

Databázové systémy a metody zpracování dat

2.přednáška

Principy Business Intelligence

- zpracování velkých objemů dat tak, aby výsledek této akce manažerům pomohl k rozhodování při řízení procesů
- výsledkem zpracování musí být relevantní informace, kterou dostanou manažeři ve správném čase
- základní zdroj dat, která se často ukládají do datových skladů jsou ERP systémy (relační DB)
- získání informací jako výsledek strukturovaných dotazů musí probíhat rychle (krátká odezva)
- používá se pro řízení na strategické, taktické u operační úrovni

Principy Business Intelligence

- **Definice 1** : Business Intelligence je sběr a analýza dat, jejímž cílem je lepší porozumění a reakce na změny, kterým organizace neustále čelí
- **Definice 2** : Business Intelligence je znalost podniku získaná za použití HW a SW technologií, která umožní přeměnit data organizace v informaci
- **Definice 3** : sada procesů, aplikací a technologií, jejichž cílem je účinně a účelně podporovat rozhodovací procesy ve firmě. Tyto procesy podporují analytické a plánovací činnosti podniků a organizací a jsou postaveny na principech multidimenzionálních pohledů na podniková data

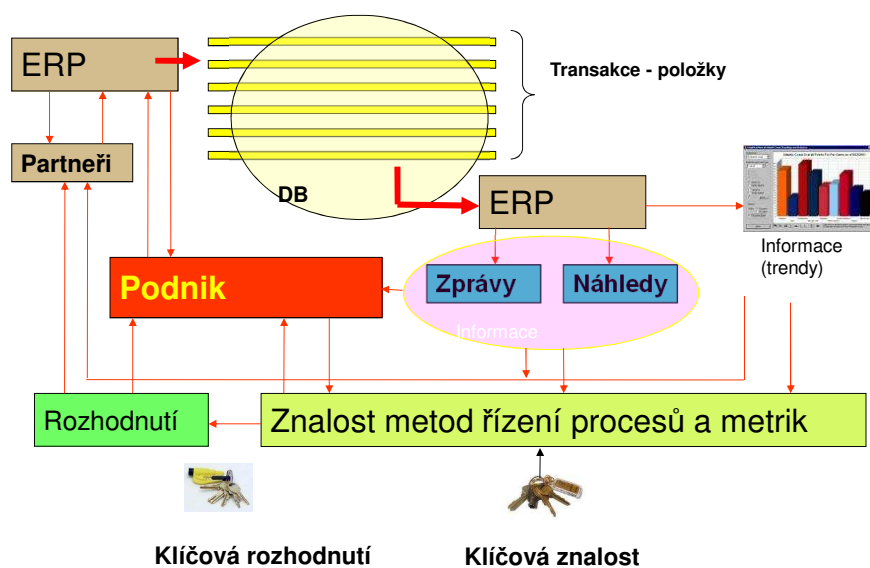
Nástroje Business Intelligence

- ERP systémy
- Dočasná úložiště (DSA: Data Staging Area)
- Operativní úložiště (ODS : Operational Data Store)
- Transformační nástroje (ETL : Extraction Transformation Loading)
- Integrační nástroje (EAI : Enterprise Application Integration)
- Datové sklady
- Datová tržiště
- OLAP
- Reportingové nástroje
- EIS (Executive Information Systém)
- Data Mining

Omezení ERP jako poskytovatele dat

- ERP jsou primárně určeny k pořizování dat a jejich aktualizaci
- Objem dat stále narůstá (za pět let se zdvojnásobí) - systém je zahlcen redundantními daty
- Problematický vícedimenzionální pohled na data - ERP databáze není pro tento pohled postavena.
 - ERP databáze neumožňují rychle a pružně měnit kritéria výběru
 - On-line přístup uživatelů k velkým objemům agregovaných dat
- Databáze, které vzniknou na základě primárních dat z ERP, jsou pro OLAP optimalizovány (*drilling a slice* operace)

