



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 4



**SLEZSKÁ
UNIVERZITA**

**OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ**

doc. Mgr. Petr Suchánek, Ph.D.

- Provozní (transakční) systémy
 - slouží k automatizaci provozních činností a obchodních procesů;
 - umožňují
 - umožňují evidovat zákazníky;
 - vystavovat objednávky;
 - rezervovat místenky či zboží ve skladě;
 - plánovat výrobu;
 - kontrolovat plnění zakázek;
 - evidovat transakce na účtu zákazníka;
 - vystavovat faktury;
 - párovat došlé platby od zákazníků;
 - převádět data do účetnictví.
-

- Provozní (transakční) systémy
 - Problémy například s
 - porovnat prodeje jednotlivých výrobků v různých regionech za posledních 12 měsíců;
 - předpovědět počet zakázek na příští dva měsíce;
 - posoudit trend v příjmech s eliminací sezónních výkyvů;
 - seskupit zákazníky do segmentů se společnou charakteristikou;
 - zjistit jací zákazníci mají tendenci přejít ke konkurenci atp.

- Provozní (transakční) systémy – proč není možné řešit úkoly uvedené na snímku č. 3?
 - Nedostatečná historie dat
 - provozní systémy udržují z kapacitních důvodů data stará pouze několik měsíců, starší data se přehrávají do archívu, kde mají velice omezenou využitelnost.
 - Nedostatečné techniky a nástroje pro zpracování dat
 - provozní systémy většinou disponují sadou připravených reportů, které jsou navíc orientovány spíše na sledování procesů a jednotlivých transakcí než na globální pohled.
 - Nemožnost zpracovat data z jiných aplikací
 - v žádné společnosti není používán pouze jeden systém byť existují velmi mohutná řešení jako je např. SAP R/3, řada dat navíc vzniká mimo provozní systémy třeba i jako původně pouze osobní aplikace nebo jsou z jiných zdrojů.
-

- Provozní (transakční) systémy – proč není možné řešit úkoly uvedené na snímku č. 3?
 - Nepříznivý dopad na výkonnost provozních systémů
 - zpracování dat v provozních systémech k souhrnným přehledům a analýzám představuje další zátěž a vede k prodlužování doby odezvy transakčních systémů pro běžné uživatele.
 - Výše uvedené důvody vedou k zásadnímu konceptu datových skladů
 - požadavky na provozní systémy a na zpracování dat pro rozhodování jsou natolik rozdílné, že vyžadují dva druhy systémů - provozní systémy a datový sklad;
 - datový sklad je fyzicky a logicky oddělen od provozních systémů;
 - data z provozních systémů se převádějí do datového skladu, kde se po transformaci ukládají způsobem, který vyhovuje analytickému a prezentačnímu zpracování výstupů.
-

- Datový sklad
 - Datový sklad představuje uložení dat které má následující charakteristiky
 - integruje data z různých zdrojů do jednoho systému;
 - obsahuje historii - jsou k dispozici data i za několik minulých let;
 - data jsou uložena na různých úrovních sumarizace;
 - data se periodicky načítají z provozních systémů (většinou v noci a o víkendech);
 - uživatelé data pouze čtou, tj. neprovádí jejich zadávání ani je nemění;
 - data uspořádána podle jednotlivých subjektů;
 - data z datového skladu se využívají pomocí širokého spektra metod pro prezentace a analýzy dat.
-

- Provozní systémy - koncepce
 - dostat data do systému;
 - uživatelé mají možnost zadávat data, měnit data, rušit data a číst data;
 - zajišťují automatizaci rutinních činností;
 - aplikace jsou v podstatě statické (požadavky na funkčnost aplikace jsou poměrně stálé);
 - podporují každodenní firemní aktivity;
 - orientované na výkonnost;
 - proces implementace a využívání je poháněn technologií (tj. impulsem k inovaci systému je nové systémové prostředí, nová verze databáze atp.).
-

- Datový sklad - koncepce
 - dostat informace ze systému;
 - uživatelé mají možnost pouze číst data;
 - umožňují kreativitu uživatelů při práci s daty (analýzy, prezentace);
 - aplikace jsou dynamické (požadavky na funkčnost aplikací se mění);
 - podporují dlouhodobé strategie firmy;
 - poskytují konkurenční výhodu;
 - proces implementace a využívání je poháněn potřebami organizace (tj. impulsem k inovaci systému jsou nové potřeby uživatelů).

