

Data Warehouse

D01. Data Warehouse



- Code: 164323-03
- Course: Information Policy
- Period: Spring 2013
- Professor: Sync Sangwon Lee, Ph. D

Contents

- 01. IT Environments
- 02. History of DW
- 03. New Information Systems
- 04. Data Warehouse



01. IT Environments

- DW(Data Warehouse)
 - A system supporting speedy & accurate decision making for elevating enterprise competitive power
 - Background of introduction
 - The needs of enterprises for analysis of large scale data
 - Speedy improvement of large scale hardware
 - Depreciation of large-scaled hardware
 - Concept change of computer science
 - A tool for competition → for survival
 - Cost center → profit center

3

01. IT Environments

- Needs of the Times for IT
 - Speedy and sensitive response against changes
 - Downsizing
 - Relational DB
 - Active application of new technologies
 - DW, simulation, modeling → analyzing market trend/tendency

4

01. IT Environments

- Needs of the Times for IT
 - Requisite implementation of CRM
 - Analysis on purchasing tendency
 - Segmentation of customers
 - Maintenance of existing customers
 - Acquisition of new customers
 - User-oriented IT environment
 - Users' direct data-manipulation without program
 - ISD-oriented → field-oriented
 - EDPS → MIS → KB

5

01. IT Environments

- Problems of Existing IT Environment
 - Difficult to analyze/forecast data fast
 - Difficult to make an integrated report
 - Long time to make unstructured documents

6

01. IT Environments

- Characteristics of Existing IT Environment
 - Batch transaction
 - Long time
 - Data deletion after batch transactions → Old data deletion
 - SAM(Sequential Access Method) file manipulation
 - SAM file = flat file, text file
 - There are no index but only data.

7

01. IT Environments

- Characteristics of Existing IT Environment
 - Integrated database
 - Problems of disintegration
 - Difficult to make integrated documents
 - Heterogeneous systems → difficult to manage by topics
 - Difficult to make real-time documents
 - Problems of integration
 - Difficult to access directly
 - Difficult to implement DB respectively for its use
 - Inconsistent correctness on data extraction methods.
 - Excessive screen design → long time for implementation
 - Frequent reprogramming for unstructured documents
 - Scant functions for analysis and forecasting

8

02. History of DW

- History of DW
 - Information Warehouse (1980s)
 - proposed by Bill Inmon, IBM
 - Actual implementation of DW (1980s)
 - Goals
 - Integration of distributed systems
 - Makeup of composite document for integrating distributed customer data
 - Target industry
 - Telecommunication, finance, logistics

9

02. History of DW

- History of DW
 - Many failures of initial DW
 - Scant OLAP tools
 - Limitation of H/W and S/W
 - Unskillfulness of large data
 - Downsizing → Transformed to C/S(Client/Server) environment
 - Low usage of OLAP(On-Line Analytical Process)

10

02. History of DW

- History of DW
 - The advent of multi-dimensional modeling
 - By Ralph Kimball
 - Introduction of ROLAP tool
 - Explosion of DW markets in America (after the middle 1990s)
 - Support of DW by DBMS
 - The advent of various OLAP tools
 - Depreciation of HW
 - Improvement of HW

11

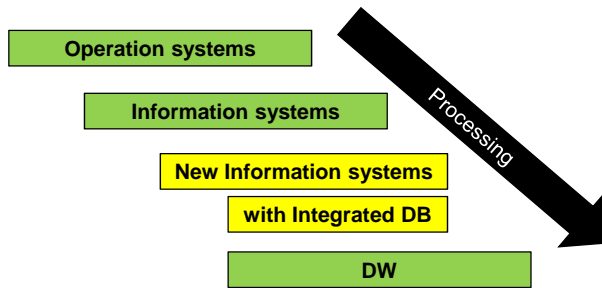
02. History of DW

- History of DW (in Korea)
 - Classification: operation system and information system
 - Trend: operation system → information system
 - Focused on formalized documents rather than various analyses
 - Differences b/t information system and DW
 - Take long to recognize the concept differences
 - Information system
 - Reprogramming whenever user's requests
 - DW
 - Accepting user's requests as they are
 - Modifying multi-dimensional modeling at the worst
 - More developed than information scope

