Managing Information Resources

Chapter 7

Outline

- Introduction
- Managing Data
  - The Three-Level Database Model
  - Four Data Models
  - Getting Corporate Data into Shape
- Managing Information
  - Four Types of Information
  - Data Warehouses
  - Document Management
  - Content Management

Concepts and Definitions

- Data
  - Facts devoid of meaning or intent
    - E.g. structured data in DB
- Information
  - Data that has meaning (data in context)
    - E.g. course selection info in a student management system, documents, voice, video...
- Knowledge:
  - Information with direction or intent
- Content
  - Term for the Web age
    - E.g. text, graphics, animation, photos, film clips etc.

Introduction

- **Managing Information Resources** initially meant managing data, first in files, then in corporate databases which were:
  - Well structured
  - Carefully defined, and
  - Controlled by IS department
- Next expanded to include “information” (data with meaning)
- Also much talk of managing knowledge
- With the emergence of the Internet, talk has now turned to managing content:
  - Text, graphics, sound, video and animation

Introduction

- As the breadth of the kinds of information resources has expanded, so has the job of managing them. The job may not start in the IS department but it invariably ends up there
  - PCs users used ‘alone’
    - Needed to share files
    - Version control, back-up etc.
  - Web sites / content
    - Initially created their own
    - Need for recovery, version control
    - Corporate consistency
  - IS to the ‘rescue’
    - Management procedures
    - Discipline

Information Resource Management Responsibilities

- Corporate databases
  - Distributed
  - Various data models
  - Data warehouse
- Information
  - Documents
  - Web contents
- Knowledge management
  - Explicit knowledge (know-what)
  - Tacit knowledge (know-how)

IS has been continually managing new forms of information resources

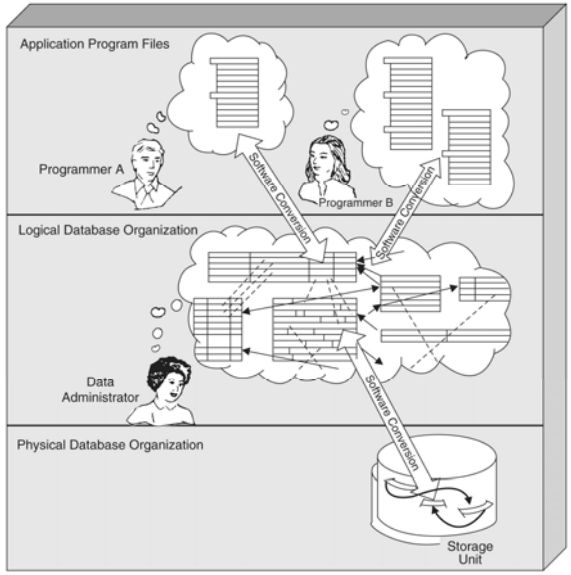
Managing Data

- Database management systems are the main tool for managing computerized corporate data
- They have been around since the 1960s and are based on two major principles:
  - A three-level conceptual model and
  - Several alternative ‘data models’ for organizing the data

The Three-Level Database Model

- **Level 1** – The External, Conceptual, or Local Level
  - Containing the various “user views” of the corporate data that each application program uses
  - Not concerned with how the data will be physically stored or what data is used by other applications
- **Level 2** – The Logical or Enterprise-data Level
  - ‘Technical’ (human) view of the database under control of the DBA
- **Level 3** – The Physical or Storage Level
  - Specifying the way the data is physically stored
  - End user not concerned with all these ‘pointers and flags’ (how the data is physically organized), they are for use by the DBMS

The Three-Level Database Model



Example of Three-Level Database Model

Level 1

Stuent ID	Student name	Course	Score
10021	Jack	Software Engineering	79
10021	Jack	Data structure	76
10022	James	Software Engineering	85
10022	James	Data structure	88

