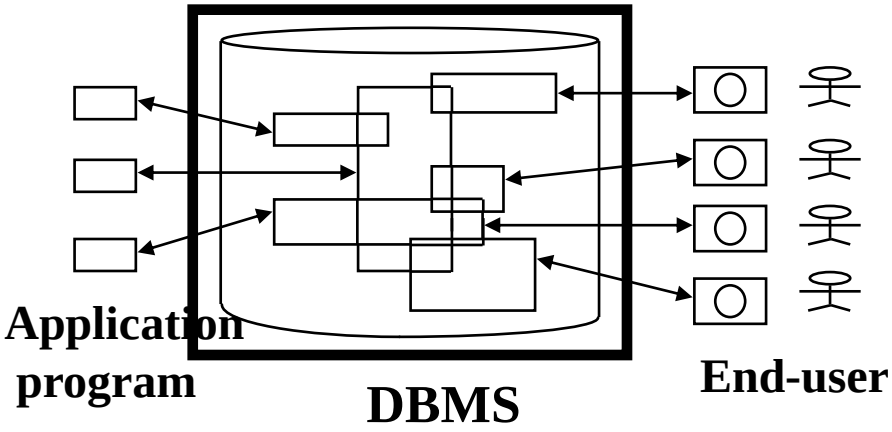
Unit 1

Introduction to DBMS

(Database Management Systems)



