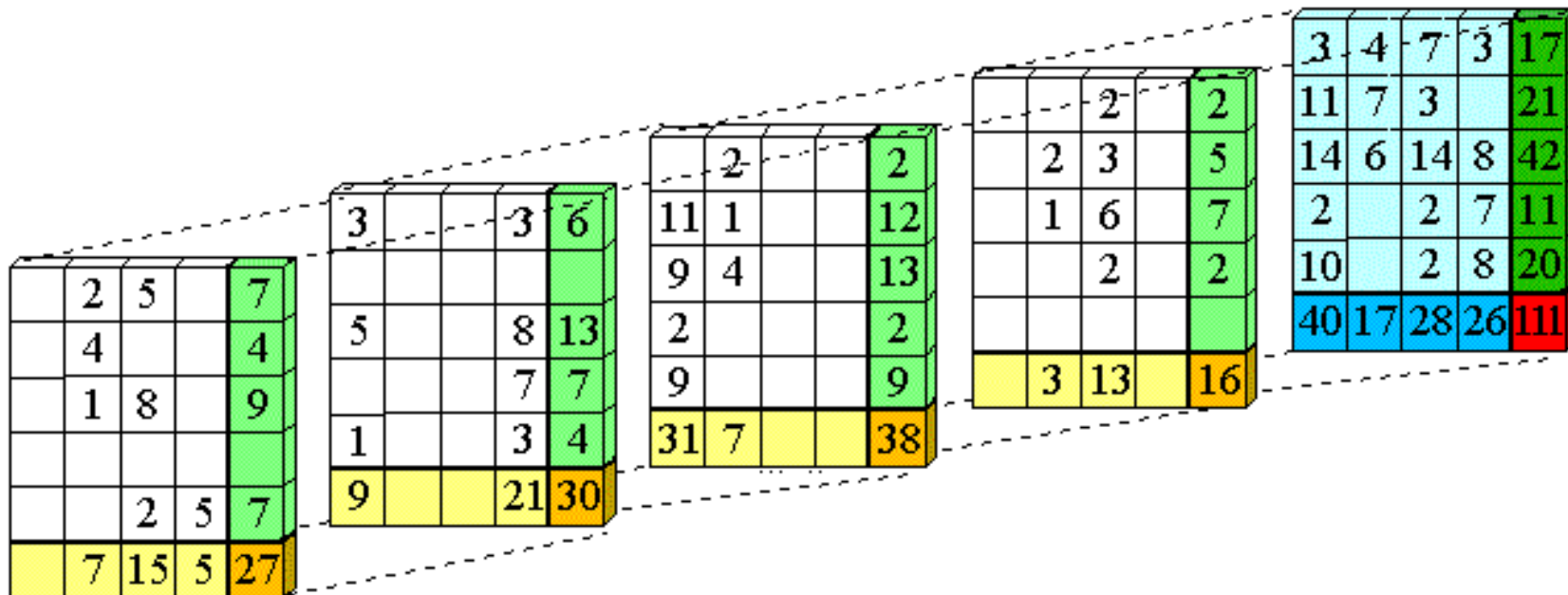
A3 ▾ = Součet z Částka

	A	B	C	D
1	Závod	Liberec ▾		
2				
3	Součet z Částka	Středisko ▾		
4	Sledované náklady ▾	stř. 01	stř. 03	Celkový součet
5	drobný inv.majetek	87	174	261
6	elektřina	125	119	244
7	materiál	145	154	299
8	pohonné hmoty	246	98	344
9	propagace	130	168	298
10	Celkový součet	733	713	1446
11				