Level 2

Table Student

StuID	StuName	Age
10021	Jack	21
10022	James	20

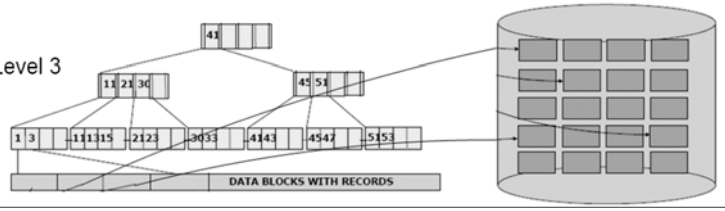
Table CourseSelect

StuID	CourseID	Score
10021	373	79
10022	373	85
10021	275	76
10022	275	88

Table Course

CourseID	CourseName	Capacity	Room
373	Software engineering	30	AQ5018
275	Data structure	40	AQ3023

Level 3



Advantages of Three-Level Database Model

- Level 2 absorbs changes made at Level 3 such as using a new physical storage device
  - Individual application programs in Level 1 do not need to be changed when the physical layer changes
- Data only needs to be stored once in Level 2, and different programs can draw on it and vary the relationships among the data

Four Data Models

- Hierarchical Model
  - Structures data so that each element is subordinate to another in a strict hierarchical manner (Parent & child)
- Network Model
  - Allows each data item to have more than one parent,
  - Relationships stated by pointers stored with the data
- Relational Model
  - Eight relational operations can be performed on this data: Select, Project, Join, Product, Intersection, Difference, Union, Division
- Object Model
  - Storing and managing data as objects
  - A competitive candidate for storing XML data

The second major concept in database management is alternate ways to define relationships among data

Relational Model

- Relational systems are not as efficient as hierarchical or network database systems, but because relational systems allow people to create relationships among data on the fly, they are much more flexible
- First used to handle end user queries – they are now widely used in high-volume transaction systems with huge files
- Hence, they have become the database technology of choice in today's systems
  - Also largely due to decrease in costs of technology: processing and disk storage

Object Models

- Object models retain traditional DBMS features including:
  - End user tools
  - High level Query languages
  - Concurrency control
  - Recovery
  - Ability to handle huge amounts of data
- Include two other major concepts
  - **Object Management** – Management of complex kinds of data such as multimedia and procedures
  - **Knowledge Management** – Management of large numbers of complex rules for reasoning and maintaining integrity constraints between data

Inconsistent Data Definitions

- **The Problem:** Management can not get consistent view across the enterprise
  - 1960s-1970: Application developed in separation → "information islands"
  - Different units in an organization developed their used their own database and their own applications
    - Inconsistent data definitions
    - Duplicate data
- **The Cause:** An application-driven approach
  - Getting applications running as quickly as possible
- **The Solution:** A data-driven approach
  - Data of interest → Data source → Applications
  - Usually evolves from the application-driven chaos

The Role of Data Administration

- Managing data as a corporate resource is more than installing a DBMS
  - Database Administrator (DBA) – Administering databases and software that manages them
  - Data Administrator – Managing enterprise-wide data resources
    - Clean up the data definitions
    - Control shared data
    - Manage data distribution, and
    - Maintain data quality

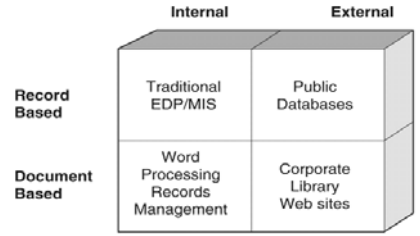
Getting Corporate Data into Shape

- **Enterprise Resource Planning (ERP)** aim to integrate all data and processes of an organization into a unified system
  - Automate and integrate the majority of business processes
  - Share common data and practices across the entire enterprise
  - Produce, access and manage information in a real-time environment
  - Configure application to meet business needs
- ERP provided the means to consolidate data to give management a corporate-wide view of operations
- Interestingly, many companies really did not take these four jobs seriously until the mid 1990s, when they needed consistent data to install a company-wide ERP package

Read Case Example P.270-272

Four Types of Information

- Internal Record-based Information
  - Such as those found in databases
- Internal Document-based Information
  - Such as reports, opinions, e-mails and proposals. Pertains to concepts: ideas, thoughts, etc.
- External Record-based Information
  - Such as acquisition from external databases.
- External Document-Based Information
  - WWW



Sources of Information

- Two sources of information: internal and external
  - Internal record-based information: traditional focus of IS
  - External record-based information: public DB
- Internal and external document-based information have received little attention from IS until recently
  - However, it is estimated that 90% of an organization's information is in documents rather than structured databases (Sprague, 1995).

