



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 7



**SLEZSKÁ
UNIVERZITA**

**OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ**

doc. Mgr. Petr Suchánek, Ph.D.

- Architektura BI
 - dána pozicí BI řešení v rámci komplexního IS/ICT řešení a uspořádáním, vazbami a parametry jednotlivých BI component;
 - V rámci architektury BI jsou definovány
 - aplikace BI;
 - technologie;
 - datové zdroje.

- U BI rozlišujeme architektury
 - procesní;
 - funkční;
 - datovou;
 - softwarovou;
 - hardwarovou;
 - technologickou.

U BI je situace složitější než u klasických architektur běžných IS/ICT, protože výstupy BI jsou více závislé na „transformačních“ technologiích (ETL, OLAP, apod.), které přenášejí a transformují data mezi různými úložišti, což zvyšuje požadavky na režii celého systému.

- V rámci definice architektury BI je nutné realizovat
 - vymežit produkční zdroje (aplikace, databáze);
 - stanovit v rámci struktury lokalizaci dočasného a operativního úložiště dat;
 - vymežit specializované nebo integrované nástroje ETL;
 - určit nástroje pro realizaci a správu datových skladů a tržišť;
 - určit nástroje pro správu OLAP databází;
 - specifikovat klientské nástroje a aplikace;

- V rámci definice architektury BI je nutné realizovat
 - určit datová rozhraní jednotlivých produktů nebo modulů v rámci komplexní struktury (jinými slovy rozhraní mezi komponentami BI a IS/ICT);
 - zajistit otevřenost řešení pro možnost implementace dalších produktů, nástrojů, modulů apod.;
 - definovat možnosti monitorování a správy provozu BI;
 - nastavit a implementovat helpdesk.

- Model BI řešení
 - základním východiskem je specifikace uživatelských požadavků a přírůstků v rámci komplexní koncepce IS/ICT řešení;
 - modely dělíme na
 - konceptuální;
 - logické;
 - fyzické.
 - při tvorbě konceptuálního a logického modelu, které jsou výchozí, se pracuje s dimenzionálním modelováním, na které navazuje modelování datového skladu.

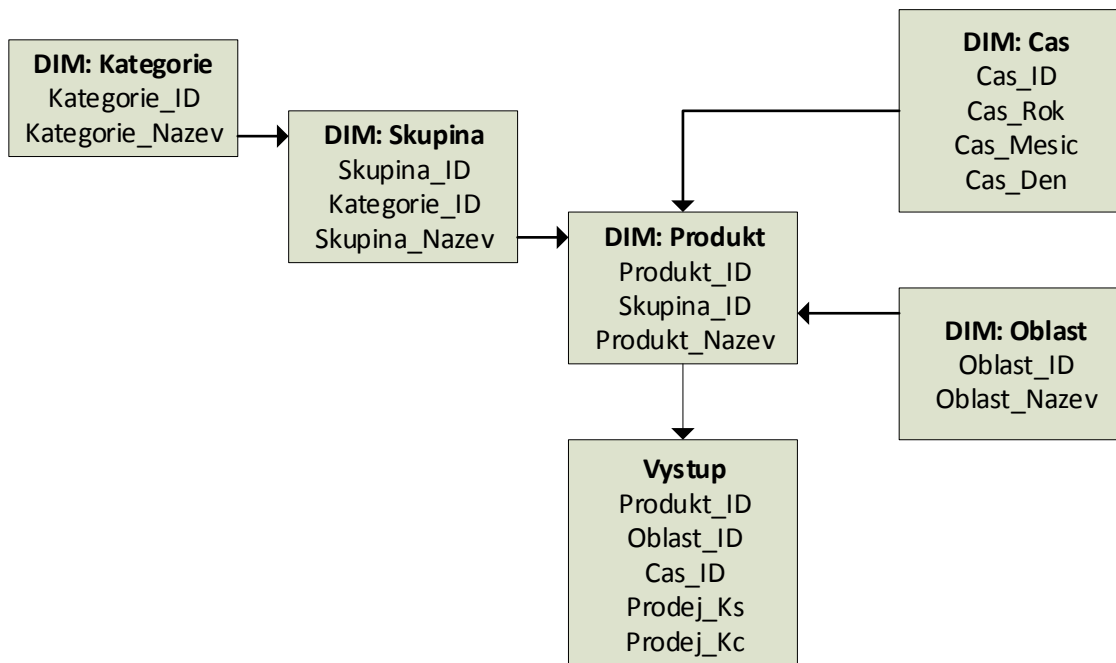
- Struktura dat
 - základem tradičního normalizovaného přístupu při tvorbě databází je optimalizace uložení dat s neexistencí redundance*.
 - při tvorbě databáze pro datový sklad oproti tomu není zcela zásadně důležité „šetření místem“, ale hlavním cílem je uložit data tak, aby se s nimi uživatelům dobře pracovalo (resp. uživatelským aplikacím).
 - V rámci dimenzionálního modelování se rozdělují činnosti do logických událostí a faktů na základě stanovených dimenzí
 - vymezení dimenzí, jejich obsahu, hierarchie prvků a charakteristik dimenzí;
 - stanovení soustavy sledovaných ukazatelů (faktů);
 - určení vazeb mezi ukazateli a dimenzemi.