Komponenty BI – multidimenzionální databáze



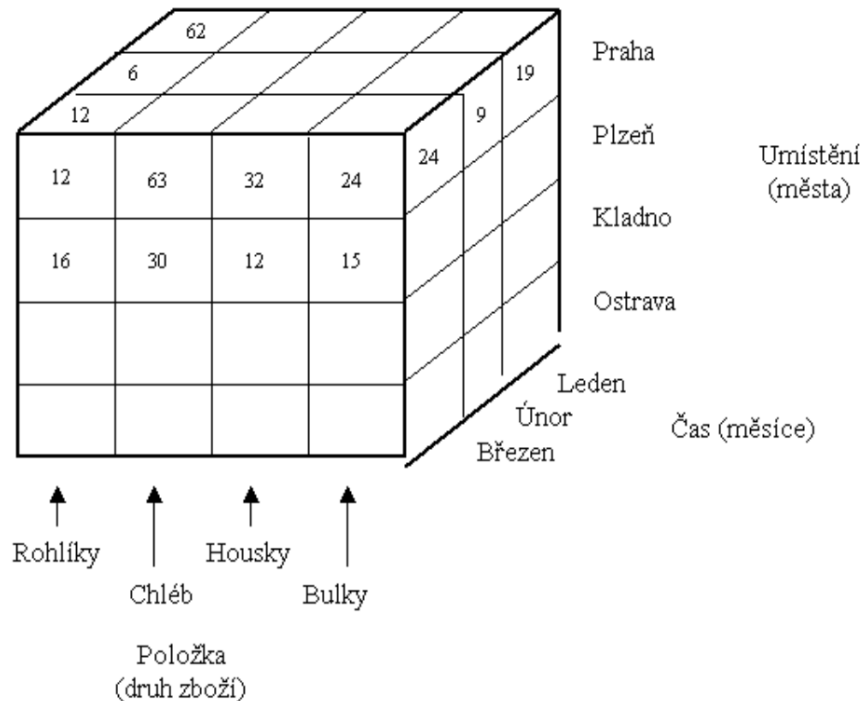
- Prezentace dat – datová kostka



Komponenty BI – multidimenzionální databáze



- Presentace dat – datová kostka



- Datová kostka
 - skládá se ze sady dimenzí a měr.
- Dimenze (rozměr) kostky
 - kategorie, vůči kterým chceme data agregovat a analyzovat;
 - vznikají z tabulek relačních databází;
 - typickými dimenzemi v multidimenzionálních databázích jsou
 - čas;
 - poloha;
 - výrobek.
 - může se skládat z řady úrovní, které dále zpřesňují údaje.

- Míry kostky
 - kvantitativní údaje, které chceme analyzovat;
 - odvozeny z tabulek relačních databází;
 - běžnými mírami jsou
 - prodeje;
 - výdaje;
 - ceny;
 - téměř každý kvantitativní údaj může být mírou multidimenzionální kostky.



- Metody uložení dat:
 - Multidimenzionální OLAP (MOLAP);
 - Relační OLAP (ROLAP);
 - Hybridní OLAP (HOLAP);
 - Dynamický OLAP (DOLAP).

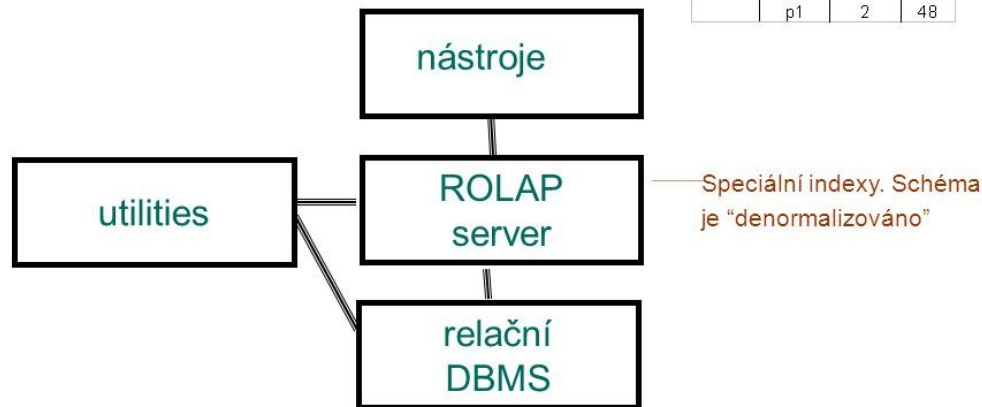
- ROLAP
 - vyžaduje uložení jak všech podrobných údajů, tak agregací v relační databázi;
 - všechna detailní data z kostky, která se najdou v tabulkách dimenzí a tabulkách faktů jsou ponechána v jejich přirozené relační databázi;
 - data se nepřesunují;
 - při ukládání agregací se vytvoří sumarizační tabulky, do kterých budou relační data ukládána službami OLAP pomocí jednoduchého SQL příkazu INSERT INTO;
 - služby OLAP vytvoří všechny tabulky a indexy samočinně;
 - podrobné údaje v kostce zůstávají beze změny.

