Cadeira de Tecnologias de Informação

Ano lectivo 2010/2011

The Database Environment

Parts of this presentation were taken from the backing material of the book

Jeffrey A. Hoffer, V. Ramesh, Heikki Topi (2011) ***Modern Database Management, 10th Edition.***

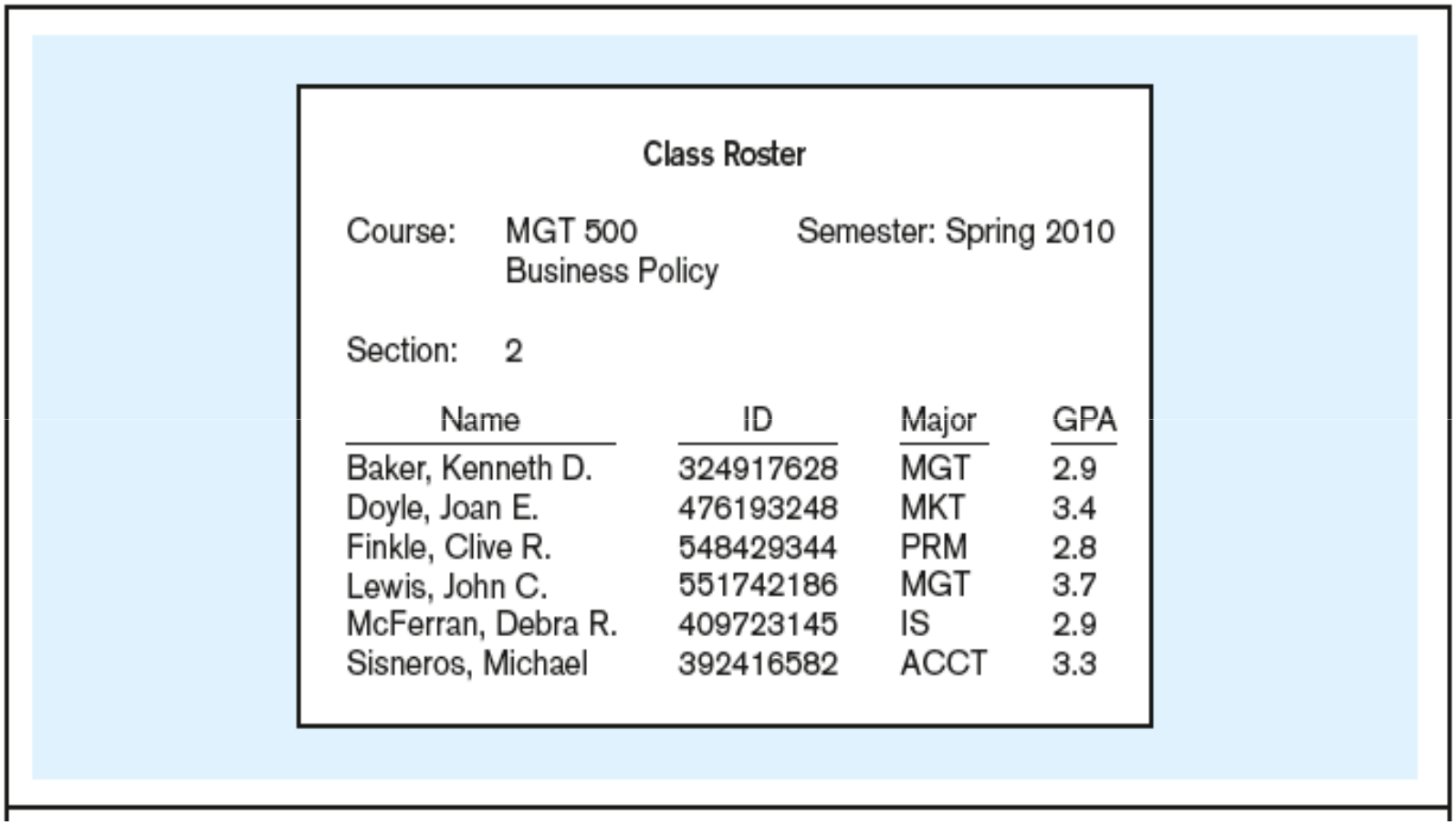
Objectives

- ✓ Define terms
- ✓ Name limitations of conventional file processing
- ✓ Explain advantages of databases
- ✓ List components of database environment
- ✓ Identify categories of database applications

Definitions

- **Database:** organized collection of logically related data
- **Data:** stored representations of meaningful objects and events
  - Structured: numbers, text, dates
  - Unstructured: images, video, documents
- **Information:** data processed to increase knowledge in the person using the data
- **Metadata:** data that describes the properties and context of user data

Data in context



Class Roster

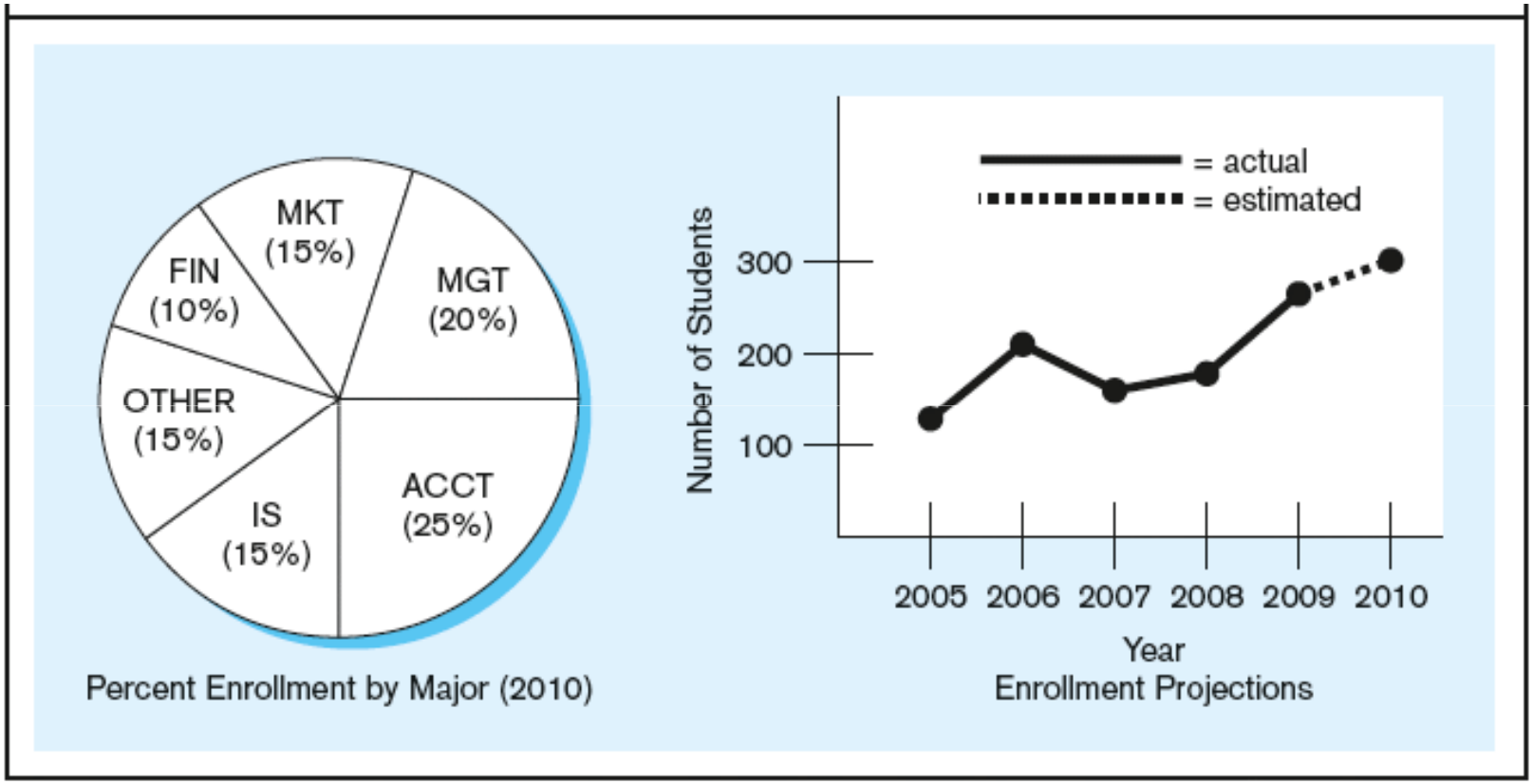
Course: MGT 500 Business Policy Semester: Spring 2010