*Informační nebo funkční nadbytek, například větší množství informace, prvků nebo zařízení, než je nezbytné.

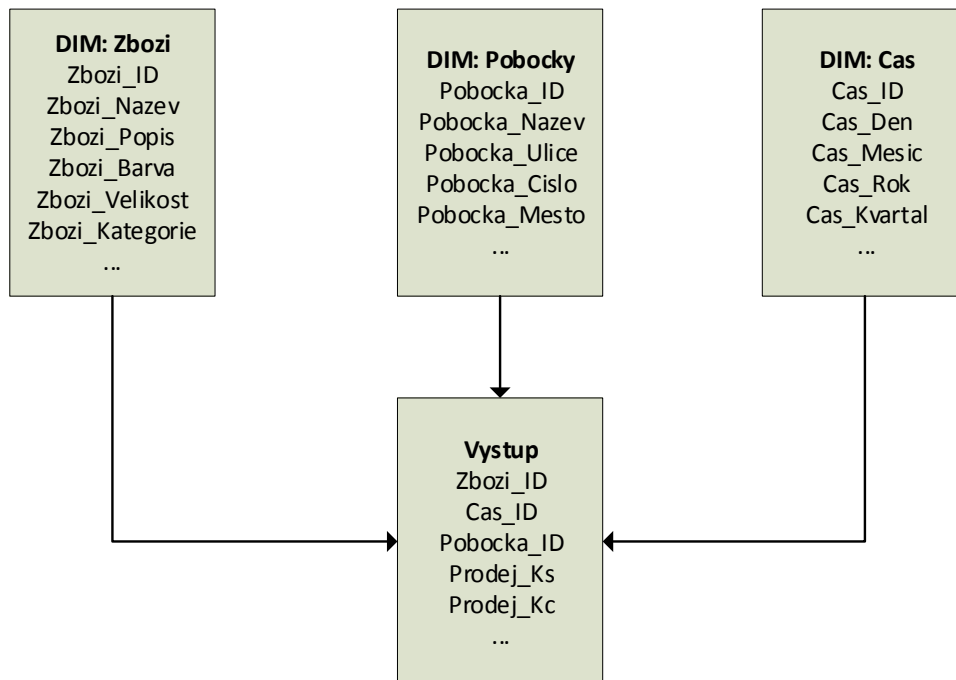
- Struktura dat – dimenzionální modelování
 - schémata se navrhují tak, že jednotlivé činnosti se rozdělují do logických událostí a faktů a nastavují se příslušné dimenze;
 - soustava sledovaných ukazatelů (faktů) je tvořena tabulkou nebo množinou tabulek obsahujících například obchodní ukazatele, u kterých má smysl sledovat časový horizont;
 - může jít například o objednávky, obraty, zisk, dodávky, bankovní operace apod.;
 - tyto tabulky jsou tvořeny nevlastními (cizími) klíči do tabulek dimenzí a množinou příslušných hodnot. Obsah tabulek faktů je neměnný, protože jde o zachycení historických hodnot.
-

- Struktura dat – dimenze
 - analytické hledisko
 - založeno na určení a hodnocení sledovaných ukazatelů.
 - infromatické hledisko;
 - zajímá nás struktura dat prezentovaná obecně jako databázová tabulka se záznamy o prvcích dimenze (zboží, zákazníci, lokality prodeje apod.).
 - dimenzionální schémata jsou realizována do struktur
 - „hvězda“ – star;
 - „sněhová vločka“ – snowflake;
 - „fact constellation“ – galaxie.