- ROLAP

ROLAP Server

- Reláční OLAP Server

prodej	výřld	datum	sum
	p1	1	82
	p2	1	19
	p1	2	48

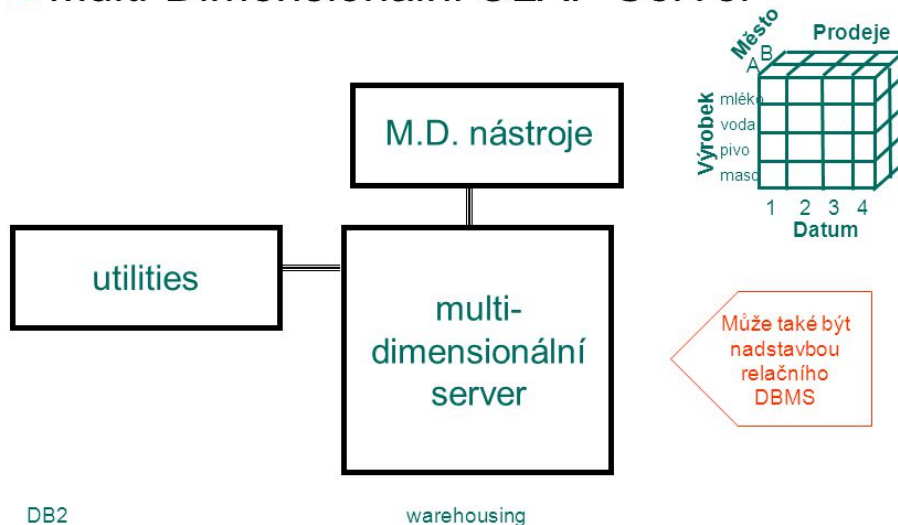


- MOLAP
 - všechny podrobné údaje z kostky ukládají ve vyhrazené multidimenzionální databázi;
 - relační data uložená v tabulkách dimenzí a v tabulkách faktů jsou zapsána do optimalizované multidimenzionální databáze;
 - k podrobným údajům o kostce do multidimenzionální databáze jsou uloženy také všechny agregace;
 - tato architektura OLAP je optimalizována pro zpracování dotazu ve službách OLAP a poskytuje nejlepší výkon ze všech čtyř metod ukládání dat;
 - vhodné pro malé a střední objemy dat.

- MOLAP

MOLAP Server

- Multi-Dimenzionální OLAP Server



- DOLAP
 - speciální typ OLAP;
 - multidimenzionální matice (kostka) je budována jako virtuální v RAM paměti;
 - výhoda
 - neomezená flexibilita.
 - nevýhoda
 - vysoké nároky na RAM paměť;
 - nutnost budovat kostku pokaždé znovu.