Section: 2

Name	ID	Major	GPA
Baker, Kenneth D.	324917628	MGT	2.9
Doyle, Joan E.	476193248	MKT	3.4
Finkle, Clive R.	548429344	PRM	2.8
Lewis, John C.	551742186	MGT	3.7
McFerran, Debra R.	409723145	IS	2.9
Sisneros, Michael	392416582	ACCT	3.3

Context helps users understand data

Summarized data



Graphical displays turn data into useful information that managers can use for decision making and interpretation

Metadata

TABLE 1-1 Example Metadata for Class Roster

Data Item		Metadata				
Name	**Type**	**Length**	**Min**	**Max**	**Description**	**Source**
Course	Alphanumeric	30			Course ID and name	Academic Unit
Section	Integer	1	1	9	Section number	Registrar
Semester	Alphanumeric	10			Semester and year	Registrar
Name	Alphanumeric	30			Student name	Student IS
ID	Integer	9			Student ID (SSN)	Student IS
Major	Alphanumeric	4			Student major	Student IS
GPA	Decimal	3	0.0	4.0	Student grade point average	Academic Unit

Descriptions of the properties or characteristics of the data, including data types, field sizes, allowable values, and data context

Disadvantages of File Processing

- **Program-Data Dependence**
  - All programs maintain metadata for each file they use
- **Duplication of Data**
  - Different systems/programs have separate copies of the same data
- **Limited Data Sharing**
  - No centralized control of data
- **Lengthy Development Times**
  - Programmers must design their own file formats
- **Excessive Program Maintenance**
  - 80% of information systems budget

Problems with Data Dependency

- Each application programmer must maintain his/her own data
- Each application program needs to include code for the metadata of each file
- Each application program must have its own processing routines for reading, inserting, updating, and deleting data
- Lack of coordination and central control
- Non-standard file formats