- Provozní systémy - technologie
 - zpracovávají velké objemy malých transakcí;
 - transakce neustále přidávají a aktualizují data;
 - důležitým hlediskem je omezení redundance dat;
 - integrita dat se zajišťuje datovým modelem a aplikacemi;
 - datové modely jsou optimalizované pro online aktualizace a rychlé zpracování transakcí;
 - používají se převážně normalizované relační datové modely;
 - zpracovávají malý počet komplexních dotazů.

- Datový sklad - technologie
 - data se načítají dávkově;
 - důležitým hlediskem je rychlý přístup datům pro účely analýz a prezentací;
 - integrita dat se zajišťuje při dávkových načítacích procesech (transformace dat);
 - datové modely jsou optimalizované pro rychlé zpracování výstupů;
 - používá se kombinace datových modelů (normalizované a denormalizované relační modely, sumarizované tabulky, star schéma datové modely, snow flake datové modely, fact constellation, multidimenzionální datové modely).
-

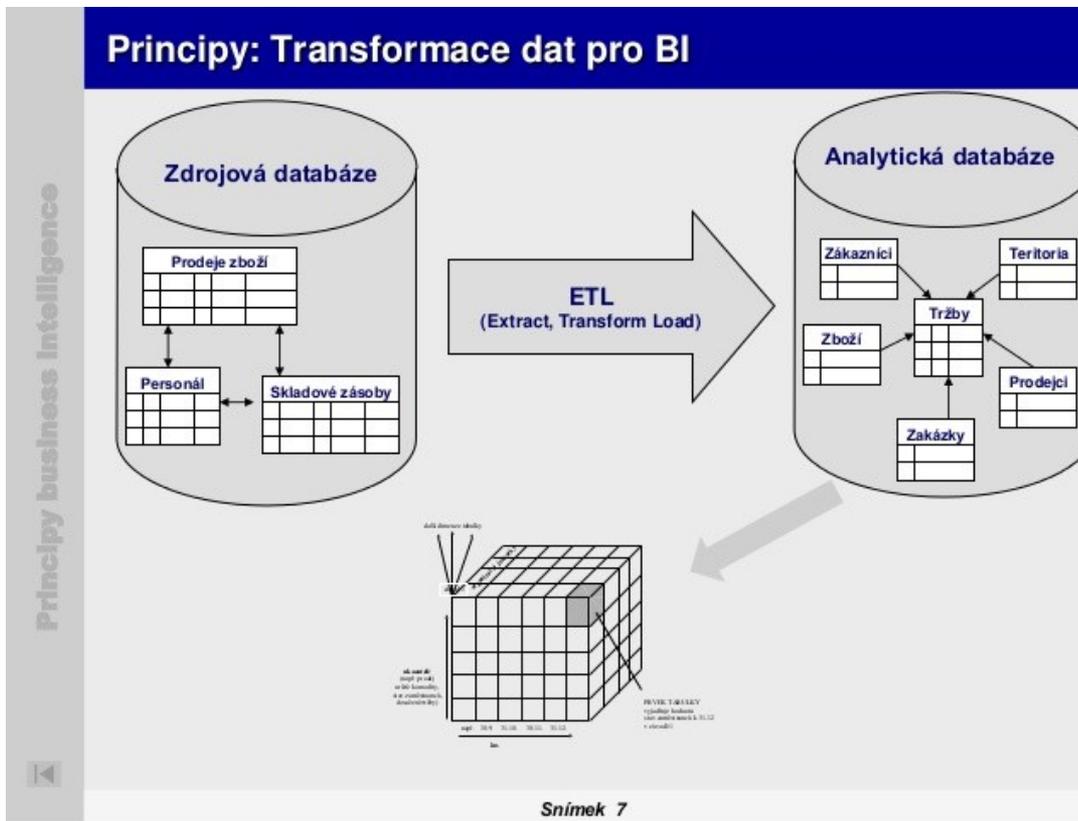
- Datový sklad – načítání dat
 - Do datového skladu se data nezadávají, ale načítají se z provozních systémů.
 - Načítání se většinou provádí v čase, kdy nejsou provozní systémy příliš zatíženy, aby se neprodložovala doby odezvy pro uživatele těchto systémů.
 - Při plnění datového skladu je nutné realizovat tyto hlavní kroky:
 - extrakci vstupních dat;
 - transformaci vstupních dat;
 - načtení dat do DW.

- Datový sklad – extrakce vstupních dat
 - Nástroje datového skladu musí umožňovat extrakci dat z provozních systémů, což ve většině případů znamená komunikovat určitým způsobem (ODBC, nativní drivery, textové soubory) s relační nebo síťovou databází či případně systémem souborů.