Outline of Unit 1

1.1 Information Systems

1.2 An Overview of a Database System

1.3 An Architecture for a Database System

1.4 Database Technology Trends

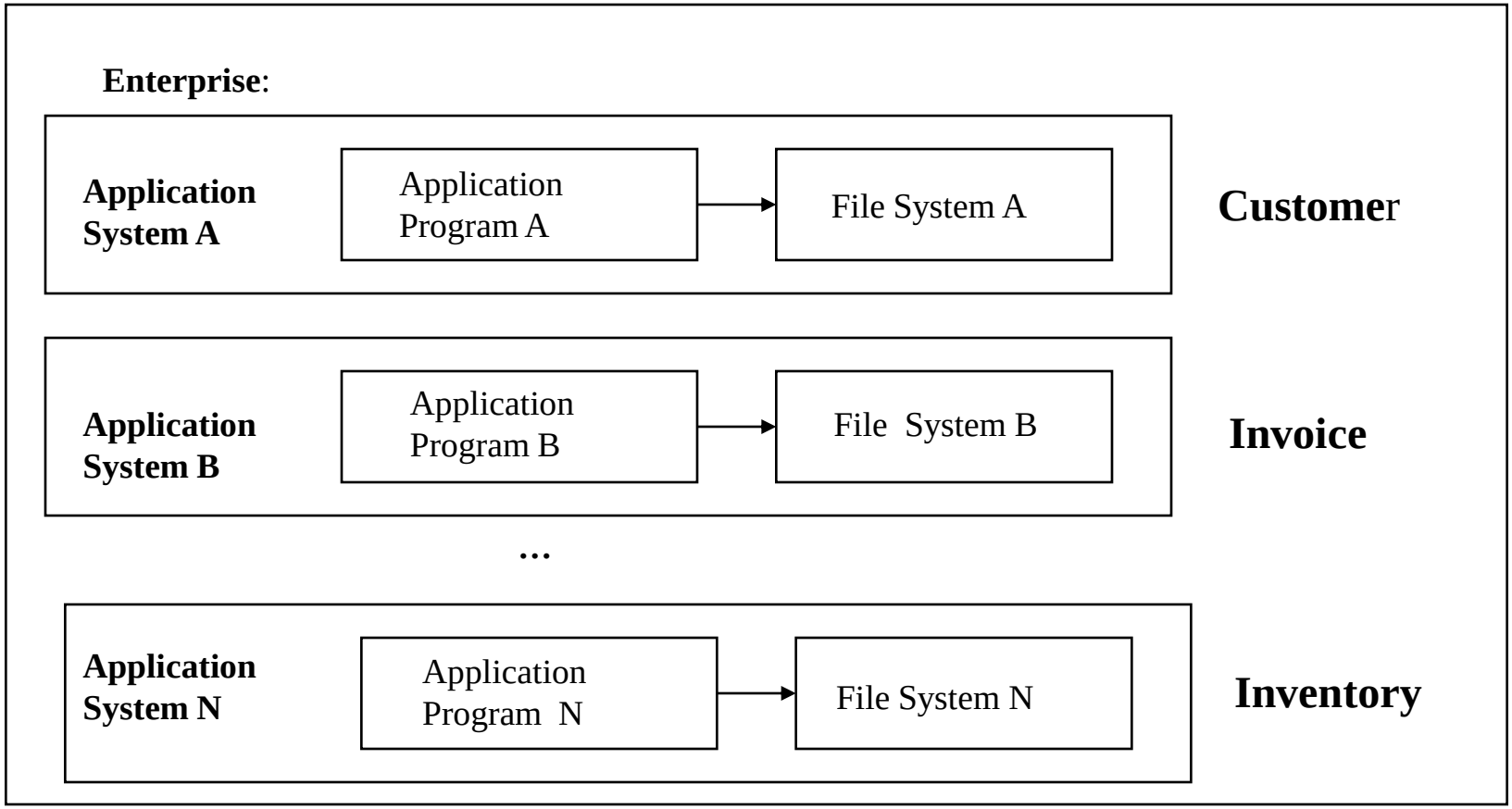
1.1 Information Systems

Stages of Information System

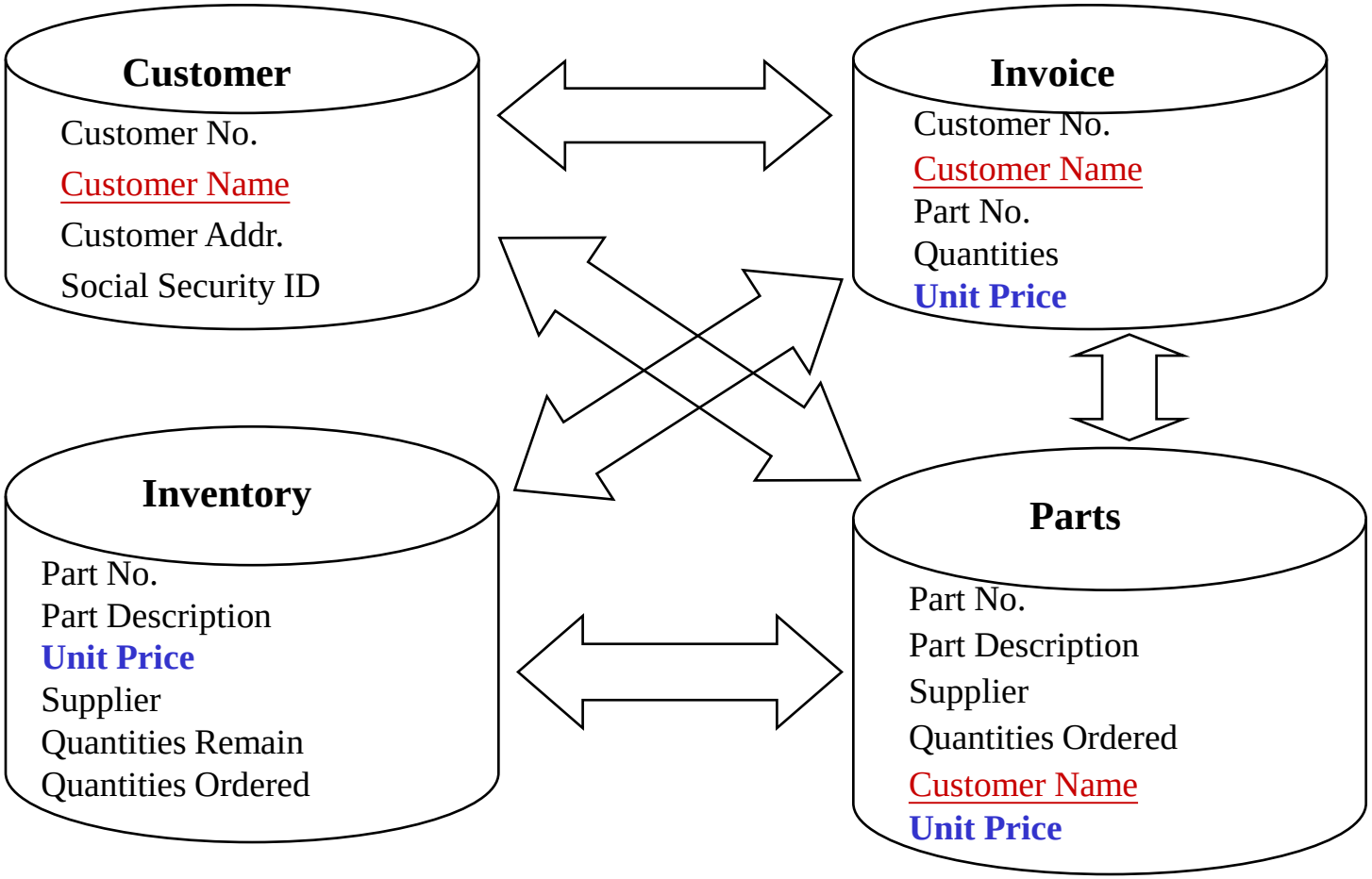
- Stage 0: Manual Information System
  - Records
  - Files
  - Index Cards
- Stage 1: Sequential Information Systems
  - Tapes
  - Files
  - slow, non-interactive, redundancy,... .
- Stage 2: File Based Information Systems
  - Disk (direct access)
  - application program has its own file ⇨ data dependence
  - data redundancy
- Stage 3: **DBMS based** Information Systems
  - Generalized data management software
  - Transaction processing