Duplicate Data

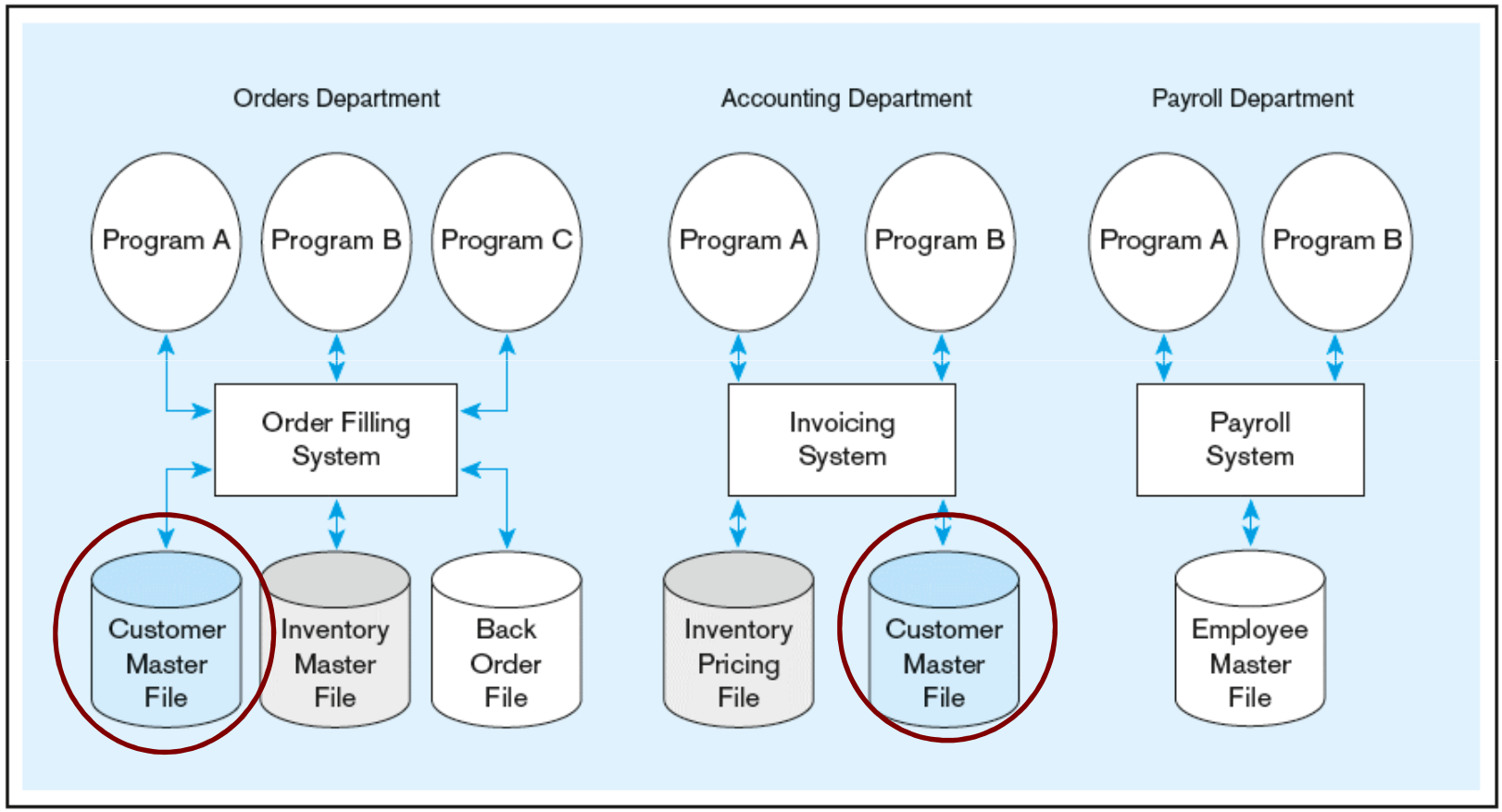


FIGURE 1-2 Old file processing systems at Pine Valley Furniture Company

Problems with Data Redundancy

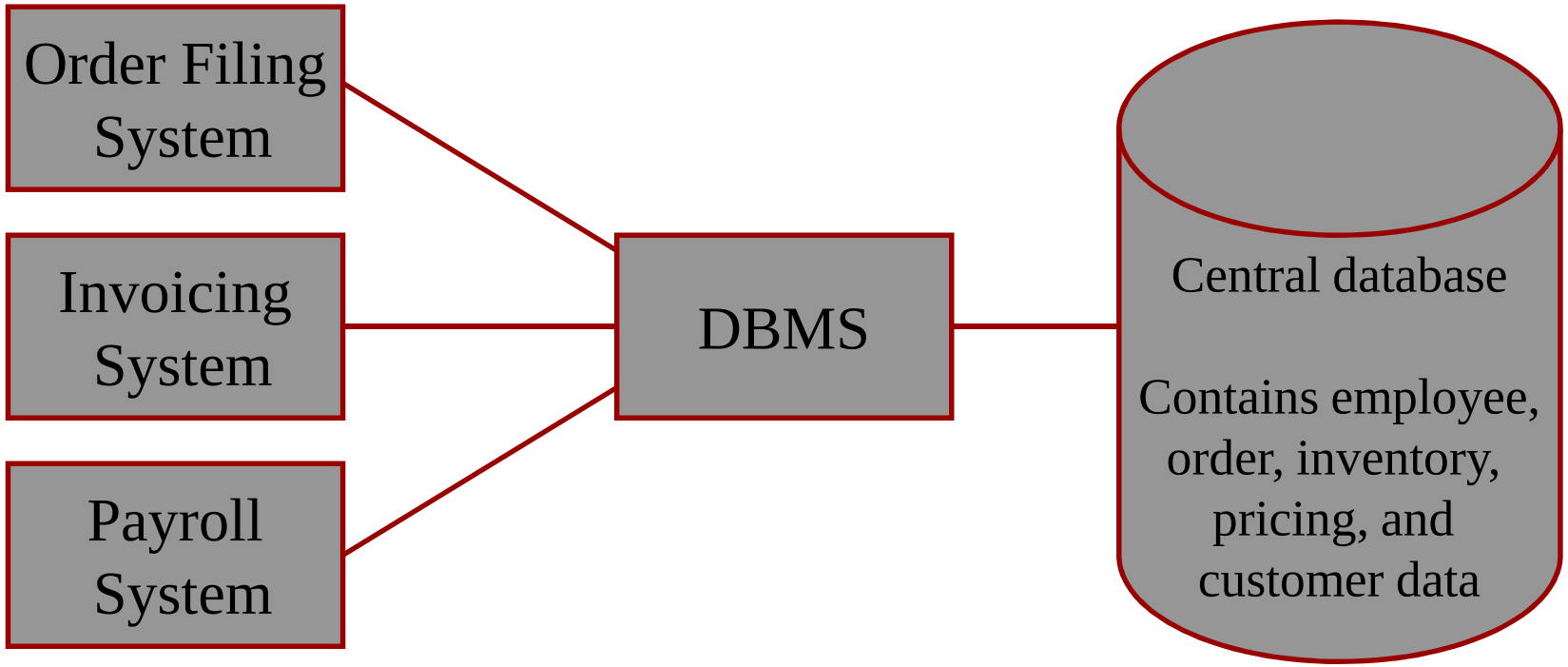
- **Waste of space to have duplicate data**
- **Causes more maintenance headaches**
- **The biggest problem:**
  - **Data changes in one file could cause inconsistencies**
  - Compromises ***data integrity***

SOLUTION:
The DATABASE Approach

- Central repository of shared data
- Data is managed by a controlling agent
- Stored in a standardized, convenient form

Requires a Database Management System (DBMS)

Database Management System
(Gestor de Bases de Dados)



A software system that is used to create, maintain, and provide controlled access to user databases

Advantages of the Database Approach

- **Program-data independence**
- Planned data redundancy
- Improved data consistency
- Improved data sharing
- Increased application development productivity
- Enforcement of standards
- Improved data quality
- Improved data accessibility and responsiveness
- Reduced program maintenance
- Improved decision support

Program-Data independence

- **Physical data independence**

  Changes to the physical level (how the data is stored, whether in arrays or linked lists etc.) must not require a change to an application based on the structure

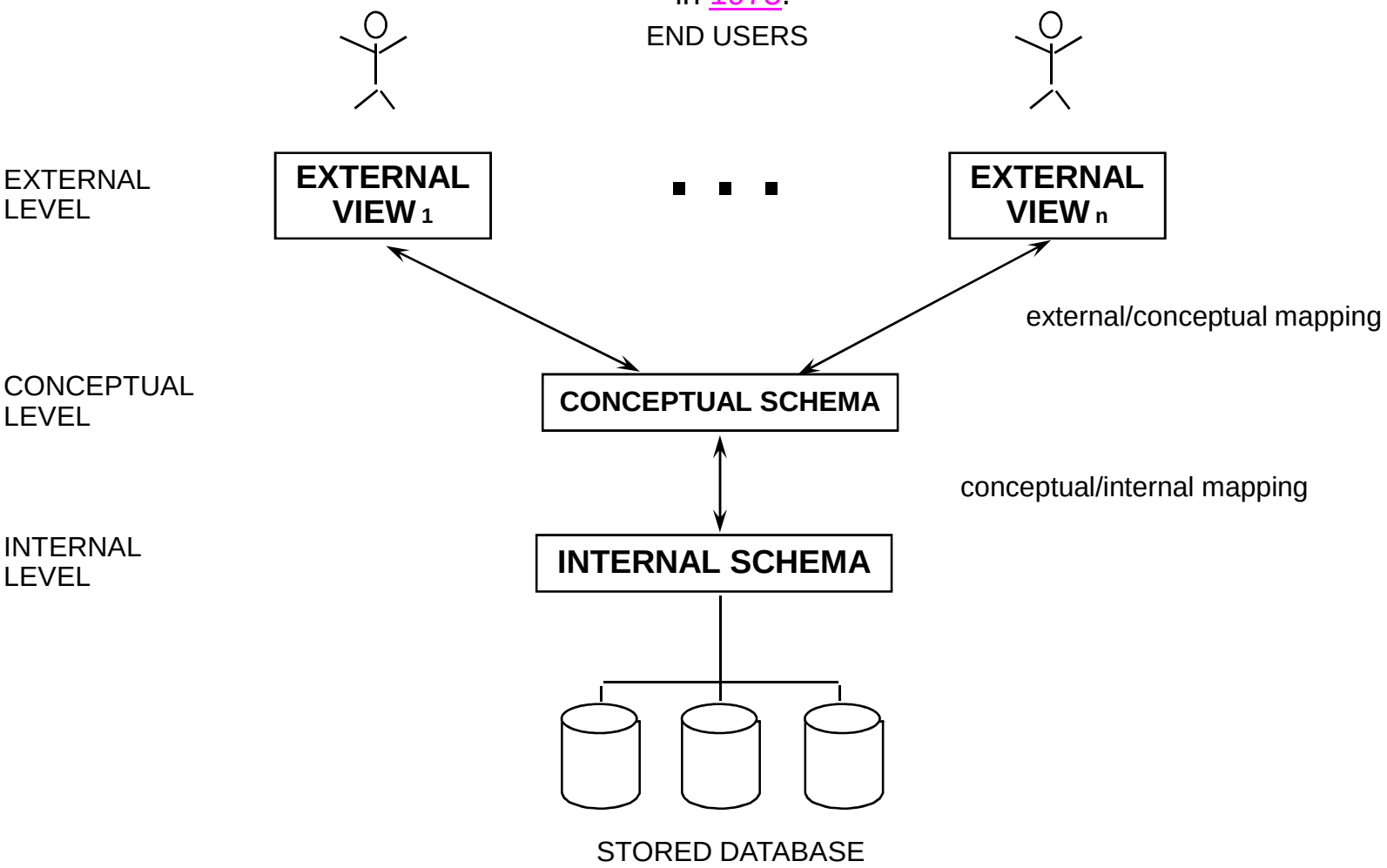
- **Logical data independence**

  Changes to the logical level (tables, columns, rows, and so on) must not require a change to an application based on the structure