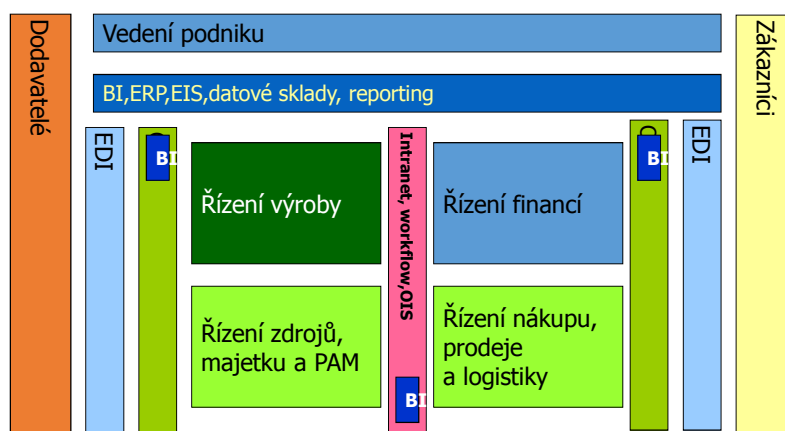
Zjednodušené schéma využívání ERP



Schematické a zjednodušené schéma BI

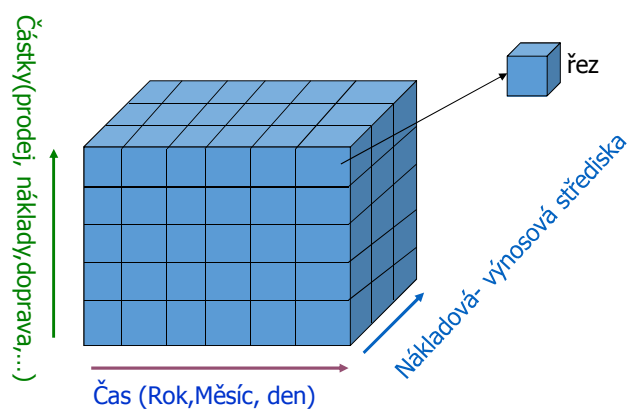


Organizačně-technologické schéma podniku

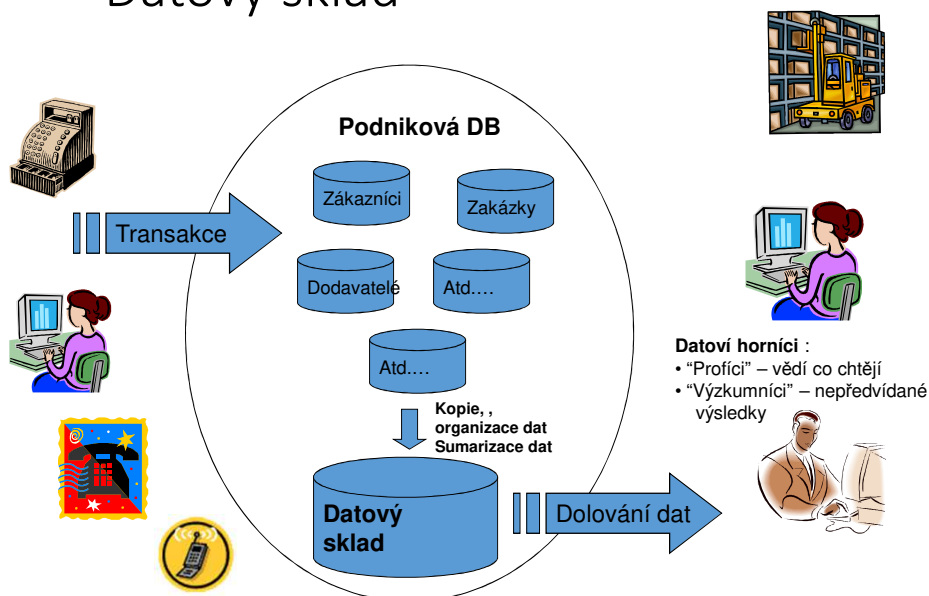


OLAP kostka

- http://www.databaseanswers.org/designing_olap_cubes.htm



Datový sklad



Definice

- Datový sklad: základní Business Intelligence
- Datové tržiště : subjektivě orientované analytické DB- součást datového skladu
- Operativní datová úložiště (ODS : Operational Data Store) : podpůrné analytické DB
- Dočasná úložiště dat (DSA: Data Staging Area) : úložiště dat před jejich zpracování do databázových komponent řešení BI

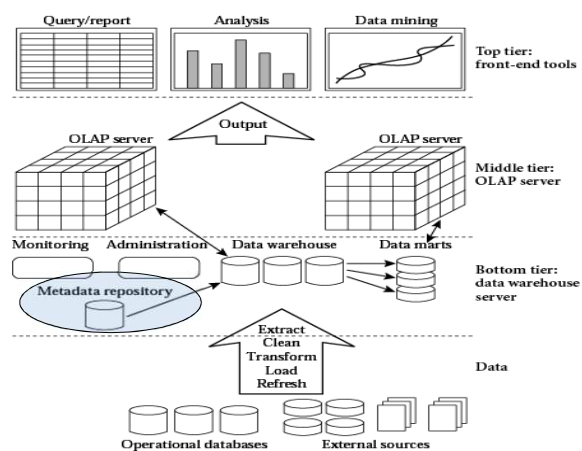
Vrstvy pro analýzu dat