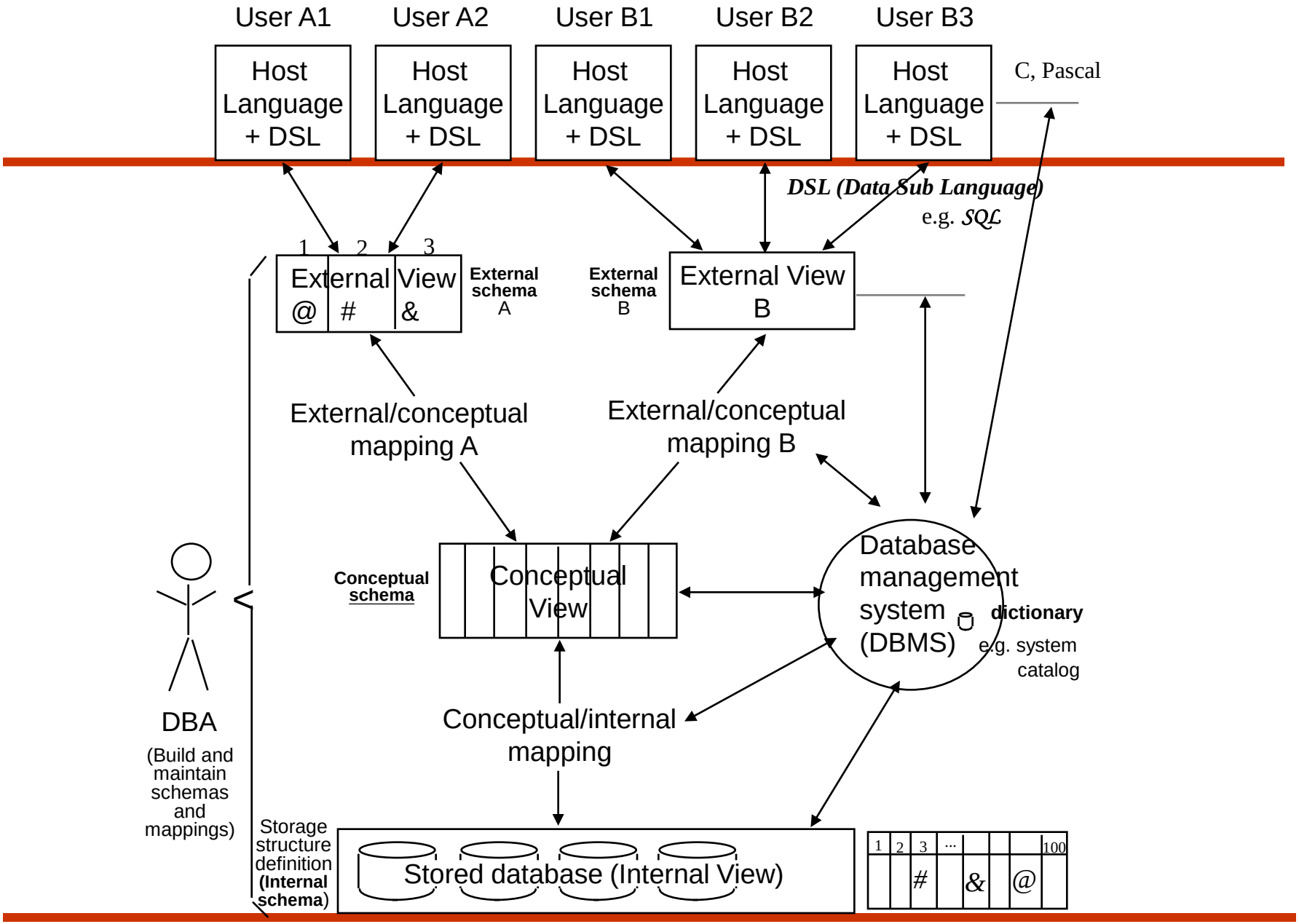
File Based Information Systems

- Conventional **Data Processing** techniques:



File Based Information Systems (cont.)