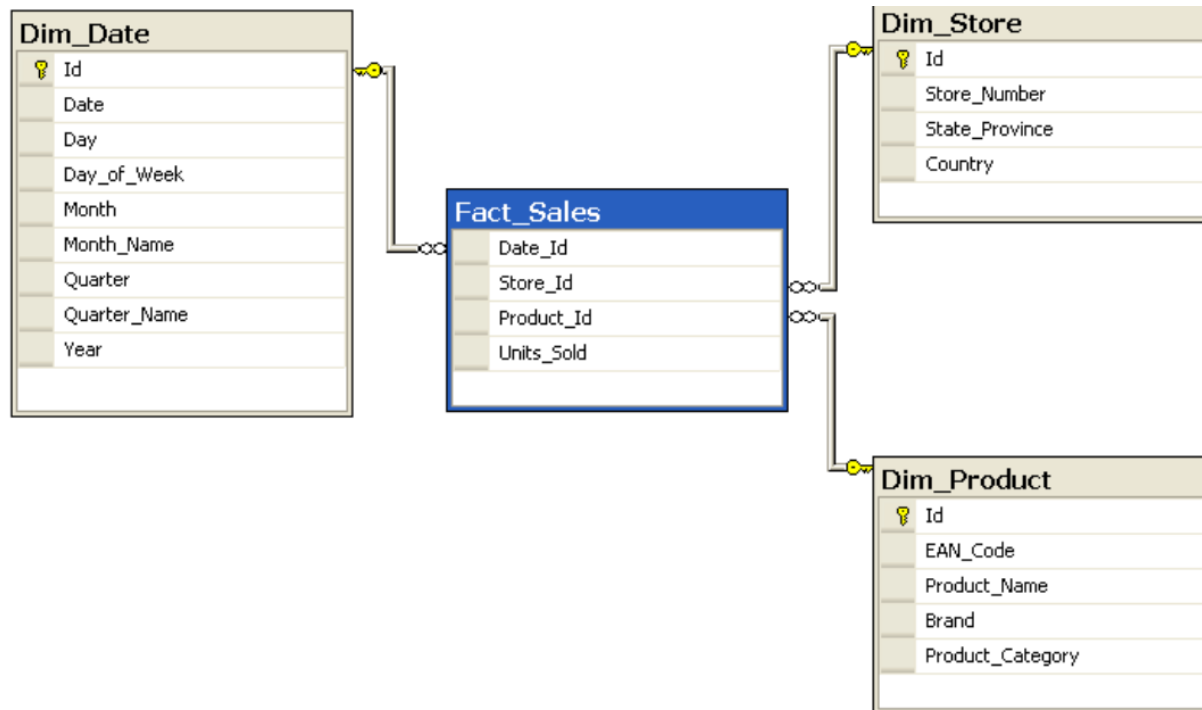
- HOLAP
 - kombinuje vlastnosti MOLAP a ROLAP;
 - Využívá vynikající zpracování MOLAP se schopností ROLAP pracovat s větším objemem dat;
 - ukládá data jak v relační, tak i multidimenzionální databázi;
 - data jsou agregována pomocí MOLAP strategie;
 - zdrojová data determinovaná objemem jsou uložena pomocí ROLAP strategie;
 - rychlé zpracování;
 - minimalizace požadavků na zpracování dat.

- STAR schéma
 - dovoluje relační DB simulovat multidimenzionální DB;
 - nejjednodušší schéma DW;
 - faktová tabulka uprostřed spojuje dimenzní tabulky okolo
 - faktové tabulky
 - nesou (zejména) číselné údaje – fakta;
 - zabírají nejvíce místa.
 - dimenzní tabulky
 - nesou atributy faktů – jejich popis, context;
 - související atributy v jedné tabulce;
 - minimalizace počtu dimenzních tabulek;
 - mají velké množství sloupců (atributů).

