

What is A Data Warehouse?

- A data warehouse, is a decision support database that is maintained **separately** from an organization's operational database.
- Data warehouses support **information processing** by providing a solid platform of **consolidated, historical** data for analysis.



Formal Definition

- “A data warehouse is a **subject-oriented**, **integrated**, **time-variant**, and **nonvolatile** collection of data in support of management’s decision-making process.”
-- William H. Inmon (the father of data warehousing)



Subject-oriented

- Organized around **major subjects** rather than the organization's ongoing operations.
 - Product, courses,
 - Customers, students,
 - Suppliers, departments
 - Sales, revenue
- Focus on the modeling & analysis of data for **decision making**, not on daily operations or transaction processing.
- Provide a simple and concise view around particular subject issue by **excluding data that are not useful in the decision support process**.



Integrated

- Integrate multiple, heterogeneous data sources
 - Relational database, flat files, on-line transaction records
- Data cleaning and data integration techniques applied.
 - Ensure consistency in naming, encoding, attribute measures, etc
 - Hotel price, currency, tax, breakfast.



Time-Variant

- Significant longer time span
 - Operational database: current data, or recent data
 - Data warehouse: historical perspective
- Every key structure in a data warehouse
 - Contains time info, explicitly or implicitly.
 - Key of operational data may not contain time info.



Nonvolatile

- A physically separate store of data transformed from operational environments
- No operational update of data
 - No transaction processing, recovery, concurrency control
- Only two operations in data accessing
 - Initial loading of data
 - Access of data

