An Overview of Data Warehousing and OLAP

Chaudhuri and Dayal paper
Elmasri and Navathe Ch 26 (3rd Ed)

Outline

- Data warehousing concepts
- Architecture
- Back-end functions

Terminology

- On-Line Transaction Processing (OLTP)
- Decision Support System (DSS)
- Data Warehouse (DW)
- On-Line Analytical Processing (OLAP)
  - Relational OLAP (ROLAP)
  - Multidimensional OLAP (MOLAP)
  - Multidimensional DB (MDDB)

What is the Problem?

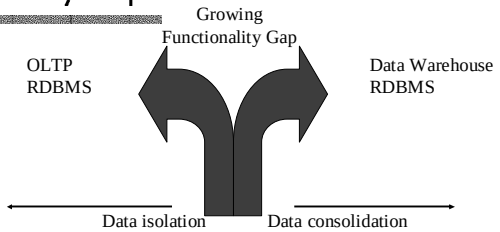
- Most enterprises now suffer from an overabundance of data
- Delivering business information becoming more complex and time consuming
- Cannot use operational data
- Data warehouse provides separate environment for extracting business info

What is Data Warehousing?

Data warehousing is a collection of decision support technologies aimed at enabling the knowledge worker to make better and faster decisions.

– Chaudhuri and Dayal, SIGMOD Record, March 1997

Why Separate OLTP and DW?



Data isolation
Limited historical content
Complex data structure
Transaction throughput

Data consolidation
Rich historical content
Comprehensible data structure
Query throughput

OLTP vs. Data Warehousing

	OLTP	Data Warehousing
Typical User	clerical	professional
Usage of System	run business	analyze business
User Interaction	pre-determined	ad-hoc
Unit of Work	transaction	query
Work Characteristics	read/write	read mostly
Records Accessed	tens	millions
Number of Users	thousands	hundreds
Focus	data in	info out

CISC 432/832 Data Warehousing Overview 7

Typical Queries

OLTP

Find all items on order for customer "Jones"

Data Warehouse

Over the past 3 years, what was the average weekly sales of HB pencils?

CISC 432/832 Data Warehousing Overview 8

Where is Data Warehousing Used?

Many Industries

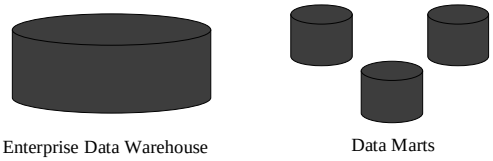
- Consumer Packaged Goods
- Retail
- Telecommunications
- Healthcare
- Transportation
- Financial Services
- Data Service Providers

Many Applications

- promotion analysis
- category management
- call/rate usage analysis
- claims analysis
- logistics management
- consumer credit analysis
- value-added data analysis

CISC 432/832 Data Warehousing Overview 9

DW Configurations

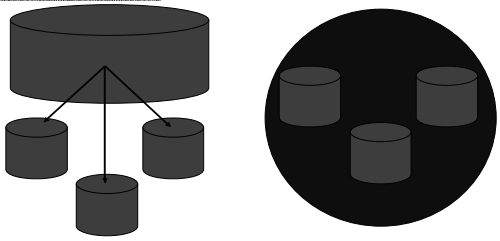


Enterprise Data Warehouse

Data Marts

CISC 432/832 Data Warehousing Overview 10

DW Configurations (cont.)

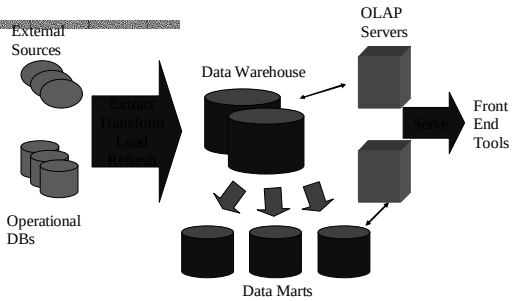


Multi-tiered Warehouse

Federated Warehouse

CISC 432/832 Data Warehousing Overview 11

Data Warehouse Architecture



CISC 432/832 Data Warehousing Overview 12

Back-end Functions

- **Data cleaning**
  - large volumes of data from multiple sources so likely to be errors and anomalies
  - 3 classes of tools
    - data migration
    - data scrubbing
    - data auditing

CISC 432/832 Data Warehousing Overview 13

Back-end Functions (cont.)

- **Load**
  - includes preprocessing like checking integrity constraints, sorting, computing aggregates / derived tables, building indices, partitioning into multiple target storage areas
  - volumes much larger than DBMS
  - efficiency key to loading within time window

CISC 432/832 Data Warehousing Overview 14

Back-end Functions (cont.)

- **Refresh**
  - propagate updates on source data to warehouse data
  - can be handled by replication servers
    - data shipping
    - transaction shipping
  - incremental update of derived tables is a research problem

CISC 432/832 Data Warehousing Overview 15