Structures of Information

- The two different structures of information are managed in different ways
  - Record-based: Facts about entities
    - Data warehouse
  - Document-based: Dealing with concepts (Housed in documents, messages, video, audio clips...)
    - Document management systems
    - Web content management

	Data Records	*Documents*
Item of interest	Entity	Concept or idea
Attribute of item	Field	Set of symbols
All attributes for item	Record	Logical paragraph
All related items	File	Document
A group of related files	Database	File cabinet
A collection of databases	Application system	Library, records center
Data models (representational approaches)	Hierarchical, relational, etc.	Keywords, hypertext

Scope of Information Management

	Typical Corporate Authority	*Information Sources*	*Technologies Used*
Internal record-based information	Information systems department	Transaction processing Organizational units	DBMS Data dictionaries Enterprise data analysis techniques
Internal document-based information	Administrative vice president Word processing center Records management	Corporate memos, letters, reports, forms, e-mail	Word processing Micrographics Reprographics Text-retrieval products
External record-based information	End users Corporate planning Financial analysis Marketing	Public databases	Internet-based services Public networks Analysis packages
External document-based information	Corporate library	Public literature News services Catalogs and indexes Subscriptions Purchased reports	Bibliographic services Environmental scanning Public networks

On Line Transaction Processing (OLTP)

- Maintains a database that is an accurate model of some real-world enterprise.
- Supports day-to-day operations.
- Characteristics:
  - Short simple transactions
  - Relatively frequent updates
  - Transactions access only a small fraction of the database

On Line Analytic Processing (OLAP)

- Uses information in database to guide strategic decisions.
- Characteristics:
  - Complex queries
  - Infrequent updates
  - Transactions access a large fraction of the database
  - Data need not be up-to-date

What is Data Warehouse?

- "A data warehouse is a **subject-oriented**, **integrated**, **time-variant**, and **nonvolatile** collection of data in support of management's decision-making process." — W. H. Inmon

Read Case Example P.277-279

Data Warehouse: Subject-Oriented

- Organized around major subjects, such as customer, product, sales
- Focusing on the modeling and analysis of data for decision makers, not on daily operations or transaction processing
- Provide a simple and concise view around particular subject issues by excluding data that are not useful in the decision support process

Data Warehouse: Integrated

- Constructed by integrating multiple, heterogeneous data sources
  - Relational databases, flat files, on-line transaction records
- Data cleaning and data integration techniques are applied.
  - Naming conventions, encoding structures, attribute measures, etc. among different data sources
  - When data is moved to the warehouse, it is converted.

Data Warehouse: Time Variant

- The time horizon for the data warehouse is significantly longer than that of operational systems
  - Operational database: current value data
  - Data warehouse data: provide information from a historical perspective (e.g., past 5-10 years)
- Every key structure in the data warehouse
  - Contains an element of time, explicitly or implicitly

Data Warehouse: Nonvolatile

- A physically separate store of data transformed from the operational environment
- Operational update of data does not occur in the data warehouse environment
  - Does not require transaction processing, recovery, and concurrency control mechanisms
  - Requires only two operations in data accessing:
    - Initial loading of data and access of data