Komponenty BI – multidimenzionální databáze



- STAR

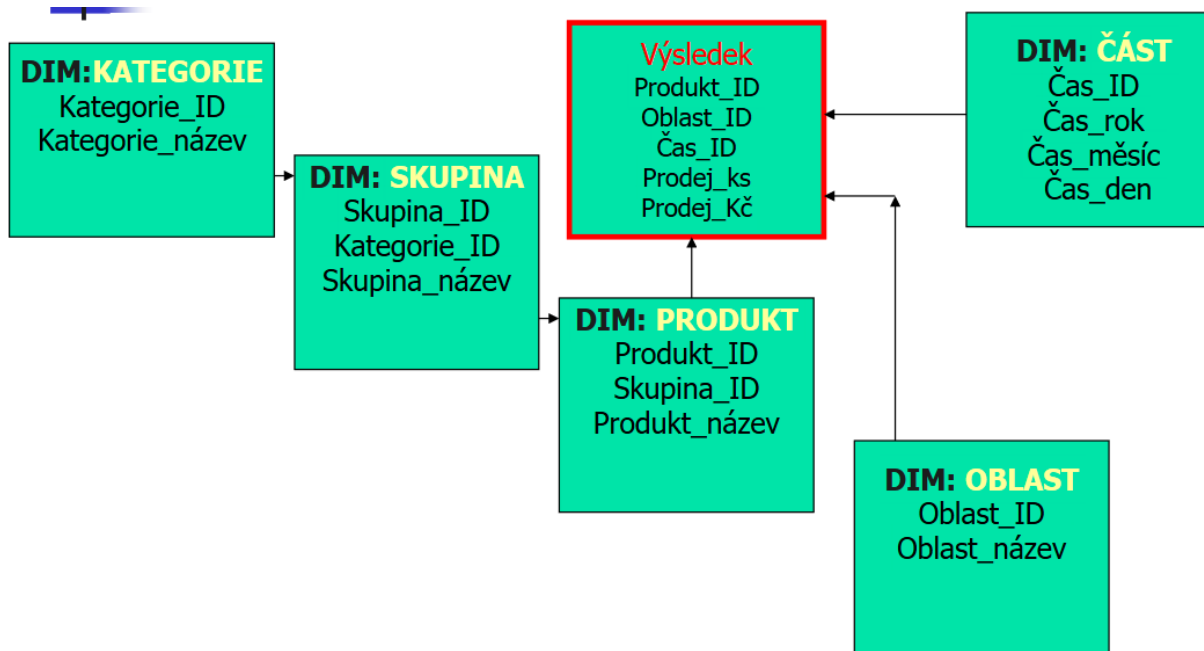


- SNOWFLAKE schéma
 - opět v centru faktové tabulky a okolo dimenzí;
 - dimenze jsou normalizované (do určité míry)
 - nikoli faktové tabulky.
 - dimenzní tabulky rozloženy procesem normalizace do několika propojených tabulek;
 - efektivní zejména
 - pro díravé (sparse) dimenze;
 - má-li dimenze velké množství atributů.