- Reporting : ad hoc dotazovací proces do DB komponent BI
- OLAP : pokročilé a dynamické analytické úlohy
- Data Mining (dolování dat) : analýza většího množství dat za účelem získání nových poznatků

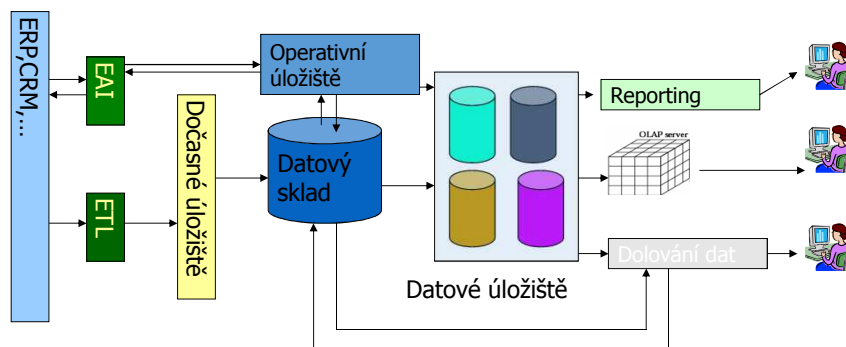
Metadata

Metadata - pomocí předem definovaných dat s jasně danou a popsanou strukturou uchováváme informace o jiných datech.

Architektura OLAP



Hlavní komponenty Business Intelligence a jejich vazby



Transformační nástroje (ETL : Extraction Transformation Loading)
Integrační nástroje (EAI : Enterprise Application Integration)

Datová pumpa (kritické místo celé aplikace)

Datová pumpa, nebo-li ETL nástroj umožňuje efektivní zpracování velkých objemů z různých zdrojů a jejich uložení do datového skladu. Každý ETL nástroj musí umět:

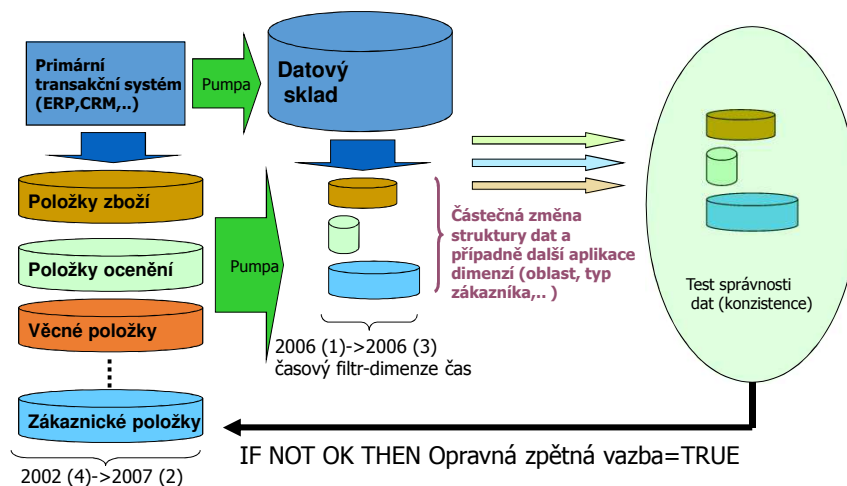
- a) zpracovávat různorodá data obvykle fyzicky umístěná na různých místech,
- b) navrhovat transformace pro přenos dat mezi různými datovými formáty



Zpracování - odstranění redundancí, agregace podle dimenzí, zapomínání dat
Zapomínání dat - úmyslné odstranění nepotřebných dat z datového skladu

Datová pumpa (kritické místo celé aplikace)

Datová pumpa = Extraction Transformation and Loading = ETL



Vytváření datového skladu

Data Warehouse (DW)

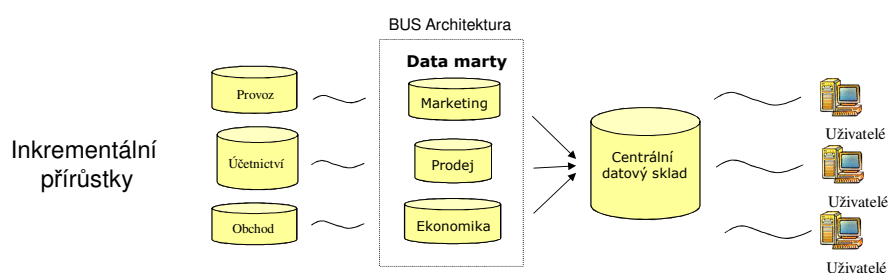
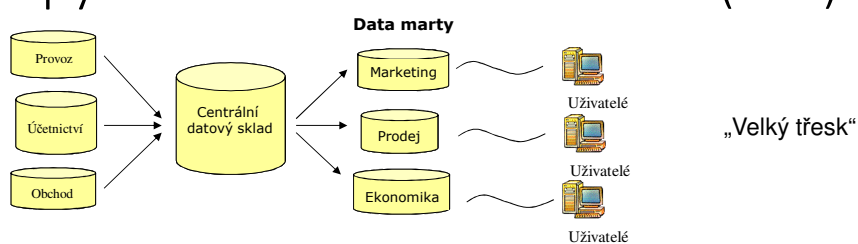
Základní způsoby vytváření DW

- Existují dva základní přístupy k vytváření DW:

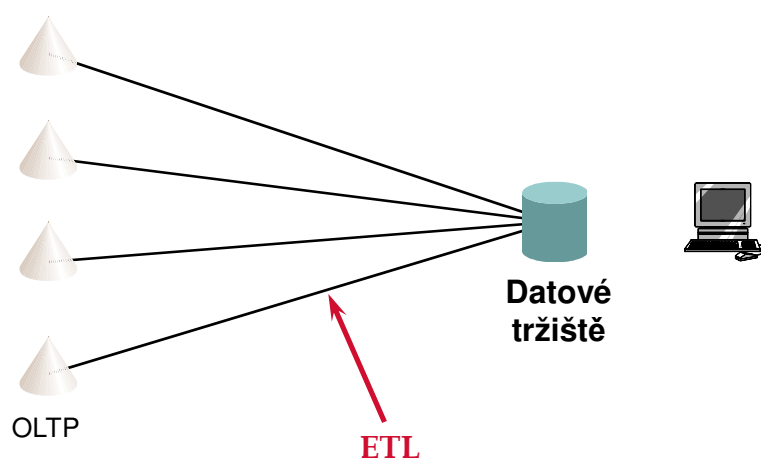
- Metoda velkého třesku

- Přírůstková metoda

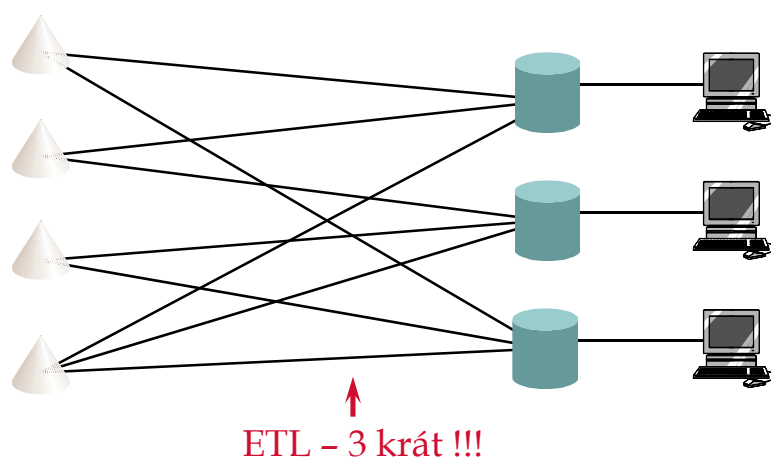
Přístupy k tvorbě datového skladu (DW)



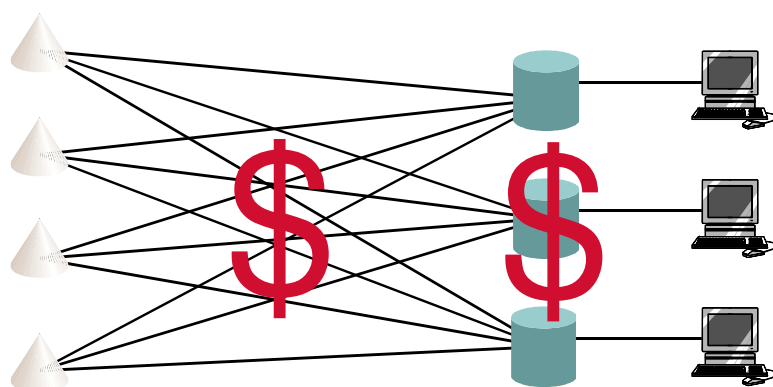
Nezávislé datové tržiště (DM)