12

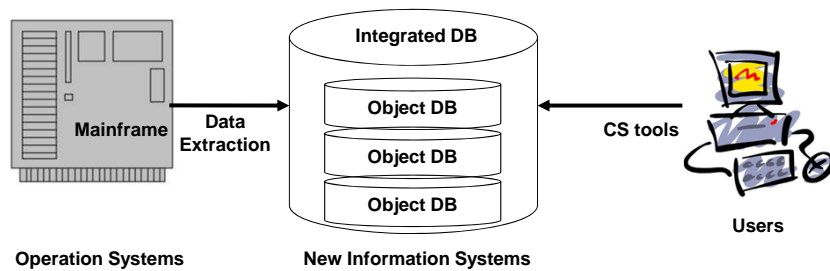
03. New Information Systems

- New Information Systems
 - Information systems with Integrated DB
 - Overcoming problems of traditional information systems
 - Problems of traditional information systems
 - Difficult to make integrated documents
 - Difficult to manage by topics
 - Difficult to make real-time documents



03. New Information Systems

- Implementation of New Information Systems



03. New Information Systems

- Characteristics of New Information Systems
 - Data extraction
 - Checking and loading log files periodically
 - Integrated DB
 - Storing all the data of operation systems
 - Difficult to access to enormous volume data
 - → Implement object DB
 - Users
 - Accessing to object DB with various programming tools

15

03. New Information Systems

- Problems of New Information Systems
 - A method of data extraction affects the degree of correctness.
 - Extravagant cost of development
 - Additional programming required for unstructured documents
 - Scant analysis and forecasting

16

03. New Information Systems

- Solutions for New Information Systems
 - An alternative → DW
 - Comparatively speedy implementation
 - Comparatively low cost
 - Easy to make up unstructured documents
 - Speedy response to user's changeful requirements

17

04. Data Warehouse

- Definition of DW
 - Integrated systems with online multi-dimensional analysis without reprogramming by integrating historical data (internal data and external data) by the subject

18

04. Data Warehouse

- Definition of DW
 - Integrated systems with online multi-dimensional analysis without reprogramming by integrating **historical** data (internal data and external data) by the subject
 - → Generally 3 to 5 years

19

04. Data Warehouse

- Definition of DW
 - Integrated systems with online multi-dimensional analysis without reprogramming by integrating historical data (**internal data** and external data) by the subject
 - → Acquired from operation systems (sales, finance, human resources management, ...)

20

04. Data Warehouse

- Definition of DW
 - Integrated systems with online multi-dimensional analysis without reprogramming by integrating historical data (internal data and external data) by the subject
 - → Acquired from secondary data

04. Data Warehouse

- Definition of DW
 - Integrated systems with online multi-dimensional analysis without reprogramming by integrating historical data (internal data and external data) by the subject
 - → Integrating data by topics

04. Data Warehouse

- Definition of DW
 - Integrated systems with online multi-dimensional analysis **without reprogramming** by integrating historical data (internal data and external data) by the subject
 - → **End-user computing**
 - A user can manage data directly.
 - No additional programming is needed.

23

04. Data Warehouse

- Definition of DW
 - Integrated systems with **online** multi-dimensional analysis without reprogramming by integrating historical data (internal data and external data) by the subject
 - → **Immediately and networked**

24

04. Data Warehouse

- Definition of DW
 - Integrated systems with online **multi-dimensional analysis** without reprogramming by integrating historical data (internal data and external data) by the subject
 - → (Theoretically) able to analyze data 60-dimensionally

25

04. Data Warehouse

- Definition of DW
 - **Integrated systems** with online multi-dimensional analysis without reprogramming by integrating historical data (internal data and external data) by the subject
 - → Linked with traditional operation systems
 - → The interface (data extraction, loading automation, ...) among systems are very important.