Komponenty BI – multidimenzionální databáze



- SNOWFLAKE





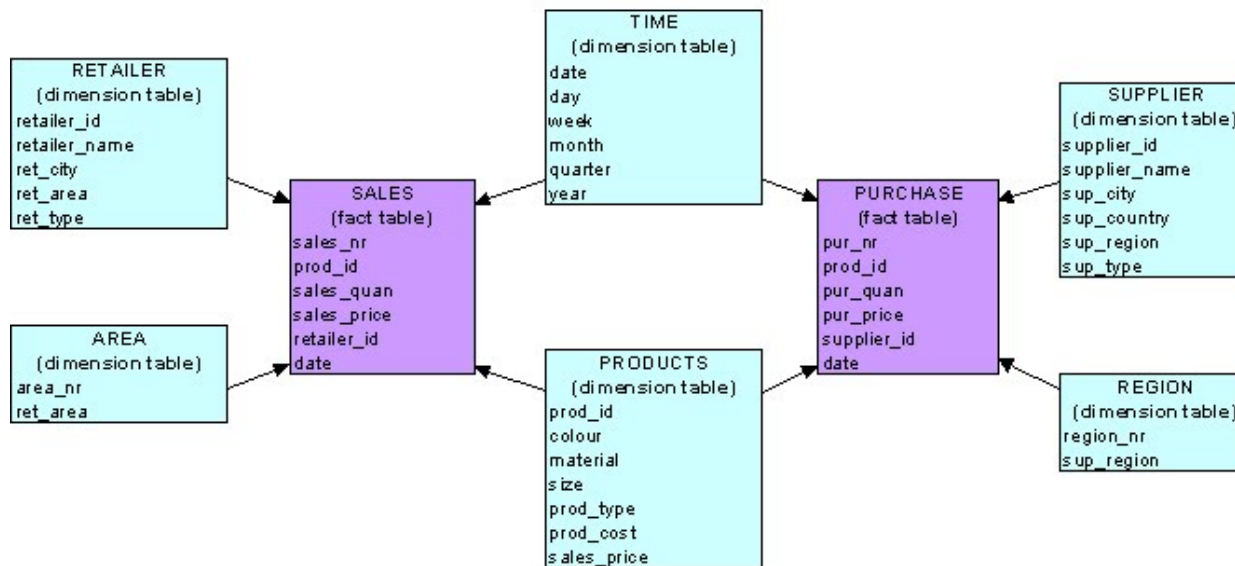
- FACT CONSTELLATION schéma
 - aplikace mohou vyžadovat více tabulek faktů, aby mohly sdílet tabulky dimenzí;
 - Toto schéma může být zobrazeno jako soubor hvězd a proto se nazývá „Constellation“ (galaxie nebo souhvězdí).

Komponenty BI – multidimenzionální databáze



- FACT CONSTELLATION schéma

Galaxy Model



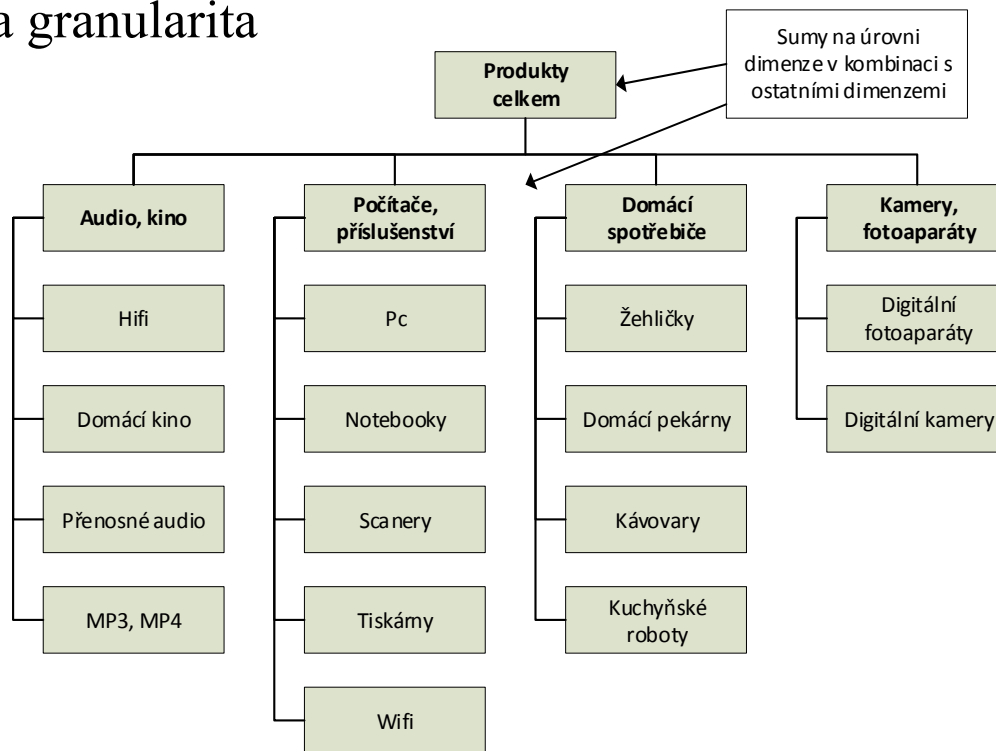
Komponenty BI – multidimenzionální databáze