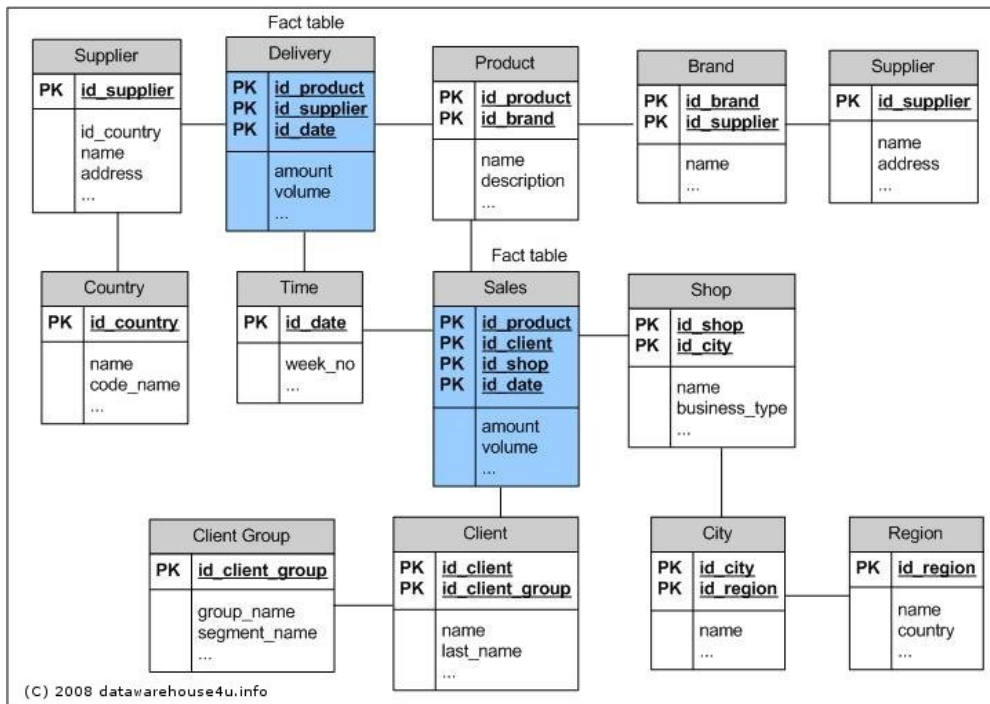
- Struktura dat – SNOWFLAKE



- Struktura dat – STAR

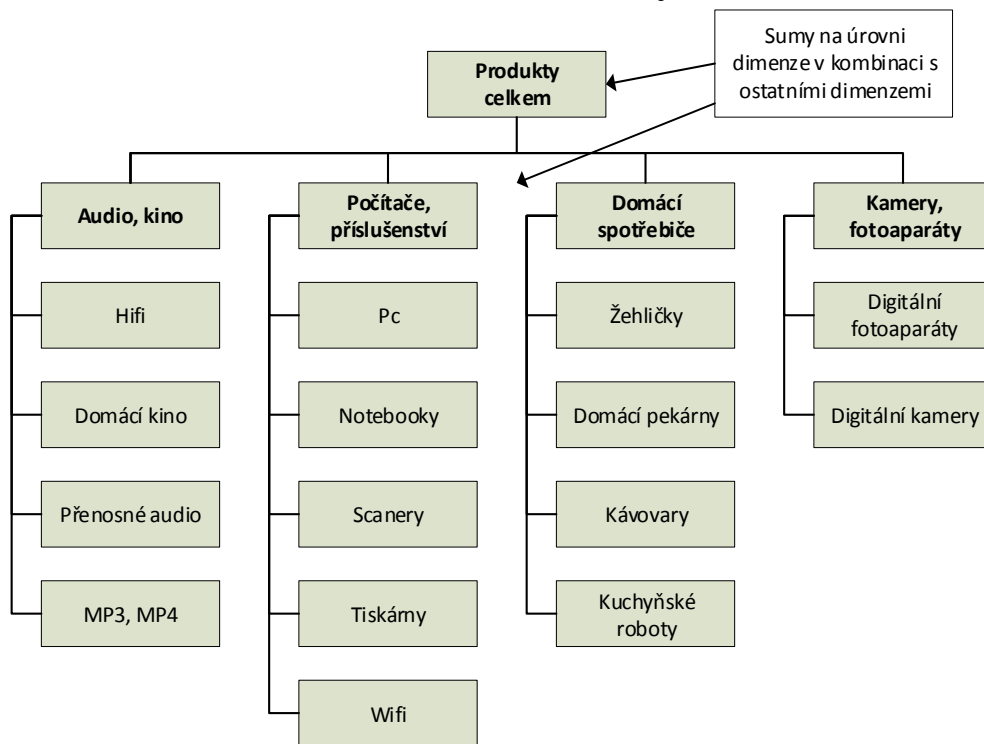


- Struktura dat – Fact Constellation



- Struktura dat – agregace a granularita
 - jednotlivé prvky dimenzí se uspořádávají do hierarchické struktury a jsou kategorizovány do skupin a podskupin;
 - úkolem BI je potom zajistit příslušné **agregace** a výpočty hodnot ukazatelů vycházejících z uživatelských požadavků;
 - BI - obsahuje tabulky agregovaných hodnot ukazatelů a to i na nižších úrovních resp. v tzv. nižší **granularitě**
 - drill down
 - zpřístupňování dat na vyšší úroveň detailu;
 - drill up
 - v opačném směru jako drill down.
-

- Struktura dat – Ukázka struktury dimenze Produkty



- Technologická platforma
 - vychází z konceptuálního a logického modelu;
 - úkolem je zajistit funkčnost celého systému zejména z hlediska rychlosti odezvy BI řešení a korektnosti výstupů;
 - Korektnost zde představuje
 - dodání odpovědi na uživatelský požadavek;
 - požadovanou kvalitu odpovědi.

- Technologická platforma – východiska
 - 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.

- Technologická platforma – úlohy při návrhu
 - definice technologické architektury s konkrétní specifikací technologických komponent a vzájemných vazeb;
 - vytvoření fyzického modelu datového skladu;
 - stanovení paměťové kapacity datového skladu s přihlédnutím ke zvyšování množství dat v jeho databázích.

- Technologická architektura
 - musí být schopna zajistit
 - přístup k datům v produkčních (zdrojových) databázích;
 - transformaci dat do formy odpovídající modelu datového skladu;
 - zpřístupnění dat uživatelům.

- Technologická architektura
 - back room
 - ta část datového skladu, ve které probíhají jeho vnitřní procesy skryté koncovým uživatelům;