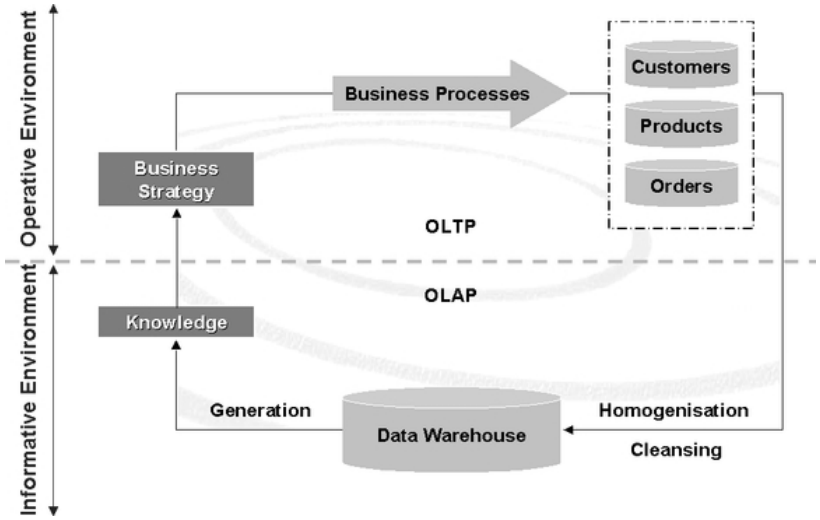
Data Warehouse vs. Heterogeneous DBMS

- Traditional heterogeneous DB integration: A query driven approach
  - Build wrappers/mediators on top of heterogeneous databases
  - When a query is posed to a client site, a meta-dictionary is used to translate the query into queries appropriate for individual heterogeneous sites involved, and the results are integrated into a global answer set
  - Complex information filtering, compete for resources
- Data warehouse: update-driven, high performance
  - Information from heterogeneous sources is integrated in advance and stored in warehouses for direct query and analysis

Data Warehouse vs. Operational DBMS

- OLTP (on-line transaction processing)
  - Major task of traditional relational DBMS
  - Day-to-day operations: purchasing, inventory, banking, manufacturing, payroll, registration, accounting, etc.
- OLAP (on-line analytical processing)
  - Major task of data warehouse system
  - Data analysis and decision making

OLTP vs. OLAP



OLTP vs. OLAP

	OLTP	OLAP
Users	Clerk, IT professional	Knowledge worker
Function	Day to day operations	Decision support
Database design	Application-oriented	subject-oriented
Data	Current, up-to-date detailed, flat relational isolated	Historical, summarized, multidimensional, integrated, consolidated
Usage	Repetitive	Ad-hoc
Access	Read/write, index/hash on primary key	Lots of scans
Unit of work	Short, simple transaction	Complex query
No. of records accessed	Tens	Millions
No. of users	Thousands	Hundreds
DB size	100MB-GB	100GB-TB
Metric	Transaction throughput	Query throughput, response

Key Concepts in Data Warehouse

- **Metadata:** The part of the warehouse that defines the data. Metadata means "data about data."
  - Metadata explains the meaning of each data element, how each element relates to each other, etc.
  - It sets the standard – without it data from different legacy systems cannot be reconciled, so the data will not be "clean"
- **Quality Data:** Is the cleaning process to adhere to metadata standards
  - The older the data the more suspect its quality
- **Data Marts:** Is a subset of data pulled off the warehouse for a specific group of users
  - In the early 1990s, one huge warehouse was envisaged, but proved un-practical due to long search times and large cost factors

Steps in a Data Warehousing Project

1. Define the business uses of the data
2. Create the data model for the warehouse
   - i.e. defining the relationships between the data elements
3. Cleanse the data
4. Select the user tools
   - Consider the users point of view by selecting the tools they will use & then training them on tool use
5. Monitor usage and system performance

Data warehouses are seen as strategic assets that can yield new insights into customer behavior, internal operations, product mixes and the like. But to gain the benefits, companies must take the step of reconciling data from numerous legacy systems

Document Management

- Estimated that 90% of an organization's information is in documents rather than structured databases
- Byte Magazine define a document as a snapshot of some set of information that can
  - Incorporate many complex information types;
  - Exist in multiple places across a network;
  - Depend on other documents for information
  - Change on the fly (as subordinate documents are updated)
  - Having an intricate structure or complex data types as full-motion video and voice annotations;
  - Be accessed and modified by many people simultaneously (if they have permission to do so)

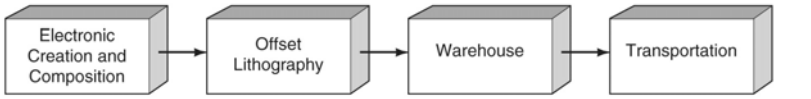
Fundamental Roles of Documents

- 4 Fundamental roles of documents
  - As a ***product***, or support for a product
  - As a fundamental mechanism for ***communication*** among people and groups within an organization and between organizations.
  - As the primary vehicle for ***business processes***
  - As an important part of ***organizational*** memory

Improving the Publishing Process

- Technology enables a major restructuring of the process of publishing and distributing paper documents
- Traditional Process – designed primarily for high volume and high quality documents
- Process has inefficiencies:
  - Infrequent long print run requires storing documents which become obsolete between runs
  - 60% of the total cost of delivering theses documents is in storage & transportation