- Datový sklad – transformace vstupních dat
 - Data v datovém skladu jsou uložena jiným způsobem než ve zdrojových systémech;
 - Provozní systémy používají v naprosté většině normalizovaný entito-relační datový model;
 - U datového skladu se naproti tomu používá kombinace několika datových modelů (schéma hvězda, schéma sněhová vločka, normalizovaný entito-relační model, denormalizovaný entito-relační model, multidimenzionální datový model).

- Datový sklad – transformace vstupních dat
 - Transformace dat se skládá z těchto dílčích operací
 - **validace** - ověření správnosti dat;
 - **čištění** - odstranění či změna nesprávných dat;
 - **integrace** - dosažení konzistence dat pocházejících z různých systémů (datové typy, formáty...);
 - **derivace** - vytvoření derivovaných dat na základě vstupních dat;
 - **denormalizace** - snížení potřeby spojování tabulek při využívání datového skladu;
 - **sumarizace**- vytvoření požadovaných souhrnů z detailních dat.

Komponenty BI – datový sklad



- Datový sklad – transformace vstupních dat
 - Prostředí DS musí poskytovat dostatek nástrojů a metod pro zvládnutí všech těchto kroků, vzhledem k povaze těchto procesů nepostačuje pouze SQL jazyk, ale je nutné používat specializované prostředky;
 - Velmi důležitou fází je validace a čištění, protože základní podmínkou využitelnosti DS je to, že obsahuje důvěryhodná a správná data;
 - Jakkoliv sofistikované metody na využití dat jsou v podstatě bezcenné, pokud nepracují se správnými daty.

- Datový sklad – uložení dat
 - Data je možné ukládat na základě různých strategií
 - pokaždé se uloží celý obsah DS znovu (použitelné pouze u velmi malých objemů dat či pro úvodní načtení);
 - ukládají se pouze přírůstky a změněná data (v tomto případě musí být k dispozici systém zajišťující rozpoznání změněných údajů).
- Datový sklad – architektura
 - Prosadily se dva základní koncepty datového skladu
 - nezávislé datamarty (virtuální datový sklad);
 - integrovaný datový sklad.

- Datový sklad – nezávislé datamarty
 - Při této koncepci se pro DS řeší potřeby jednotlivých útvarů či aplikací víceméně odděleně a vytváří se samostatná datová úložiště tzv. datamarty, která se někdy označují jako útvarové datové sklady.
 - Výhody tohoto uspořádání jsou následující
 - snazší a rychlejší implementace;
 - rychlejší přínosy pro uživatele.
 - Nevýhody
 - může docházet k nekonzistencím mezi jednotlivými datamarty;
 - komplikované načítací procesy (velký počet, náročné na údržbu).
 - S rostoucí velikostí datového skladu převažují nevýhody nad výhodami a proto byl vytvořen druhý přístup.
-

- Datový sklad – integrovaný datový sklad
 - Při této koncepci se data z provozních systémů ukládají do centrálního datového úložiště, ze kterého se následně odvozují datamarty pro potřeby jednotlivých útvarů či aplikací.
 - Výhody tohoto uspořádání jsou následující
 - konzistentní obsah datového skladu;
 - menší počet načítacích procesů z provozních systémů (primární načítací procesy);
 - jednodušší správa načítacích procesů;
 - snazší vytváření nových datamartů (detailní data jsou již k dispozici v DS).

- Datový sklad – integrovaný datový sklad
 - Nevýhody
 - složitější realizace;
 - pomalejší implementace (lze eliminovat vhodnou metodologií);
 - sekundární načítací procesy (z centrálního DS do datamartu).
 - Vzhledem k tomu, že požadavek na konzistentnost obsahu datového skladu je naprosto zásadní (DS musí poskytovat "jedinou verzi pravdy") tento přístup v současnosti převládá.