 - diskový prostor, ve kterém se zpracovávají zdrojová data do podoby definované dimenzionálním modelem;
 - obsahuje katalog metadat - popisuje obsah datového skladu, udává zdroje dat a jsou v něm zakotvena pravidla pro transformaci dat (řízení extrakce, čištění, ukládání).
 - front room
 - prezentační servery, na kterých jsou uložena data za účelem přímého přístupu a dotazování uživateli.

- Technologická architektura
 - východisko pro tvorbu datového skladu;
 - prvním krokem je sběr a sumarizace požadavků na technologickou architekturu;
 - co je nutné vědět
 - Jaké softwarové nástroje jsou v podniku upřednostňovány?
 - Jaké jsou v podniku využívány stávající technologické platformy a standardy aplikací, databází, apod.?
 - Jaké jsou plány, priority a cíle v oblasti IS/IVT v podniku?

- Technologická architektura
 - Tvorba modelu technologické architektury
 - grafické zobrazení technologické architektury s textovou dokumentací zachycující všechny nutné komponenty;
 - Stanovení fází implementace technologické architektury
 - pro tuto část je nutné vzít v potaz
 - všechny komponenty architektury;
 - předpokládané lidské zdroje;
 - finanční zdroje;
 - priority z hlediska celkového harmonogramu implementace.
 - u každé fáze musí být jednoznačně určené konkrétní výstupy.

- Technologická architektura
 - Finální dokument plánu technologické architektury - obsahuje
 - souhrn požadavků na technologickou architekturu;
 - model technologické architektury;
 - detailní popis fází implementace modelu.
 - Tvorba plánu infrastruktury
 - detailní plán infrastruktury zachycující všechny souvislosti a potřeby technologické architektury v návaznosti na
 - prostor;
 - energii;
 - software;
 - hardware;
 - síť;
 - apod.