Nezávislé datové tržiště



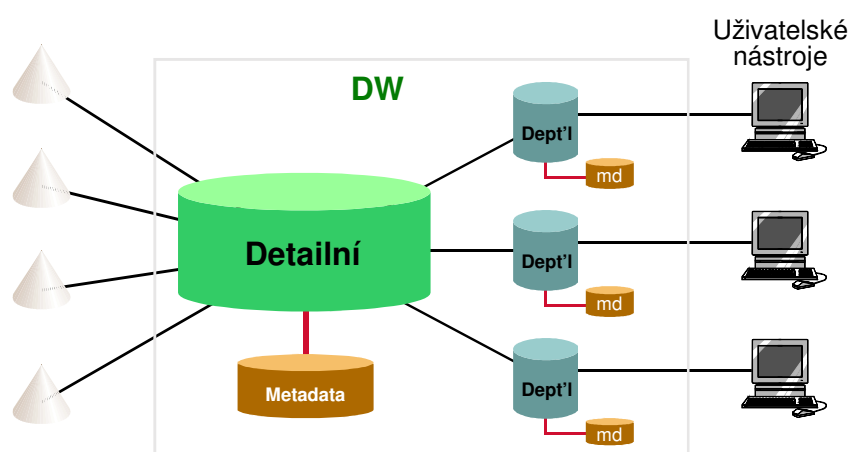
Nezávislé datové tržiště



Duplikace práce na ETL
Časově i finančně náročné

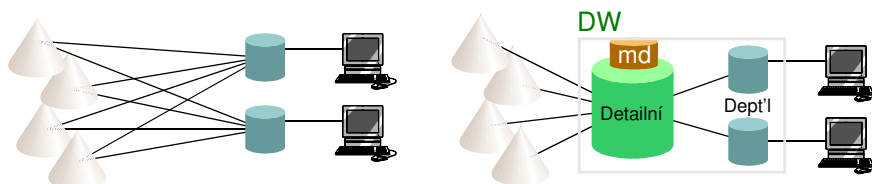
Údržba nezávislých DM je
náročná, těžkopádná

DW architektura



Závislé datové tržiště s odpovídajícími
metadata

DW architektura



- ✉ Rychlý vývoj, ale ...
- ? Budou ETL konzistentní?
- ? Jak je řízena redundance?

- ♦ Komplexní architektura
- ♦ Konzistence dat
- ♦ Detailní i agregovaná data
- ♦ Metadata jsou konzistentní

Základní způsob budování DW

- Aktuálně preferována přírůstková metoda
- Zaručuje:
 - Projektovou zvládnutelnost řešení
 - Reálné časové horizonty jednotlivých etap (2-4 měsíce)
 - Řešení aktuálních uživatelských (obchodních) potřeb
 - Zpětnou vazbu uživatelů

Cíle datového skladu

- Zajistit dostupnost firemních informací
- Zajistit konzistenci firemních informací
- Vytvořit adaptivní a pružný zdroj informací
- Zabezpečit ochranu firemních informací
- Vytvořit základnu pro firemní podporu rozhodování (analytické centrum)