- Datový sklad – využití datového skladu
 - Operativní dotazy (tj., předem nepřipravené dotazy na určité hodnoty);
 - Sestavy (jak standardní generované dávkově, tak operativní vytvářené podle potřeby);
 - Multidimenzionální analýza (OLAP, tj. rychlé prohlížení dat sumarizovaných na různých úrovních z různých pohledů neboli dimenzí);
 - Statistické analýzy (např. zjišťování závislosti veličin, identifikace důležitých proměnných, vytváření segmentů);
 - Finanční analýzy (např. ekonometrické modelování, termínové modely);
 - Analýzy časových řad a tvorbu předpovědí (např. předpovědi budoucích hodnot a identifikace sezónních výkyvů).
-

- Datový sklad – využití datového skladu
 - Vizualizaci dat (prohlížení dat v dynamicky provázaných grafech pro např. identifikaci neobvyklých a extrémních hodnot a závislostí mezi daty);
 - Dolování dat (data mining, specializované techniky pro zpracování velkých objemů dat a hledání skrytých vzorů a souvislostí);
 - Geografické informační systémy (převádění hodnot proměnných na geografickou prezentaci - např. zbarvení okresů podle počtu zákazníků);
 - Manažerské informační systémy - EIS (připravené aplikace pro vedoucí pracovníky se snadným ovládním kombinující OLAP, reporting, přehledné zobrazení kritických veličin, jednoduché předpovědi);
 - Aplikace vytvořené podle konkrétních potřeb.
-

- Datový sklad – co je důležité při tvorbě DS
 - Rozšiřitelnost a škálovatelnost řešení;
 - I když DS bude zpočátku nevelký rozsahem a objemem dat, je nutné mít na paměti jeho budoucí růst a tedy nutné mít možnost přecházet na výkonnější platformy (bez nutnosti přepracovávat aplikace);
 - Dostupnost pro hardwarové a softwarové platformy;
 - Podpora pro architekturu klient/server a pro webové technologie;
 - Nástroje pro extrakci dat (databázové systémy, systémy souborů a datové soubory);
 - Nástroje pro transformace dat (validace, čištění, integrace, derivace, denormalizace, sumarizace).
-

- Datový sklad – co je důležité při tvorbě DS
 - Způsoby uložení dat v centrálním datovém skladu a v datamartech;
 - Otevřenost na vstupu - z jakých datových zdrojů (formátů) lze data načítat;
 - Otevřenost na výstupu - jak lze přistupovat k datovému skladu z externího prostředí;
 - Využití metadat;
 - Prostředky pro zajištění bezpečnosti dat;
 - Nástroj na správu datového skladu;
 - Nástroje na využití datového skladu;
 - Prostředí pro vytváření aplikací.
-

- Provozní systémy – způsob realizace
 - Lze využívat standardizovaná řešení předem vytvořená na základě funkční analýzy řešené problematiky;
 - Na základě analýzy funkčních požadavků je možné vytvořit:
 - celkový datový model ve fázi návrhu;
 - kompletní řešení ve formě standardizovaného produktu či systému.
 - Vstup dat do systému je plně pod kontrolou produktu (tj. není nutné načítat dat z jiných systémů, rozhodně ne ve velkém rozsahu);
 - Vytvořené řešení je možné následně implementovat u různých zákazníků převážně pomocí nastavování parametrů, naplnění číselníků a voleb u připravené aplikace.
-