- Technologická architektura - fyzický model DS
 - DS
 - podstatné zajistit uložení dat tak, aby byly minimalizovány I/O operace, které jsou v celé řadě případů hlavní podíl na prodloužení doby zpracování uživatelského požadavku resp. dotazu.
 - cílem tedy je
 - zorganizovat,
 - uspořádat,
 - doplnit podpůrnými informacemi (mohou být například indexy).
 - data tak, aby v návaznosti na předpokládané funkce, aplikace a dotazy byla zajištěna co nejmenší doba odezvy.
-

- Technologická architektura - fyzický model DS – podpora minimální doby odezvy
 - verifikací nebo změnou granularity ukládaných dat;
 - rozdělením tabulek (partitioning) do logicky souvisejících celků s cílem zrychlení přístupu k datům (zejména u velkých tabulek, s dlouhou dobou scanování resp. procházení jednotlivých záznamů);
 - sloučením tabulek (merging) (snížení přístupů do různých tabulek);
 - vytvořením datových polí (array) (jednoznačná uspořádaná množina, kdy se obvykle k jednotlivým prvkům přiřazují konkrétní I/O operace
 - obvykle se využívá v případech, kdy se v pravidelných intervalech zpracovávají stejné dotazy).

- Technologická architektura - fyzický model DS – podpora minimální doby odezvy
 - duplikováním dat (redundance) (u datových skladů obvykle nejsou problémy s paměťovou kapacitou, proto mohou být duplicity v nutných a vhodných případech využity);
 - využitím odvozených dat;
 - kategorizace dat podle předpokladu četnosti přístupu k těmto datům;
 - předběžným zpracováním dat (například kalkulace);
 - indexováním dat (standardní i tzv. kreativní, což jsou indexy označující nastavené, jinými slovy dopředu definované pohledy z hlediska požadavků uživatelů).
-

- Technologická architektura – velikost DS
 - velikost DS – dynamická – vesměs rostoucí tendence;
 - pro výpočet velikosti datového skladu je možné využít CASE nástroje a vypočítat potřeby zcela přesně
 - to předpokládá mít tento software k dispozici, což není nezbytné.
 - primární je vypočítat nezbytnou minimální velikost
 - klíčovým ukazatelem je, zda výpočet budeme realizovat pro předpokládanou implementaci
 - relační databáze;
 - multidimenzionální databáze.

- Technologická architektura – velikost DS – příklad
 - obchodní firma
 - síť 250 prodejen;
 - datová strukturu se třemi dimenzemi;
 - předpoklad udržování tříleté historie dat.
 - v následujících tabulkách nejsou obsaženy výpočty beroucí v potaz nároky na tabulky dimenzí resp. číselníky, jejichž velikost je však vzhledem k celkové velikosti databáze významně menší (velikost odvodit z předchozích řešení);
 - vedle číselníků je nutné k celkové minimální velikosti ještě připočítat nároky na logy, dočasné soubory (temporary files), rollback apod.

- Technologická architektura – velikosti relační databáze

	Parametr	Vstup	Výstup
1	Dimenze CAS	3 roky, (3 x 365 dnů)	1095 dnů
2	Dimenze PRODEJNY	250 prodejen (data předávaná denně)	250 prodejen
3	Dimenze PRODUKTY	23 000 druhů (denně cca 2 300)	2300 produktů
4	Počet záznamů v tabulce faktů	1095 dnů * 250 prodejen*2300 produktů	629 625 000 záznamů
5	Délka záznamu, velikost atributů v tabulce faktů	3 atributy*4B + 5 atributů*8B	52 bytů
6	Velikost tabulky faktů	629 625 000 záznamů*52 bytů	32 740 500 000 bytů tj. cca 33 GB
7	Velikost indexů	cca 200 % základu	cca 65 GB
8	Celkem	základ + indexy	cca 98 GB

- velikost multidimenzionální databáze

	Parametr	Vstup	Výstup
1	Dimenze CAS (4 úrovně hierarchie)	3 roky, (3 x 365 dnů)	1095 dnů
2	Dimenze PRODEJNY (6 úrovní hierarchie)	250 prodejen (data předávaná denně)	250 prodejen
3	Dimenze PRODUKTY (7 úrovní hierarchie)	23 000 druhů (denně cca 2 300)	2300 produktů
4	Počet záznamů v tabulce faktů	1095 dnů * 250 prodejen*2300 produktů	629 625 000 záznamů
5	Velikost buňky	32 B	52 bytů
6	Základní velikost MDDB	629 625 000 záznamů*32 bytů	20 148 000 000 bytů tj. cca 20 GB
7	Velikost skutečné MDDB (např. pro MS OLAP Services)	$((2*\text{počet všech úrovní})+(4*\text{počet ukazatelů}))*\text{počet řádků}/3$ tj.: $((2*(4+6+7)+4*4)*629\,625\,000)/3$	10 493 750 000 (cca 10,5 GB)
8	Dočasné soubory	100 % základu	cca 10,5 GB
9	Celkem	základ + dočasné soubory	cca 21 GB

- Technologická architektura – velikost DS
 - proč jsou nároky na velikost relační databáze obecně vyšší než u využití databáze multidimenzionální?
 - u multidimenzionální databáze jsou data uložena s využitím principu vícerozměrné matice, kdy jednotlivé hodnoty jsou přístupné přímo pro danou kombinaci prvku dimenzí.