- Agregace a granularita
 - jednotlivé prvky dimenzí se uspořádávají do hierarchické struktury;
 - jsou kategorizovány do skupin a podskupin;
 - úkolem BI je potom zajistit příslušné agregace a výpočty hodnot ukazatelů vycházejících z uživatelských požadavků;
 - databáze obsahují tabulky agregovaných hodnot ukazatelů a to i na nižších úrovních resp. v tzv. nižší granularitě (tento přístup umožní BI aplikacím rychlejší odezvu na analytické požadavky uživatelů);
 - princip granularity umožňuje dále využívat principy zvyšování a snižování tzv. úrovně detailu, kdy zpřístupňování dat na vyšší úroveň detailu se označuje jako drill down, v opačném směru jako drill up;
 - jako dimenze se standardně využívají ukazatele
 - čas (dny, měsíce, roky), obchodní zástupci, zákazníci, objednávky, produkty, dodavatelé, konkurence (resp. konkurenci), oblasti prodeje a celá řada dalších.
-

Komponenty BI – multidimenzionální databáze

- Agregace a granularita



- Technologická platforma
 - technologická platforma musí vycházet z definice konceptuálního a logického modelu;
 - Východiskem pro návrh technologické platformy tedy je:
 - stanovení granularity dat a jejich optimalizace v rámci datových skladů a datových tržišť v přímé vazbě na požadavky řízení podniku;
 - určení odhadu množství dat uložených v datových skladech a tržištích s přihlédnutím k časovému hledisku (stárnutí dat);
 - množství dat z předchozího bodu se odvíjí od nutnosti udržování dat z historie pro účely tvorby časových řad a od nich se odvíjejících analýz – vzniká zde potřeba určení požadavků na historii dat;
 - výběr databázového systému, technologie (OLAP) a potřebných nástrojů.

- NOVOTNÝ, O., POUR, J. a D. SLÁNSKÝ, 2005. *Business Intelligence – Jak využít bohatství ve vašich datech*. Praha: Grada. ISBN 978-80-247-6685-0.
 - LABERGE, R., 2012. *Datové sklady – Agilní metody a business intelligence*. Praha: Computer Press. ISBN 978-80-251-3729-1.
 - <http://isict.wikidot.com/wiki:operativni-informace>
 - https://cs.wikipedia.org/wiki/Online_Transaction_Processing
 - https://cs.wikipedia.org/wiki/Online_Analytical_Processing
 - www.ksi.mff.cuni.cz/~pokorny/dj/prezentace/3_67.pptx
 - [https://technet.microsoft.com/cs-cz/library/hh916543\(v=sc.12\).aspx](https://technet.microsoft.com/cs-cz/library/hh916543(v=sc.12).aspx)
 - <http://informacni-technologie.studentske.cz/2009/03/685-multidimenzionalni-kostka.html>
 - <http://slideplayer.cz/slide/3810636/>
 - http://homel.vsb.cz/~dan11/is_skripta/IS%202010%20-%20Danel%20-%20OLAP.pdf
 - http://zcu.arcao.com/kiv/db2/zkouska/zvesela/DB2%20-%20Kupka/DP_Marketa_Vlaskova_2006.pdf
 - <http://blog-mstechnology.blogspot.cz/2010/06/bi-dimensional-model-fact-constellation.html>
 - https://is.muni.cz/el/1433/podzim2012/PV005/um/PV005_11_data_warehouse.pdf
-



Děkuji za pozornost

Otázky?