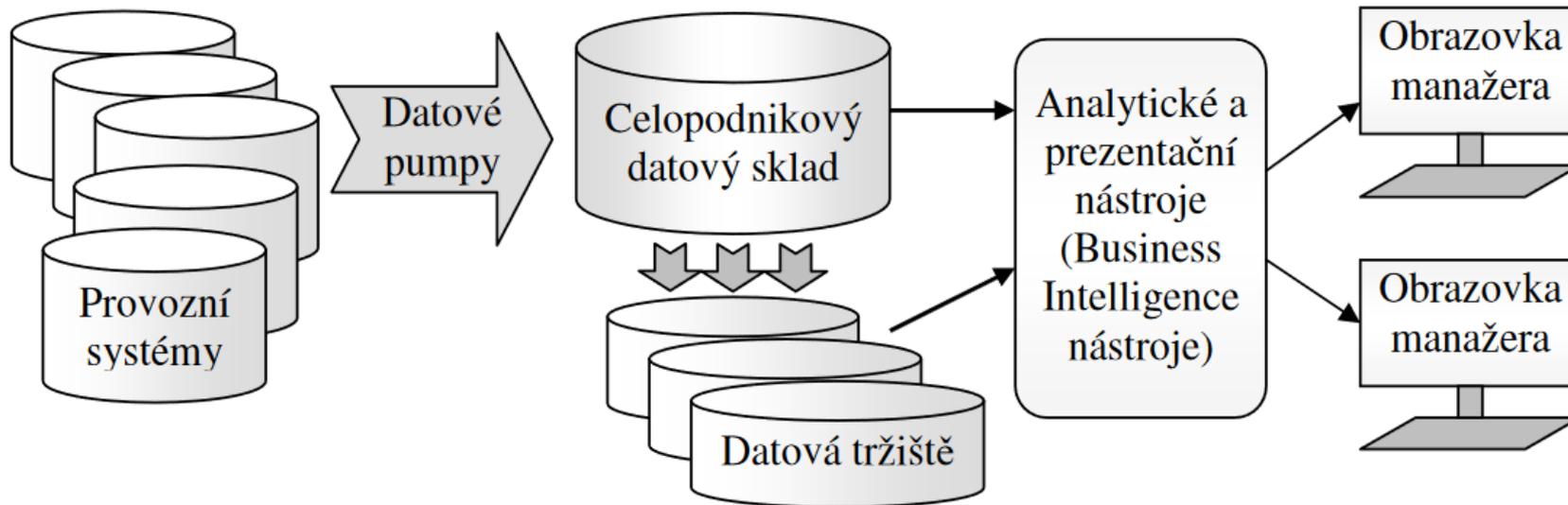
- Datový sklad – způsob realizace
 - Datový sklad je pro každou organizaci individuální, což znamená, že nelze mít připravenou aplikaci datový sklad, kterou implementuje u různých zákazníků tím, že nastavujeme její parametry;
 - Každá organizace má individuální kombinaci systémového prostředí (HW platformy, databázové systémy) a provozních systémů (které slouží jako zdroj dat pro DS);
 - Každá organizace má individuální požadavky manažerů a analytiků na analýzy, prezentace dat a aplikace (např. jiný styl řízení znamená jiné kritické veličiny, které je potřeba sledovat a analyzovat);

- Datový sklad – způsob realizace
 - Požadavky uživatelů na zpracování se neustále mění, neboť chtějí zpracovávat další data či kombinace dat, používat další metody atp. (typickým příkladem je marketingové oddělení, jehož náplní je neustálé hledání nových způsobů jak zpracovat data o zákaznících, produktech a službách a dozvědět se o nich něco nového);
 - DS nelze chápat jako jednu velkou aplikaci, ale jako proces poskytování dat, nástrojů na jejich zpracování a měnících se aplikací;
 - Realizace DS tady probíhá postupně, po jednotlivých etapách či projektech, kdy každá iterace zpřístupní určitou množinu dat, dodá sadu použitelných nástrojů a konkrétní aplikaci, kterou je nutné individuálně vytvořit;
-

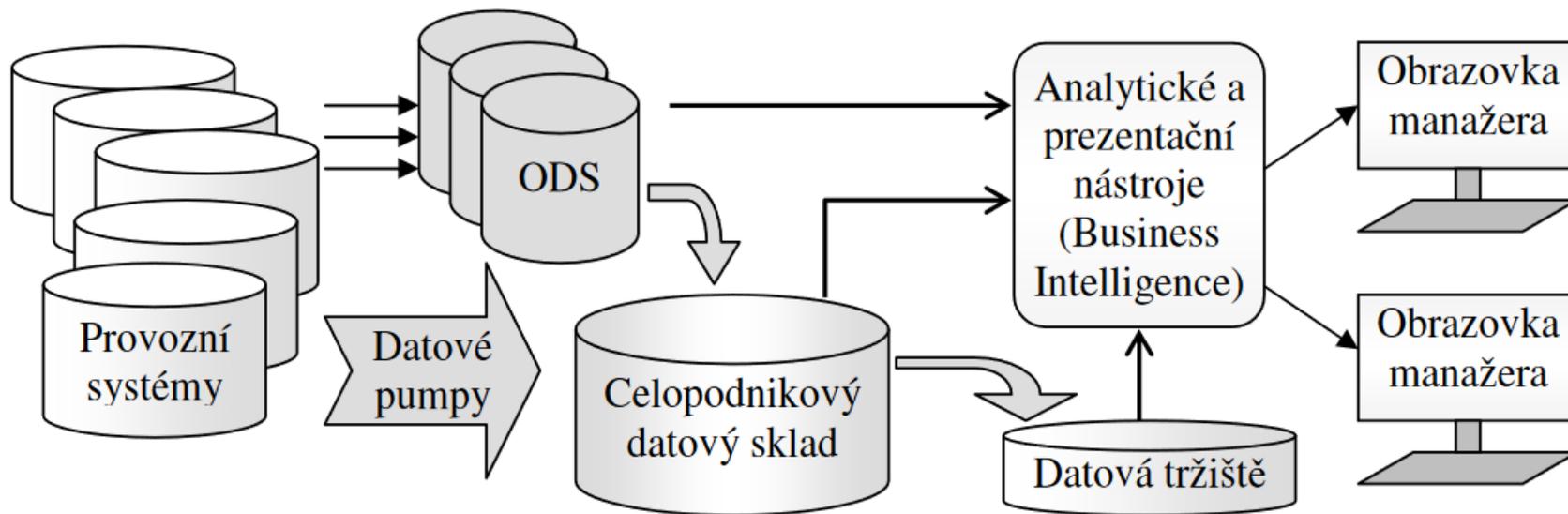
- Datový sklad – způsob realizace
 - Přestože se postupuje po etapách, tak uživatelé dostávají rychle celou škálu možností jak, uspokojující jejich potřeby na práci s daty. To je dáno především tím, že
 - nástroje pro práci s daty v DS jsou okamžitě použitelné a svou funkcí umožňují uživatelům provádět řadu věcí, pro které je nutné v transakčních systémech vytvářet aplikace;
 - vývojové prostředí pro aplikace DS umožňuje velmi rychlou realizaci.
 - Časově nenáročnější částí projektů DS je převedení dat z provozních systémů do DS (načtení, transformace, uložení), u většiny případů tato část vždy zabírá 70 – 80 % realizace projektu;

