



Data Warehouse

Jakub Božoň, Ján Šimal'

Čo je dátový sklad?

- Dátový sklad (Data Warehouse) je súbor technológií na podporu rozhodovania, ktorý umožňuje robiť lepšie a rýchlejšie rozhodnutia.
- Dátový sklad je nástroj pre transformovanie veľkého množstva dát do použiteľných a spoľahlivých informácií

Úloha D.W.

- Podpora rozhodovania na všetkých úrovniach riadenia
- Zvýšenie výkonu spracovania dát (pre podporu rozhodovania)
- Vytvorenie jednotnej, homogénnej dátovej základne (firma)

Definícia

- W. H. Inmon definoval D.W. nasledovne:

Údajový sklad je **kolekcia zjednotených, predmetovo orientovaných databáz** navrhnutých za účelom poskytovať **informácie** požadované pre **rozhodovanie**.

Kolekcia a zjednotený

- Údajový sklad obsahuje údaje **získané z mnohých systémov** a zároveň môže byť plnený tiež **externými údajmi**.
- Údajový sklad **spája údaje z mnohým rôznorodých systémov** a poskytuje integrovaný pohľad na subjekt a na plnú šírku jeho vzťahov s okolím, ktoré zachytávame.

Predmetovo orientovaný

- Údajový sklad priniesol tradičné informačné pohľady na **celopodnikové subjekty**, ako sú zákazníci, predaje či zisky. Tieto subjekty ohraničujú ako **hranice organizačné**, tak **procesné** a pre poskytnutie kompletného obrazu vyžadujú informácie z viacerých zdrojov.

Databáza

- Jadrom D.W. je rozsiahla databáza,
- Integrované podnikové údaje
- Dáta získavajú z interných a externých zdrojov

Využitie

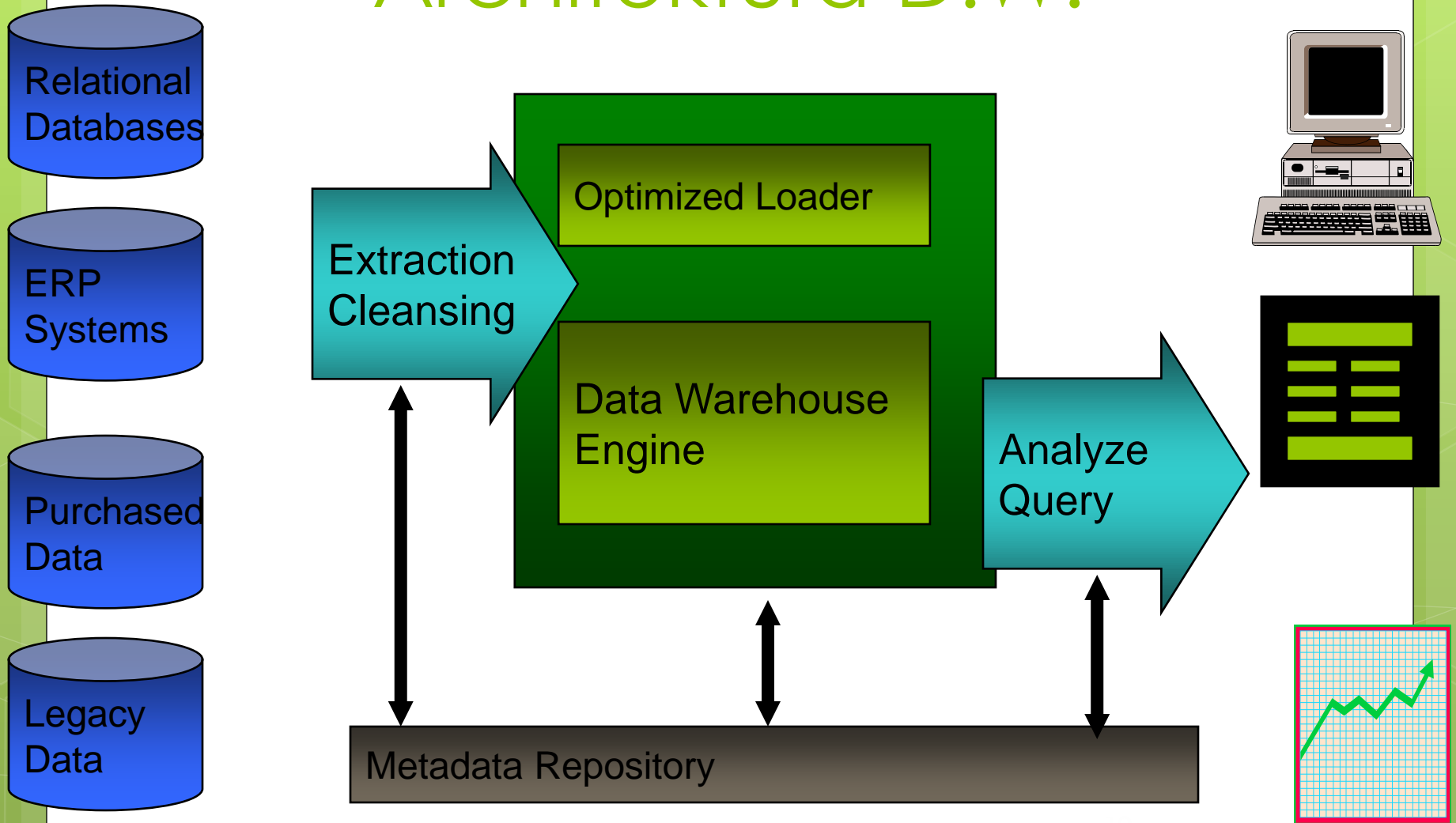
- **finančné služby**
- **priemyselná výroba**
- **Maloobchod**
- **zdravotníctvo**