26

04. Data Warehouse

- Technical Backgrounds of DW
 - Valid ROI
 - High performance of hardware
 - Low cost of hardware
 - High cost of traditional information repository
 - → Difficult to implementation enterprise-widely

27

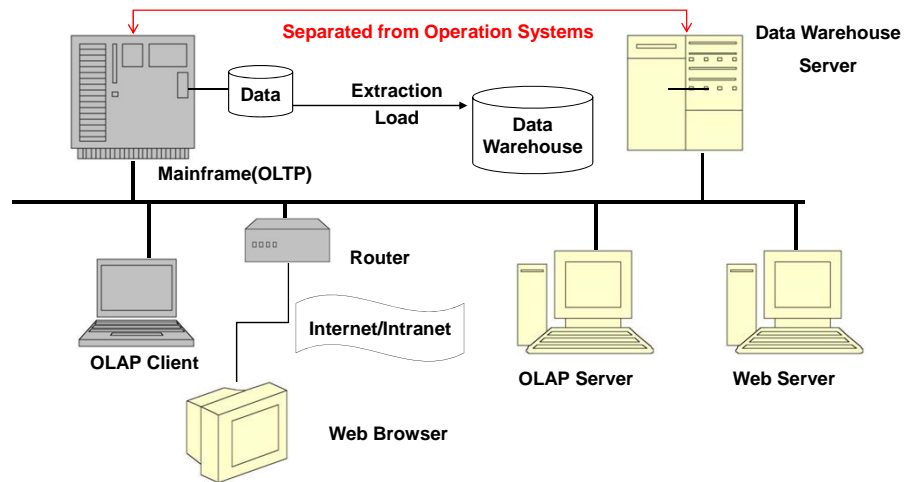
04. Data Warehouse

- Technical Backgrounds of DW
 - Introduction of new hardware technologies
 - SMP(Symmetric Multi Processing)
 - MPP(Massively Parallel Processing)
 - RDBMS(Relational DataBase Management Systems)
 - Introduction of new software technologies
 - DW tools
 - ETT(Extraction, Transformation, Transportation)
 - Multi-dimensional modeling
 - OLAP(On-Line Analytical Processing)

28

04. Data Warehouse

- Basic Structure of DW



29

04. Data Warehouse

- Effects of DW
 - Speedy structured documents
 - Reflecting user's requirements
 - By implementing summary tables
 - Speedy unstructured documents
 - No additional programming
 - User-oriented data-management
 - By setting up only some factors
 - Data analysis and forecasting
 - What-If analysis
 - Using OLAP tools, data mining tools, ...

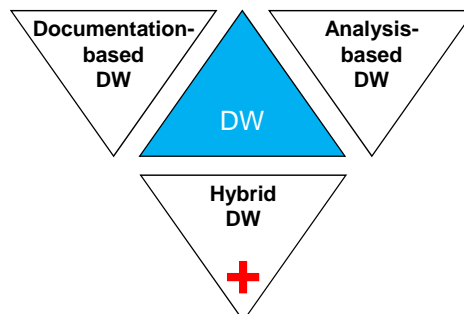
30

04. Data Warehouse

- Effects of DW
 - DW is a major enterprise infrastructure.
 - The core of integrated information systems
 - Cutting off simple documentations and analyses
 - Diversification of information
 - Access to integrated data by various people
 - → Consistent decision making
 - → Enterprise-wide

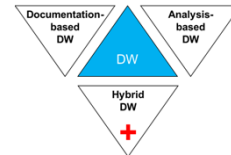
04. Data Warehouse

- Classification of DW
 - Documentation-Based DW
 - Analysis-Based DW
 - Hybrid DW



04. Data Warehouse

- Classification of DW
 - Documentation-Based DW
 - Managing temporary documentation and queries
 - Based on integrated DB
 - Extended offline transactions
 - Analysis-Based DW
 - Various OLAP tools are needed.
 - Trend analysis, What-If analysis,
 - It could be apart from DW.
 - → Data mart.
 - Hybrid DW
 - Analysis-based DW after Documentation-based DW



04. Data Warehouse

- Limitations of DW
 - Limitations of data quality
 - Low data quality of operation systems
 - There exists a prejudice that data of operation systems are always correct.
 - Distrustful data cleansing
 - Limitations of OLAP tools
 - OLAP has SQL limitation for all documents.
 - OLAP could require additional programming if necessary.
 - BPR(Business Process Reengineering) could be considered.