- Datový sklad – způsob realizace
 - Jakmile jsou data v DS, lze jednak používat nástroje a jednak velmi rychle vytvořit požadovanou aplikaci podle specifických potřeb organizace.

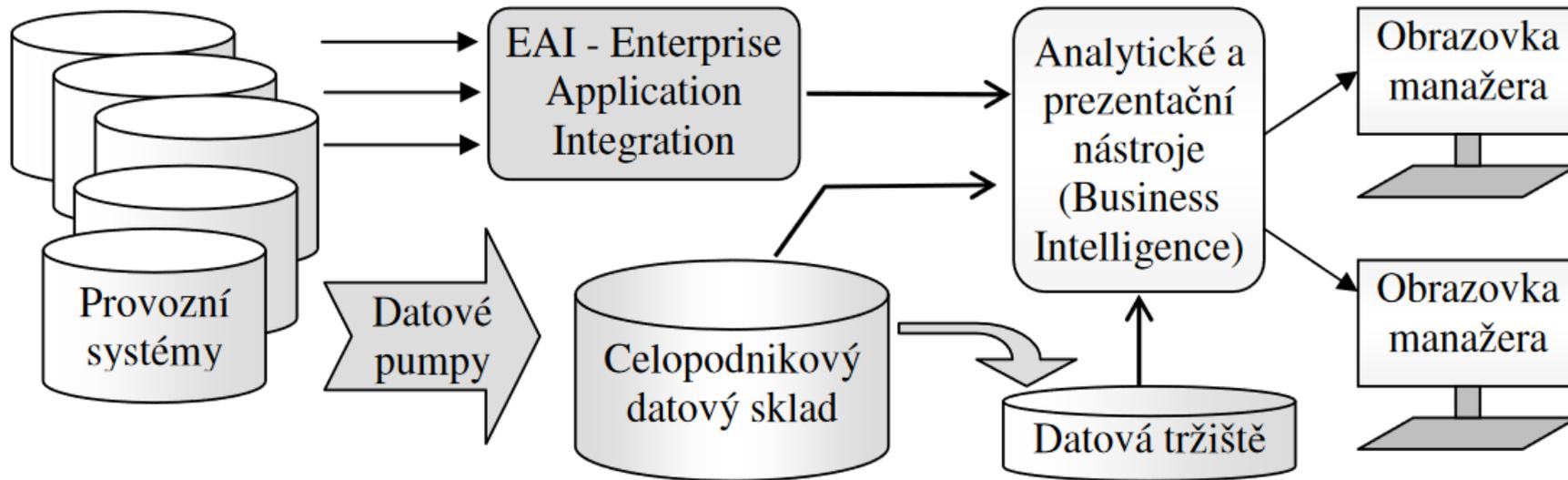
Komponenty BI – klasická koncepce DS



Komponenty BI – operativní datová úložiště



Komponenty BI – koncepce Active Data Warehouse



- 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.
- <https://www.systemonline.cz/clanky/hlavni-principy-datovych-skladu-a-proces-jejich-vytvareni.htm>
- <https://www.slideshare.net/OKsystem/bi-forum-2009-16335714>



Děkuji za pozornost

Otázky?