- Relaçní a multidimenzionální databáze

Relaçní databáze	Multidimenzionální databáze
Delší doba odezvy.	Kratší doba odezvy.
U analytických aplikací neposkytuje dostatečnou flexibilitu a vývoj aplikací je tudíž náročnější.	Nabízí efektivnější vývoj analytických aplikací.
Velké objemy dat s uloženíím na vyšší úrovni detailu.	Nižší nároky na objem dat a pracuje na nižší úrovni detailu.
Je vhodná pro větší objemy dat.	Není vhodná pro extrémně velké objemy dat.
Lepší flexibilita při změnách v dimenzích.	Složitější realizace změn dimenzí.
Složitější aktualizace agregovaných hodnot.	Snadná aktualizace agregovaných hodnot.

- **Klientské aplikace**
 - uživatelské rozhraní, které vedle délky odezvy sehrává důležitou roli při hodnocení systému jako celku uživateli;
 - důležité je
 - jaký komfort aplikace nabízejí uživatelům při definici dotazů;
 - jak se výsledky dotazů uživatelům zobrazí;
 - jaký je stupeň flexibility umožňující uživatelům rychlé a snadné změny v definici
 - změn dimenzí;
 - třídění dat;
 - formy a formáty jejich zobrazení;
 - přepočty dat;
 - změny analytických pravidel;
 - apod.

- **Klientské aplikace**
 - nutná existence množiny standardních reportů
 - žádaných uživateli ve stejné struktuře v pravidelných časových intervalech (denně, týdně, měsíčně);
 - v okamžiku změny nějaké aktuální hodnoty vyžadující regulační zásah (například překročení limitních stavů zásob, disponibilních kapacit apod.
 - realizovány jako automaticky generované zprávy).
 - primární požadavek
 - snadná ovladatelnost (často hovoře-no o požadavku na intuitivnost);
 - kvalitní možnosti prezentace (grafy, různé typy a formáty tabulek, klikací mapy atd.).

- **Klientské aplikace – využití pro různé účely**
 - **Analýzy a reporty**
 - provádějí se obvykle v SQL přímo nad datovým skladem s využitím například Esperant nebo ROLAP nástrojů typu Oracle Discoverer, Star Tracker nebo dalších.
 - **Analytické aplikace**
 - realizované nad multidimenzionální databázi (OLAP) poskytující zejména kontingenční tabulky, grafy, klikací mapy apod. s možností snadné dynamické konfigurace z hlediska různých agregačních úrovní dat, výběrů, výjimečných stavů, scénářů, vše se striktním požadavkem na minimální dobu odezvy.

- **Klientské aplikace – využití pro různé účely**
 - **Klientské nástroje**
 - představují rozhraní k relačním a multidimenzionálním databázím datového skladu;
 - jsou primárně určeny k vizualizaci dat s maximální jednoduchostí definice požadavků ze strany uživatelů;
 - jsou koncipovány tak, aby nevyžadovaly znalost nějakého dotazovacího jazyka (SQL, MDX), ale samozřejmě se nejde vyhnout požadavku na znalost syntaxe a sémantiky dat v datovém skladu;
 - definice dimenzí, úrovní, ukazatelů apod. jsou samozřejmě na uživateli.
 - například ProClarity, Oracle Discoverer, Essbase/IBM DB2 OLAP, MS Reporting Services, Business Objects, Microstrategy, PowerPlay, Impromptu (Cognos), Media (Speedware), Express (Oracle), Power Dimensions (Sybase) a další.
-

- **Klientské aplikace**
 - nutno klást důraz na
 - vysokou flexibilitu;
 - poskytování na požádání nebo automaticky standardních reportů;
 - možnost snadné definice nových specifických reportů;
 - intuitivnost ovládání a obecně uživatelskou přívětivost;
 - možnost provozování aplikace v různých technologických prostředích;
 - požadavek na kvalitní dokumentaci popisující ovládání aplikace včetně možné interpretace vybraných analýz.



- 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://datawarehouse4u.info/Data-warehouse-schema-architecture-fact-constellation-schema.html>



Děkuji za pozornost

Otázky?