  Logical data independence is more difficult to achieve than physical data independence

ANSI-SPARC Architecture

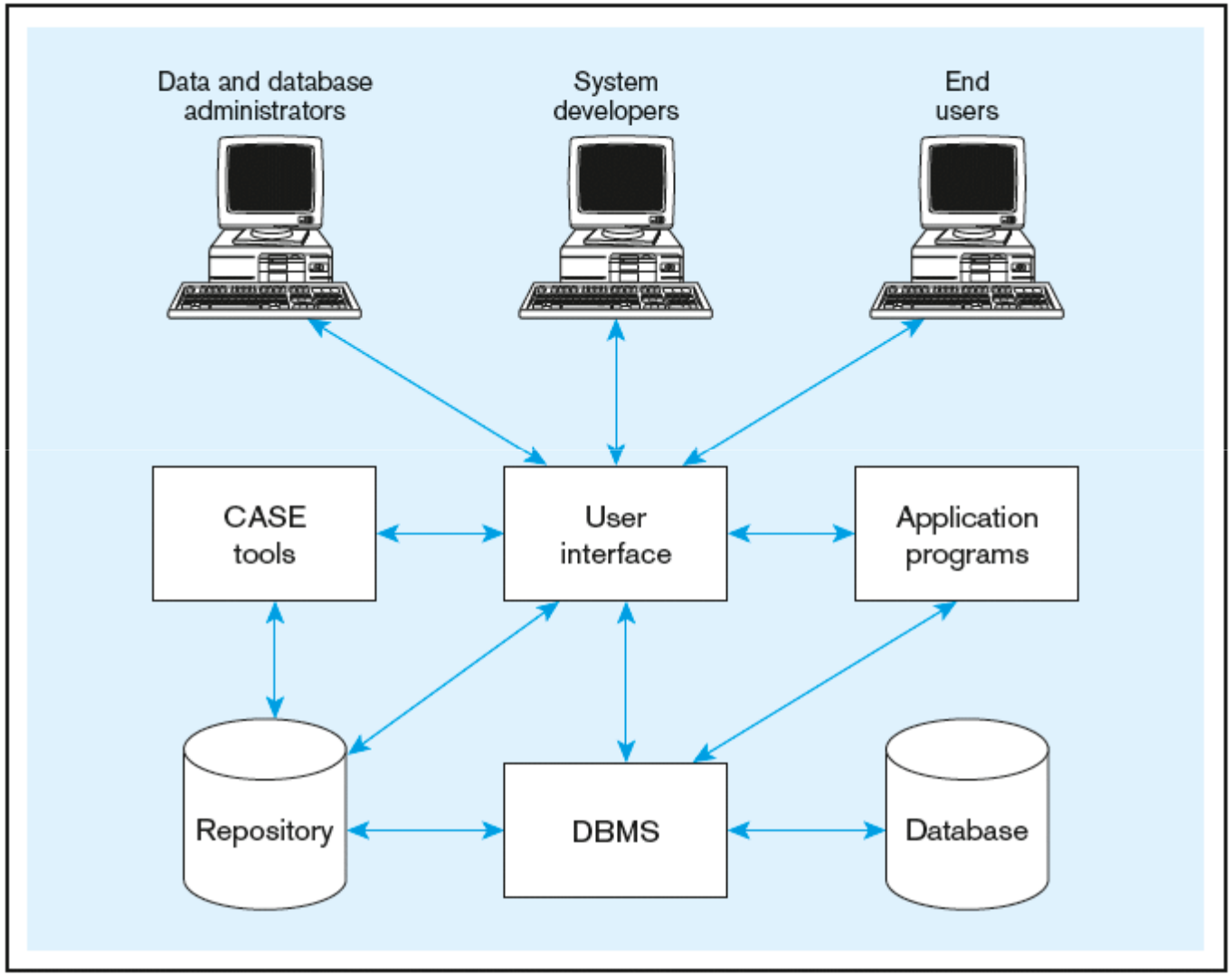
American National Standards Institute, Standards Planning And Requirements Committee) is an abstract design standard for a Database Management System (DBMS), first proposed in 1975.



Elements of the Database Approach

- **Data models**
  - Graphical system capturing nature and relationship of data
  - Enterprise Data Model–high-level entities and relationships for the organization
  - Project Data Model–more detailed view, matching data structure in database or data warehouse
- **Entities**
  - Noun form describing a person, place, object, event, or concept
  - Composed of attributes
- **Relationships**
  - Between entities
  - Usually one-to-many (1:M) or many-to-many (M:N)
- **Relational Databases**
  - Database technology involving tables (relations) representing entities and primary/foreign keys representing relationships