FIGURE 7-6 Traditional Publishing Process



Improving the Publishing Process

- The revised publishing/distribution process using newer technologies
  - Documents are stored electronically, shipped over a network, and printed when they are needed
- The major benefits result from reducing obsolescence, eliminating warehouse costs & reducing or eliminating delivery time

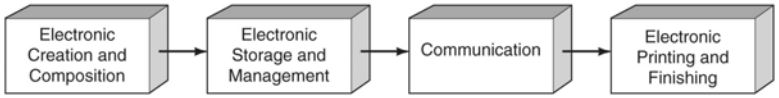


FIGURE 7-7 Reengineered Publishing Process

Read Case Example P.282

Supporting Communication among People and Groups

- The value of documents is that they transfer information across time and space
  - Internet can help but often still rely on 'paper' documents
- Electronic Document Management (EDM) can be used to facilitate such communications among people and groups

Read Case Example P.283-285

Supporting Organizational Processes

- Documents are still the vehicle for accomplishing most processes in organizations
  - Many such "Workflow systems" heavily based on the physical circulation of paper forms
- The use of technology to support processes generates significant value in reducing physical space for handling forms, faster routing of forms, and managing and tracking forms flow & workload
- In addition to improving transaction-oriented business processes with EDM, many organizations are improving the management processes of reporting, control, decision making, and problem solving as well

Read Case Example P.286-287

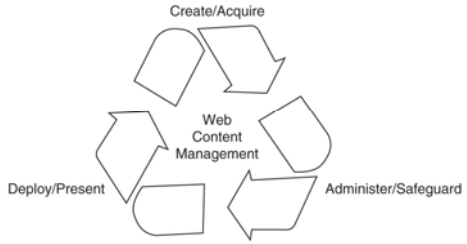
Content Management

- Corporate intranets now house documents that were previously paper-based
- A major reason content has become important to CIO is because it is a core management discipline underlying online business
  - Without production-level Web content, management processes, and technologies, large-scale e-business is not possible
  - Use of XML moves Web content from being in a human-only readable format to being in a computer-readable format
  - Thus, the content can be passed to back-end transaction processing systems and cause an action to take place
- Content is no longer static; it is active

Content Management

- To create a content management strategy, need to understand the three phases of the content management life cycle:
  - Managing Content Creation and Acquisition
  - Content Administration and Safeguarding
  - Content Deployment and Presentation

FIGURE 7-8 The Web Content Management Life Cycle



Source: Tueber, Chuck, *Dealing in Web Currency*, Gartner EXP, 56 Top Gallant, Stamford, CT, June 2001.

Managing Content Creation and Acquisition

- Content creation and acquisition need to focus on creating content quality
  - That's why it might be wise to buy some content from specialists – which is called syndicated content – rather than create it in-house
- The best organizational structure is to distribute content creation and maintenance to content-expert employees
  - But to avoid anarchy, these dispersed experts should be directed centrally, and use centrally-created formats and an automated workflow system that moves their work along

Content Administration and Safeguarding

- The emphasis in this phase, like any operational phase, is efficiency
  - The goal is achieve the most with the least effort
- Content management tools can be used to identify types of content and the business rules that apply to each type
- Whereas content creation should be distributed, content administration should be centralized

Content Deployment and Presentation

- Emphasis on effectiveness
  - i.e. Presenting the content so that it attracts visitors, allows them to navigate the site easily, and leads them to the desired actions
- Features to attract and keep visitors
  - Personalization: allowing visitors to customize how they view the page
  - Localization: tailoring a site to a culture, market or locale
  - Multichannel distribution: appropriate display for various devices

Read Case Example P.290-292

Managing Blogs

- **Blog** (Web Log) is a web site where an individual makes intermittent Web postings (an online journal)
- Powerful tools
  - Compete with major media
  - Influence on politics and company policy (e.g. 2004 U.S. elections)
- Some forward thinking companies have recognized the power of this immediate form of publishing and communication
  - Used for crisis management?
  - Employees need to be careful else trouble
  - Opportunities and challenges for organizations

Read Case Example P.293-294