Vytváranie dátových skladov

Proces premeny základných dát na warehouse zahrnuje niekoľko krokov:

- **zber**
- **konsolidovanie**
- **Filtrovanie**
- **Čistenie**
- **konverzia**
- **agregácia**

Architektúra D.W.



Architektúra dátových skladov

- Nezávislé Data-marty
- Prepojené Data-marty
- Integrovaný dátový sklad

Nezávislé Data-marty

- Budované pre konkrétnu pracovnú skupinu, alebo oddelením, ktorým sú aj riadené
- Rýchla implementácia
- Nie je možné dosiahnuť globálny pohľad na dáta

Prepojené Data-marty

- Zvyčajne distribuovaná implementácia
- Poskytuje určitý (naplánovaný) stupeň integrácie, na najvyššom stupni integrácie sa stáva integrovaným globálnym DW
- Zo zvyšovaním integrácie nárast nákladov

Globálny integrovaný DW

- Podporuje celú organizáciu
- Široká použiteľnosť
- Prístupný v rámci celej organizácie
- Umožňuje celopodnikový pohľad na dáta
- Návrh finančne a časovo veľmi náročný

Implementácia

- Implementácia zhora-nadol
- Implementácia zdola-nahor
- Kombinovaná implementácia

Implementácia zhora-nadol

- Väčšina plánovania a návrhu pred začatím projektu
- Zvyčajne sa používa pri globálnom integrovanom DW
- Výhody: Konzistentnejšie definície dát, zavedenie presných pravidiel
- Nevýhody: Cena, časové nároky

Implementácia zdola-nahor

- Nečaká sa na vytvorenie globálnej infraštruktúry
- Najprv sa implementujú data marty, z nich sa odvodzuje globálny DW
- Nevýhody: Redundantnosť a nekonzistentnosť viacerých datamartov, problematická integrácia do globálneho prostredia
- Výhody: Cena, rýchla implementácia

Kombinovaná implementácia

- Kombinácia predchádzajúcich prístupov
- Data marty budované zdola nahor
- Hlavná úloha je odhadnúť stupeň plánovania

OLTP VS D.W.

OLTP

- Aplikačne orientovaný
- Používa sa na prevádzku
- Detailne dáta
- Aktuálne dáta
- Izolované dáta
- Opakovaný prístup
- Administrátor

D.W.

- Predmetne orientovaný
- Používa sa na analýzu
- Sumarizácia dát
- Obrazové dáta
- Integrované dáta
- Ad-hoc prístup
- Manager

OLTP VS D.W.

- OLTP slúži na „**beh**“ biznisu
- Data Warehouse slúži na „**optimalizáciu**“ biznisu

Ďakujeme za pozornost'