Components of the Database Environment



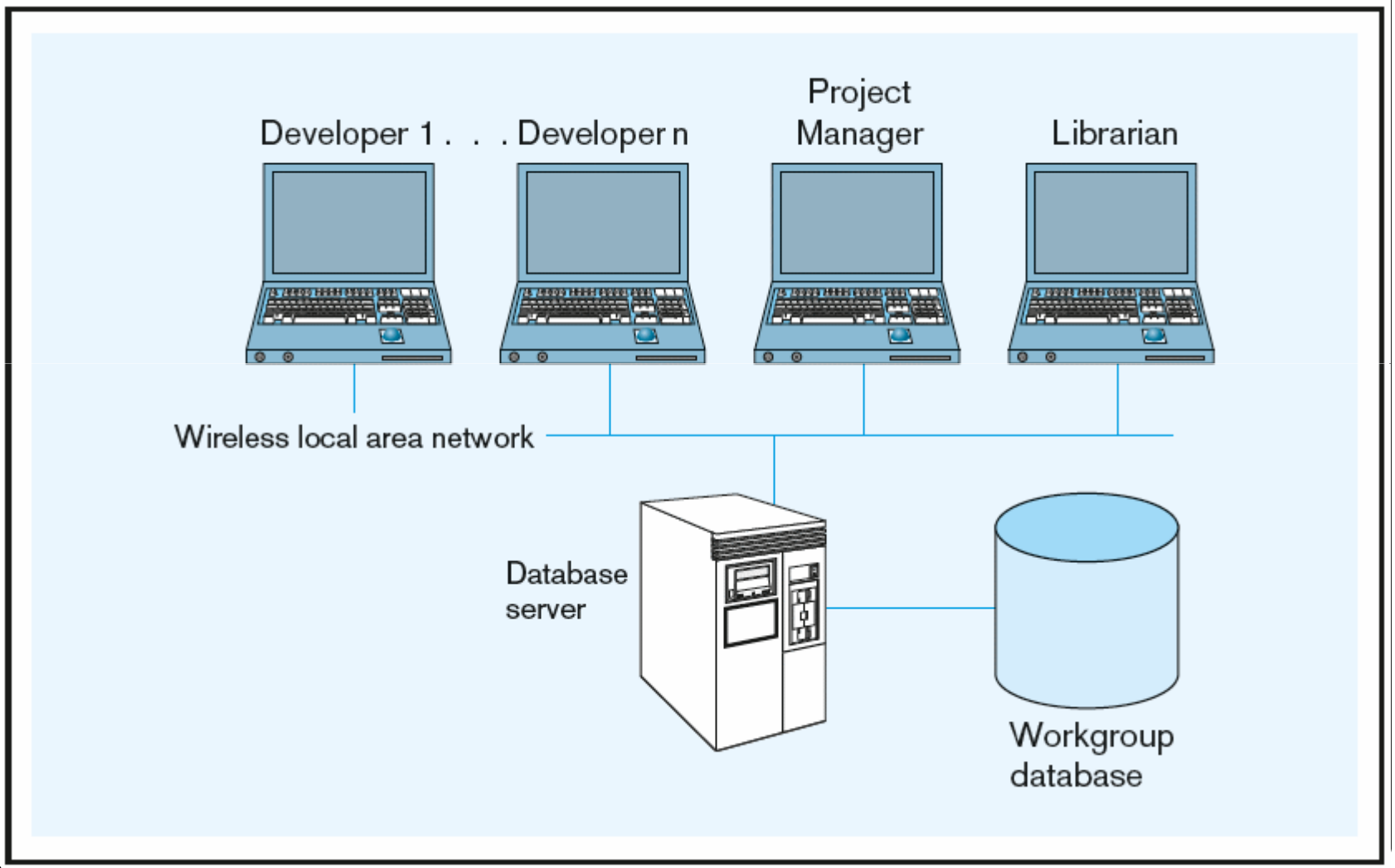
Components of the Database Environment

- CASE Tools–computer-aided software engineering
- Repository–centralized storehouse of metadata
- Database Management System (DBMS) –software for managing the database
- Database–storehouse of the data
- Application Programs–software using the data
- User Interface–text and graphical displays to users
- Data/Database Administrators–personnel responsible for maintaining the database
- System Developers–personnel responsible for designing databases and software
- End Users–people who use the applications and databases

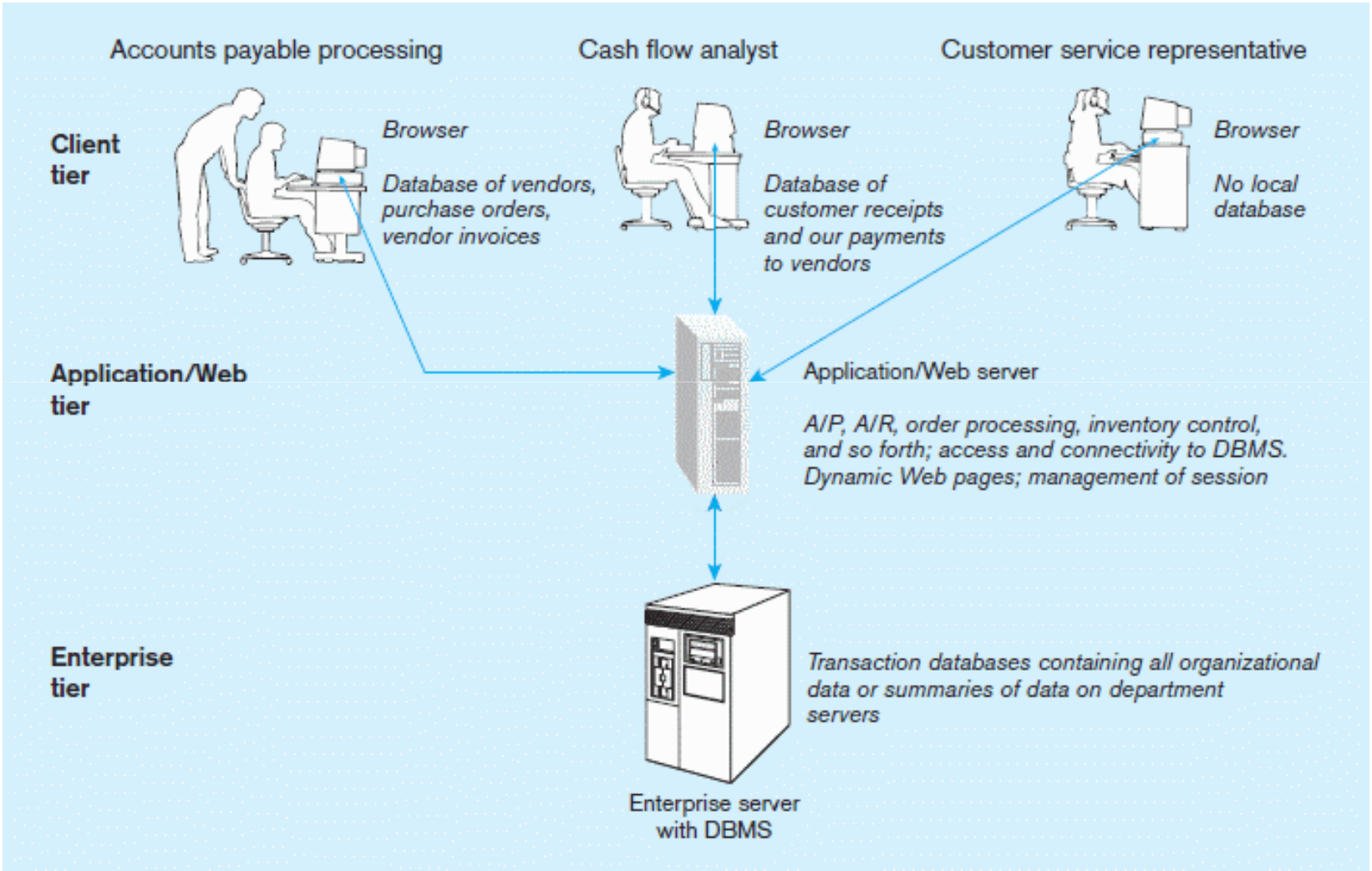
Classification of Database Applications
Perspective I - Scope

- **Personal databases**
- **Two-tier Client/Server databases**
- **Multitier Client/Server databases**

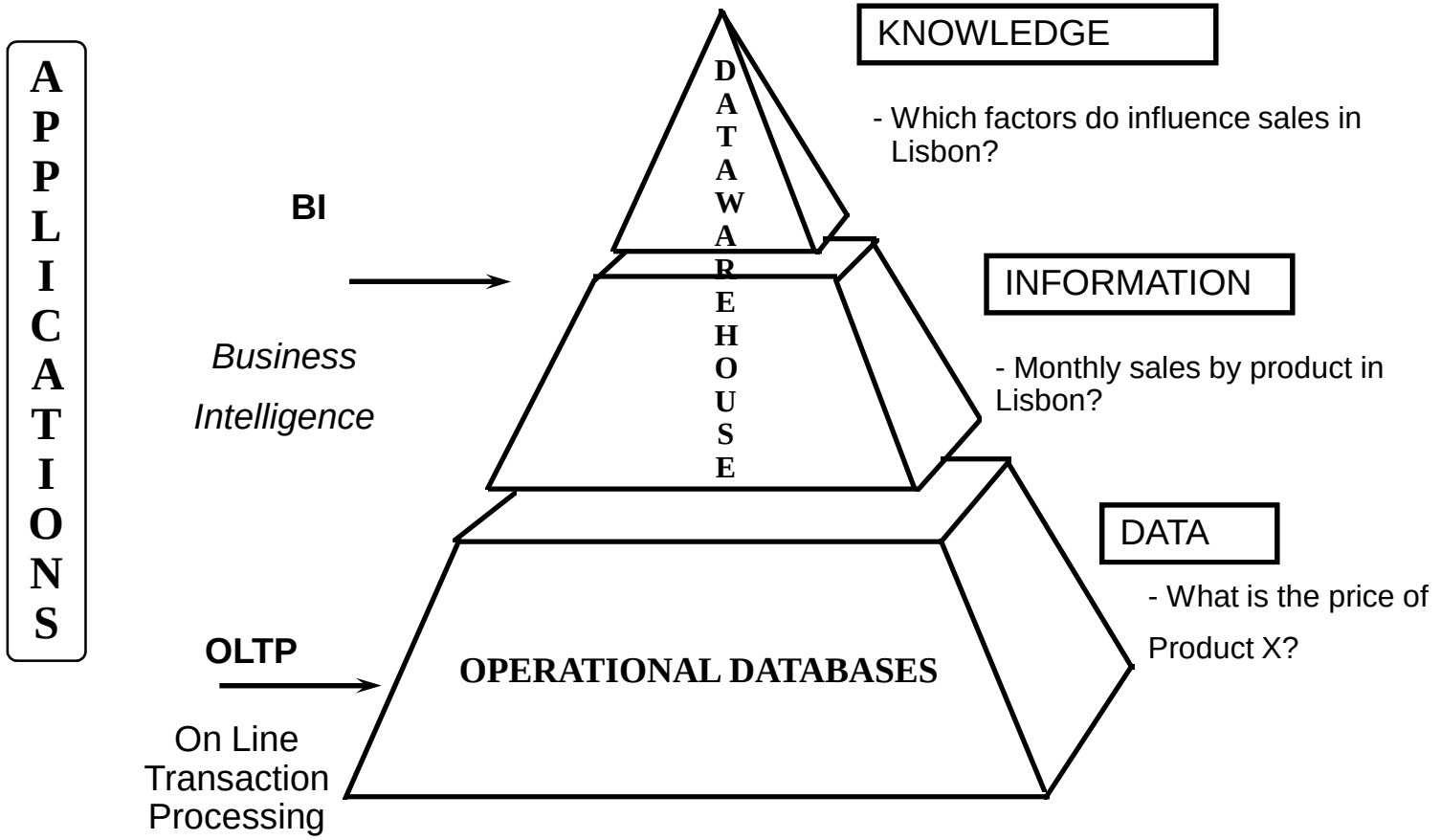
Two-tier database with local area network