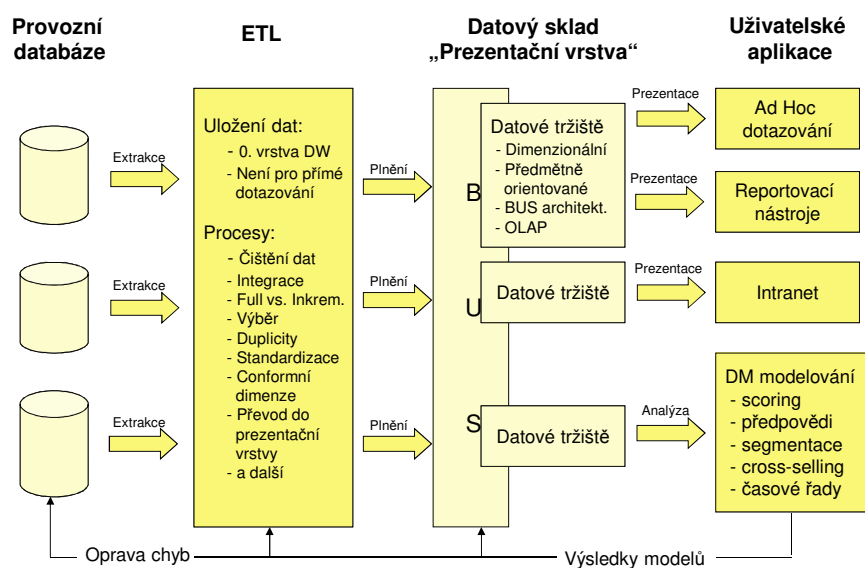
DW procesy

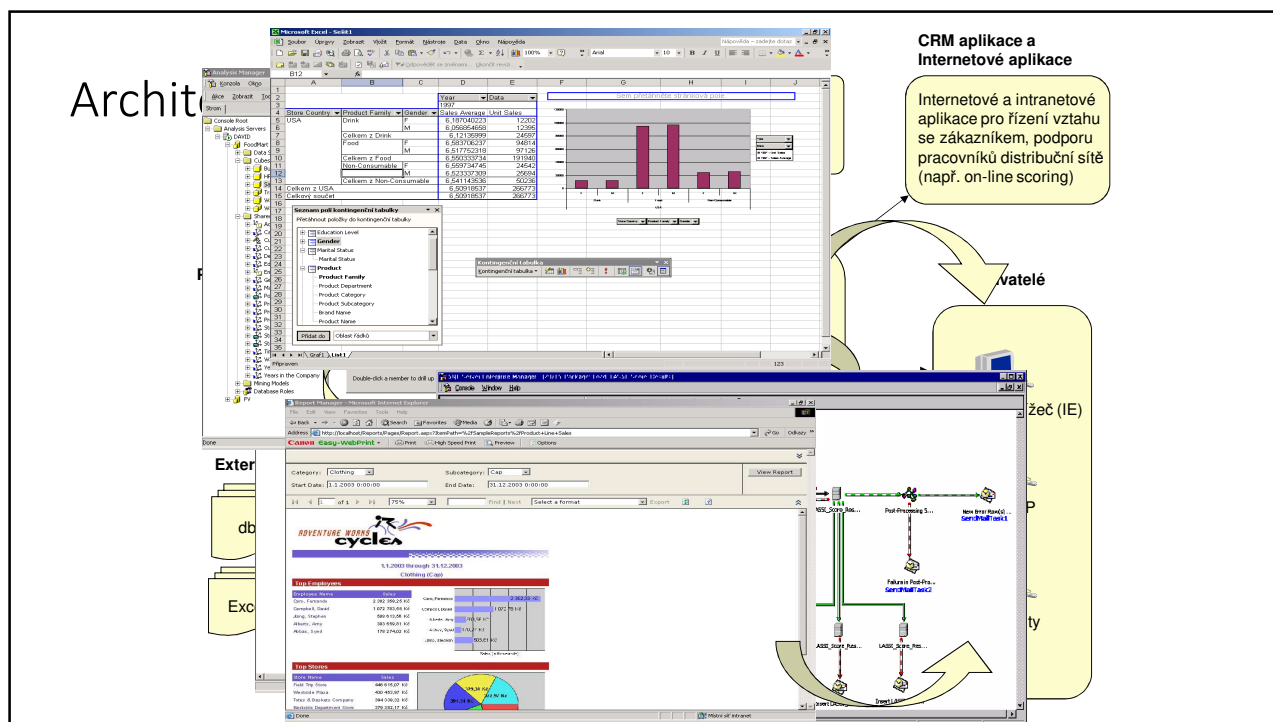
- Hlavní proces při tvorbě datového skladu
- Podprocesy:
 - Extrakce
 - Transformace
 - Čištění dat
 - Výběr dat
 - Integrace
 - Umělé klíče
 - Agregace
 - Načtení (Loading) a tvorba indexů
 - Data Quality Assurance

DW procesy

- Další procesy v DW
 - Publikace dat (prezentační server)
 - Update dat
 - Dotazování
 - Zpětná vazba (čistá dat do OLTP, data z DM do DW)
 - Audit dat
 - Bezpečnost
 - Zálohování a obnova