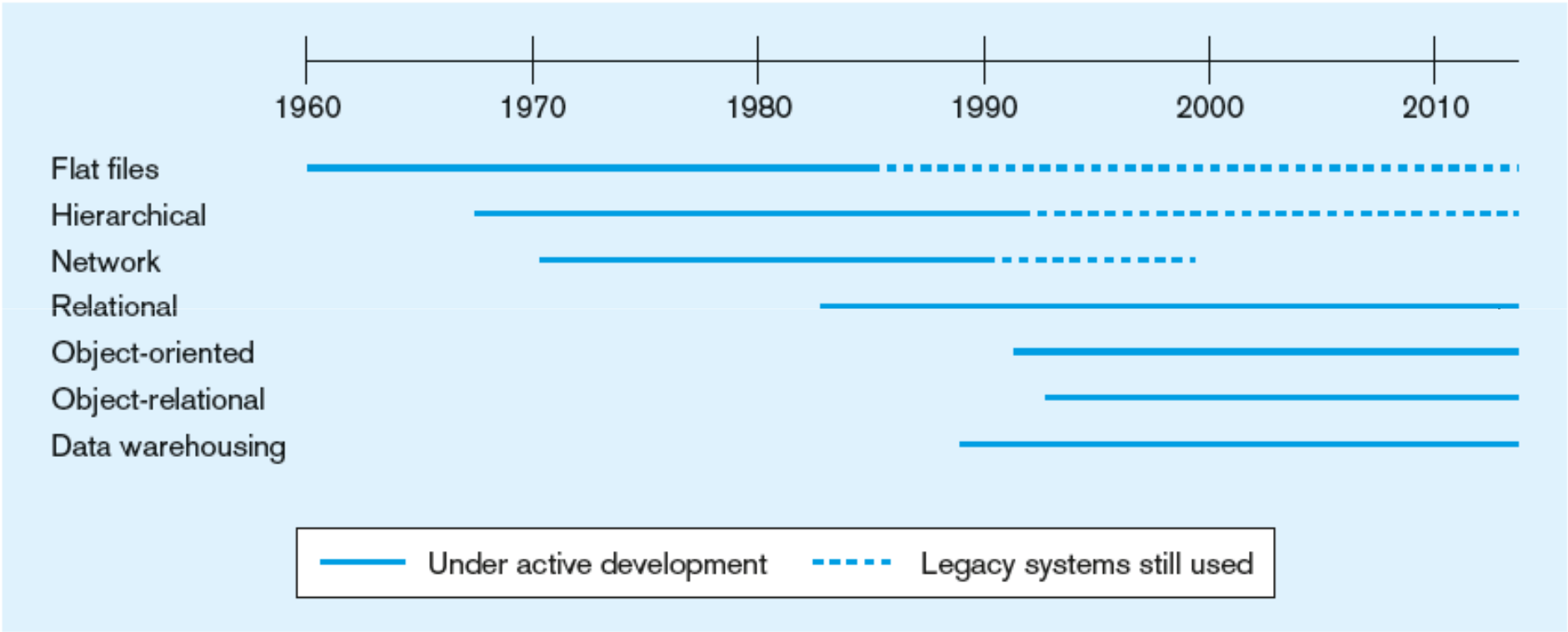
Three-tiered client/server database architecture



Classification of Databases
Perspective II - Type of use



Evolution of database technologies



Commercial DBMS

- Oracle
- DB2
- SQL Server
- Informix
- Sybase
- Ingres
- MySQL
- Postgres
- Access
- .....

Enterprise Data Model

- First step in database development
- Specifies scope and general content
- Overall picture of organizational data at high level of abstraction
- Identify High Level Entities
- Identify High Level Business Processes
- Identify Business rules

Example business function-to-data entity matrix

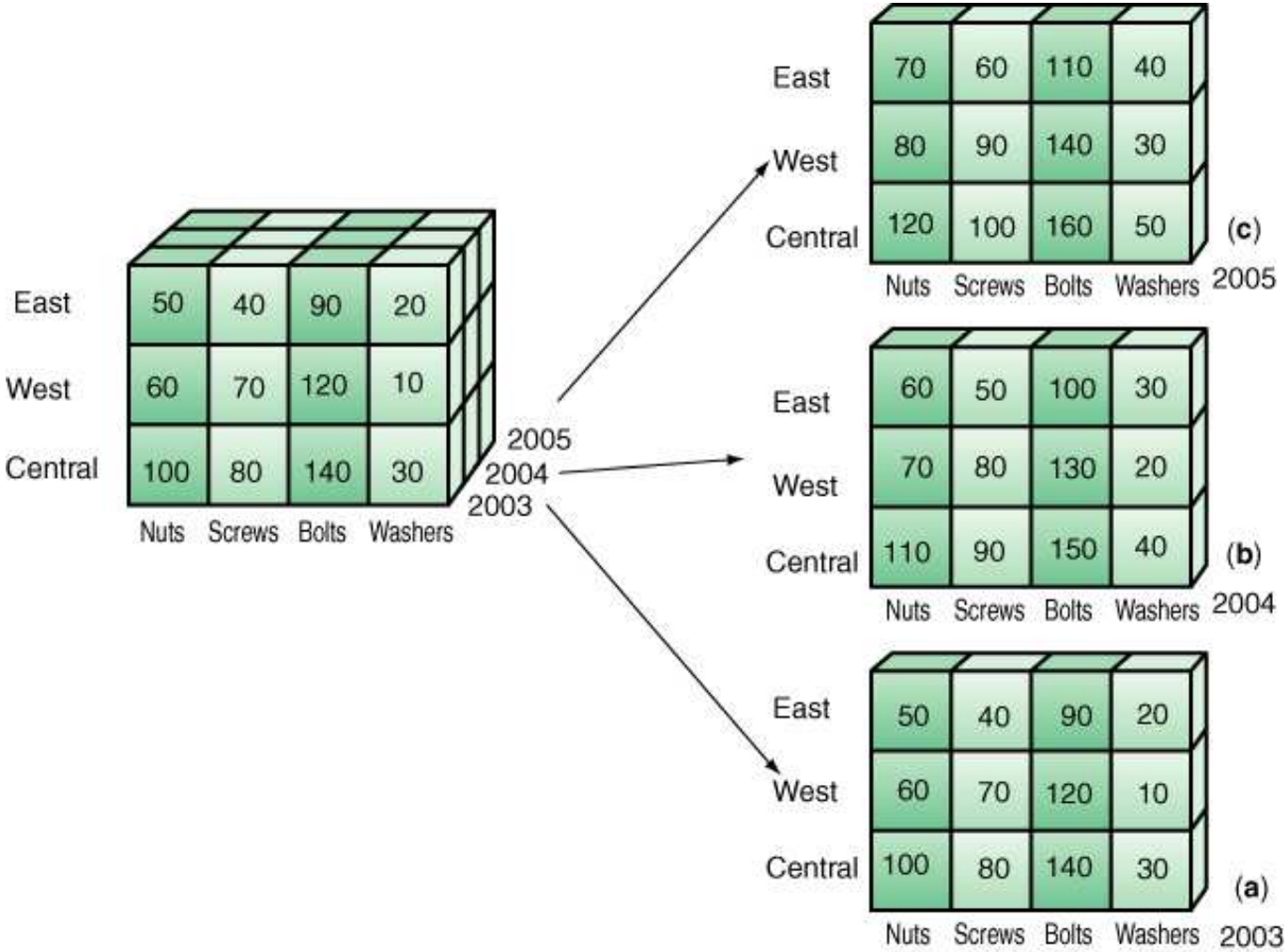
Business Functions \ Data Entity Types	Customer	Product	Raw Material	Order	Work Center	Work Order	Invoice	Equipment	Employee
Business Planning	X	X						X	X
Product Development		X	X		X			X	
Materials Management		X	X	X	X	X		X	
Order Fulfillment	X	X	X	X	X	X	X	X	X
Order Shipment	X	X		X	X		X		X
Sales Summarization	X	X		X			X		X
Production Operations		X	X	X	X	X		X	X
Finance and Accounting	X	X	X	X	X		X	X	X

X = data entity is used within business function

Data Warehouse

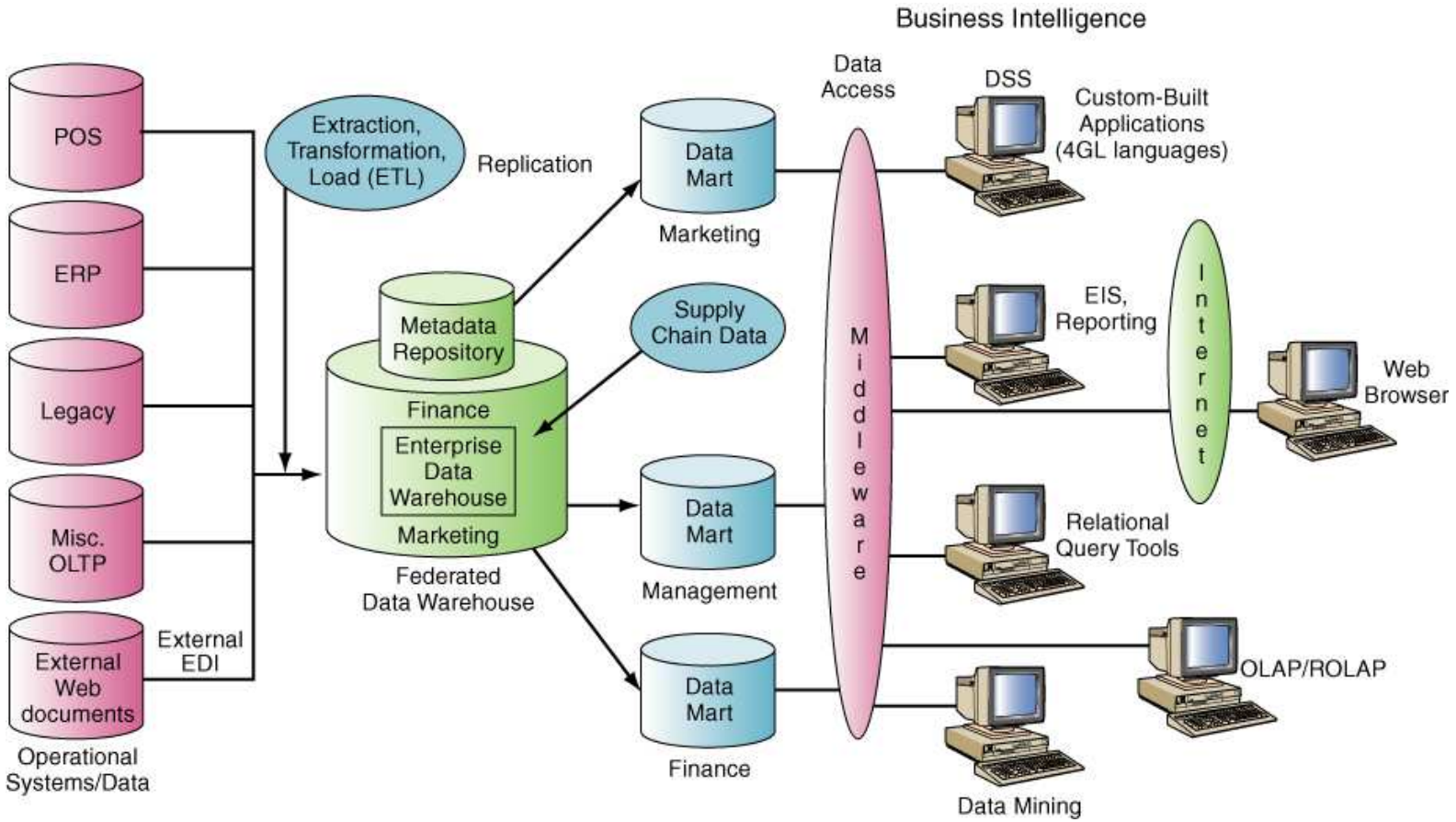
- **Um *Data Warehouse* (DW) é um repositório de dados históricos, organizados por assunto (clientes, produtos, etc) cujo objectivo é suportar a tomada de decisões**
- **Os dados e informações existentes no DW são provenientes das bases de dados operacionais e são extraídos através de ferramentas ETL *(Extraction, Transformation and Loading)***
- **Os dados do DW são normalmente apresentados aos utilizadores na forma multidimensional**

Base de Dados Multidimensional



Reiner, R.K.; Turban, E.; Potter, R.E. (2007). *Introduction to Information Systems – Supporting and Transforming Business*, John Wiley.

Data Warehouse Framework & Views



Reiner, R.K.; Turban, E.; Potter, R.E. (2007). *Introduction to Information Systems – Supporting and Transforming Business*, John Wiley.