Architektura



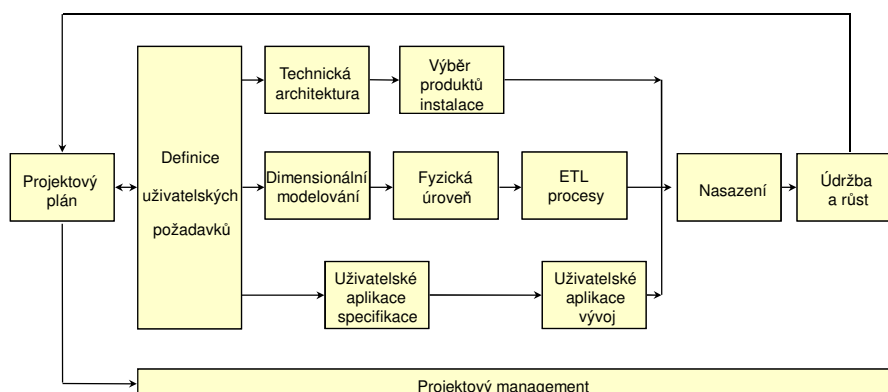


Produkty pro BI

- **Nástroje pro BI**
 - Microsoft (MS SQL Server, Analysis Services, Reporting Services)
 - Oracle
 - Sybase IQ
 - IBM DB2, DB2 OLAP Server
 - Microstrategy
 - SPSS
 - SAS
 - SAP – Business Warehouse
- **Klientské nástroje**
 - ProClarity
 - Oracle Discoverer
 - MS Excel – Power BI
 - Business Objects
 - Cognos: PowerPlay, Impromptu
 - Brio: Brio Query
 - Quadbase – EspressoReport

Business Dimensional LifeCycle

- Standardní metodologie vývoje datového skladu od Ralpha Kimballa



Popis metodologie

- Standardní metodologie tvorby datového skladu
- Modifikovaná metodika tvorby IS
 - Vyžaduje speciální znalosti analytiků/implementátorů
- Jednotlivé fáze mají rozdílnou váhu v jednotlivých etapách
 - 1. etapa – nastartování tvorby datového skladu (HW, SW infrastruktura)
 - Další etapy – dopady řešení na stávající infrastrukturu

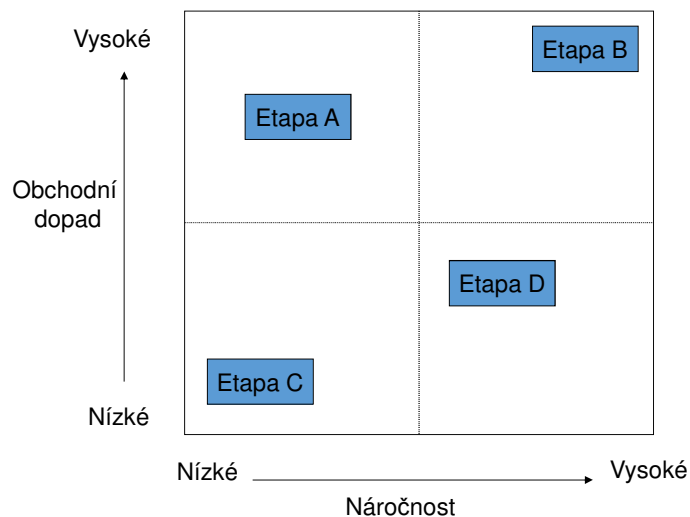
Plán projektu a projektový managm.

- Existuje poptávka po DW, od koho, proč?
 - Poptávka jediného oddělení
 - Informatika
 - Mnoho oddělení
- Ohodnotit připravenost pro projekt DW
 - Silný business sponzor
 - Pocit potřeby podpory businessu
 - Stupeň práce s informacemi dnes, ochota do budoucnosti („analytická kultura“)
 - Stav IS/IT
 - Proveditelnost (např. Existují zdroje dat?)
- Kritický faktor úspěchu podpora managementu

Plán projektu a projektový managm.

- Jak odstranit nepřipravenost?
 - Popsat hlavní potřeby business na konceptuální úrovni
 - Potřeby managementu
 - Prioritizace business potřeb
 - Proof of concept
 - Pozor potom na přehnaná očekávání

Plán projektu a projektový managm.



Plán projektu a projektový managm.

- Definice rozsahu projektu – 1. etapy
 - Řízeno obchodními potřebami ne harmonogramem
 - Spolupráce IT a business
 - Doporučeno jednoduchý obchodní problém řešitelný z jednoho zdroje dat
 - Limit na počet uživatelů (do 25)
 - Určení kriteria úspěchu realizace
- Největší riziko: neporozumění kvalitě a problémům v datech