Porquê *Data Warehouse?*

- **Disponibiliza uma visão consolidada dos dados organizacionais**
- **Suporta a tomada de decisão, através da utilização de ferramentas de Business Intelligence**

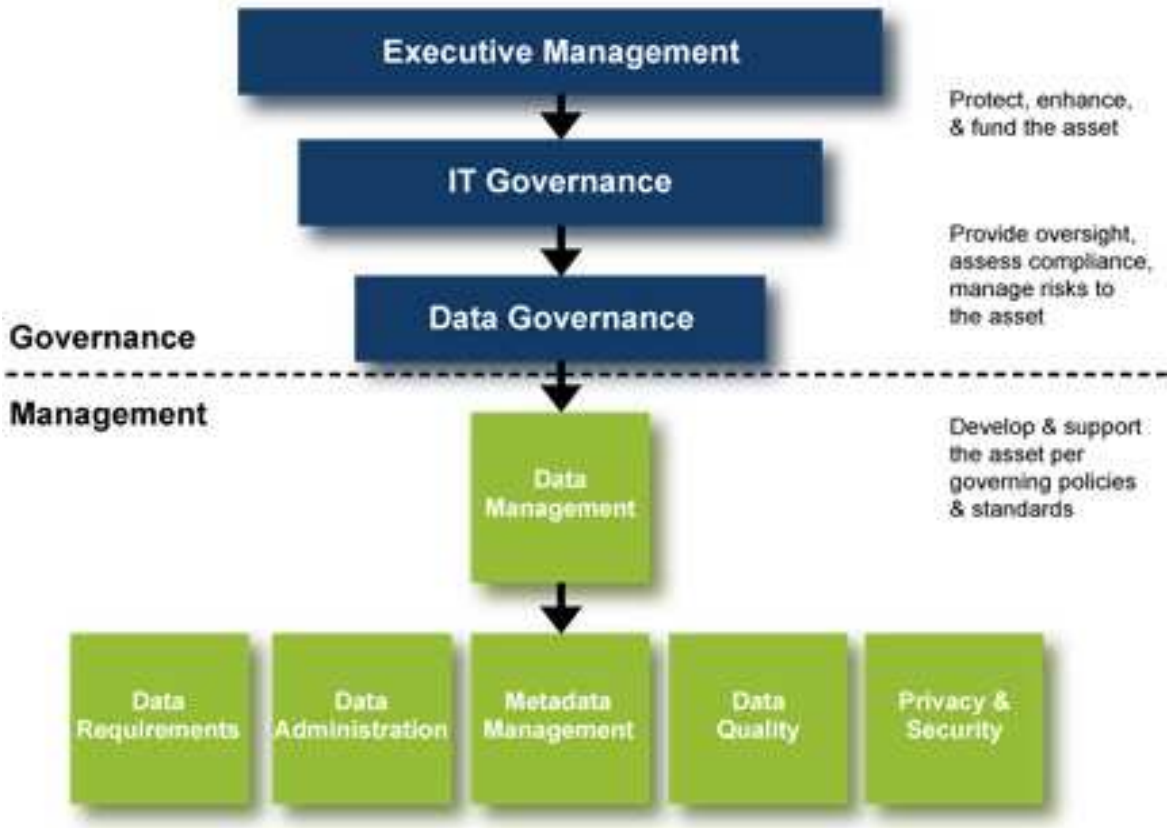
Data Marts

- **Um *data mart* é um sub-cojunto de um *datawarehouse* desenhado para satisfazer as necessidades de uma área de negócio, de um departamento, de uma área geográfica, etc**
- **O *datawarehouse* *pode ser construído bottom up a partir de um conjunto de data marts* ou *top down* em que os *data marts* são cópias de subconjuntos do *datawarehouse***

Data Governance (Governança dos Dados)

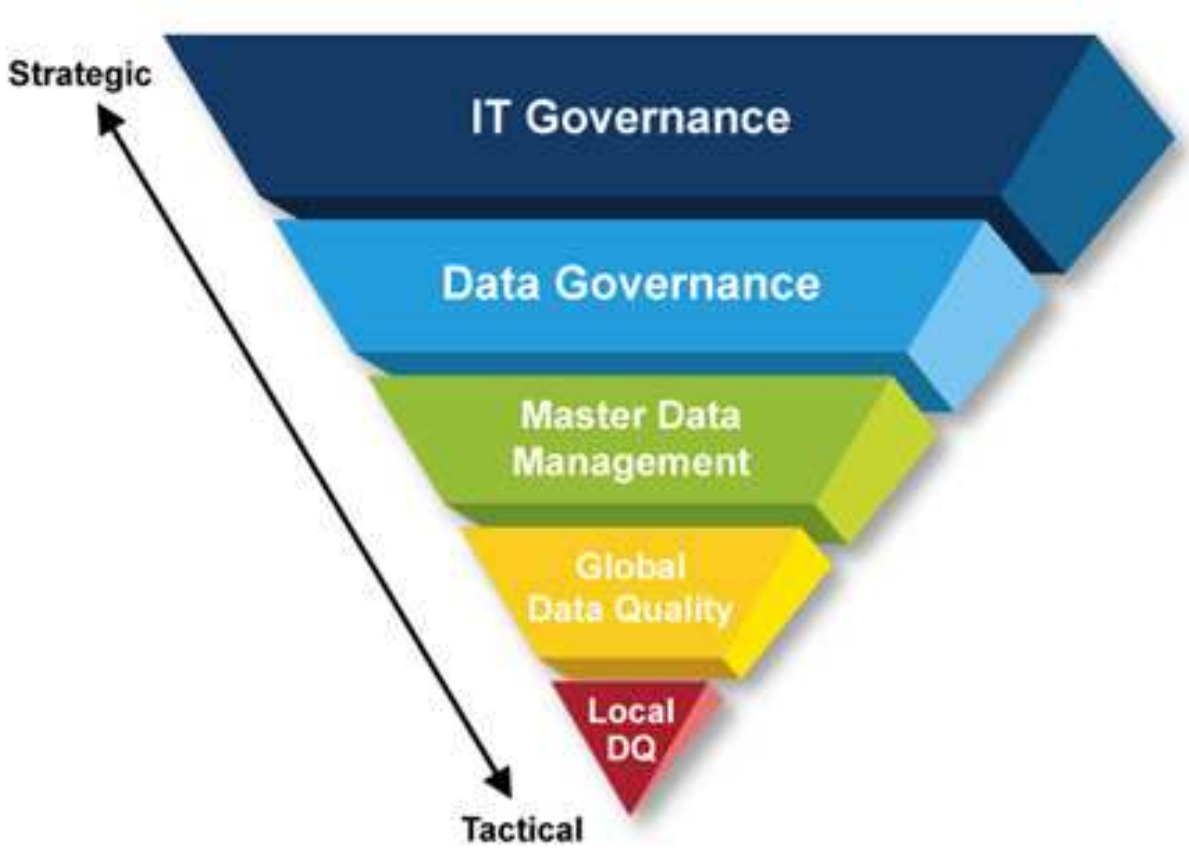
- ***Data Governance* (DG) é um conjunto de processos que asseguram que os dados, que são importantes activos, são formalmente geridos em toda a organização**
- DG é uma disciplina emergente com uma definição ainda em evolução
- **A disciplina contempla uma convergência da qualidade dos dados, gestão de dados, gestão de processos empresariais e gestão de riscos envolvendo a manipulação de dados de uma organização**

Data Governance



Reiner, R.K.; Turban, E.; Potter, R.E. (2007). *Introduction to Information Systems – Supporting and Transforming Business*, John Wiley.

Data Governance (Governança dos Dados)



Reiner, R.K.; Turban, E.; Potter, R.E. (2007). *Introduction to Information Systems – Supporting and Transforming Business*, John Wiley.