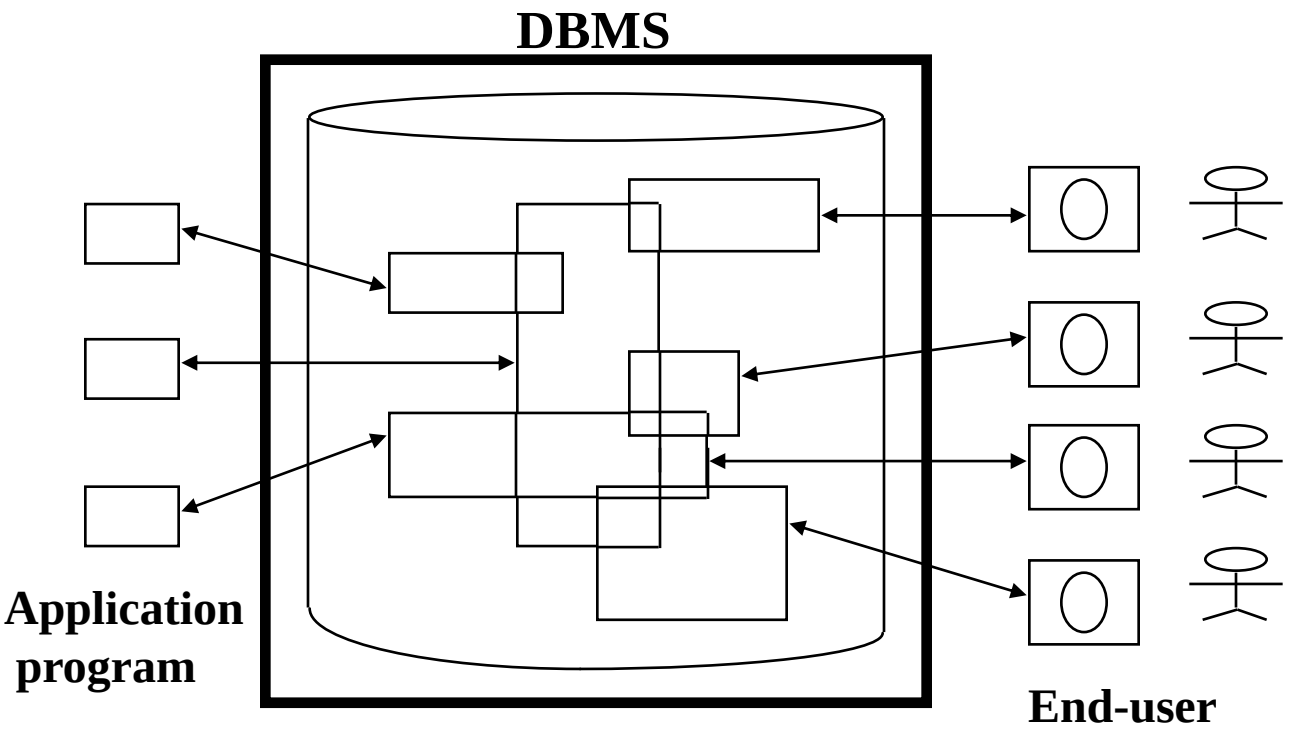



User A1
User A2
User B1
User B2
User B3
Host Language + DSL
Host Language + DSL
Host Language + DSL
Host Language + DSL
Host Language + DSL
C, Pascal
DSL (Data Sub Language)
e.g. SQL
1 2 3
External @ # View &
External schema A
External schema B
External View B
External/conceptual mapping A
External/conceptual mapping B
Conceptual schema
Conceptual View
Database management system (DBMS)
dictionary
e.g. system catalog
DBA
(Build and maintain schemas and mappings)
Conceptual/internal mapping
Storage structure definition (Internal schema)
Stored database (Internal View)
1 2 3 ... 100
& @

DBMS based Information Systems:

Basic Approach - Integration

- (1) Integration of Information
  - Description of the *integrated view* of data is the "*Conceptual Schema*" of the database



DBMS based Information Systems:

Basic Approach – Simple views and High level language

- (2) Provide simple views (External Schema) and high level language (e.g. SQL) for users to manipulate (handle) data
  - High level language: e.g. **SQL** (Structured Query Language)

    <e.g.>:
    ```
    SELECT SNAME
    FROM S
    WHERE S#= 'S4';
    ```
  - Description of user's view of data is the "external schema" or "subschema" or "view".
  - High-level languages (Query Language): **SQL**

    (1) Data Definition Language:
    define format

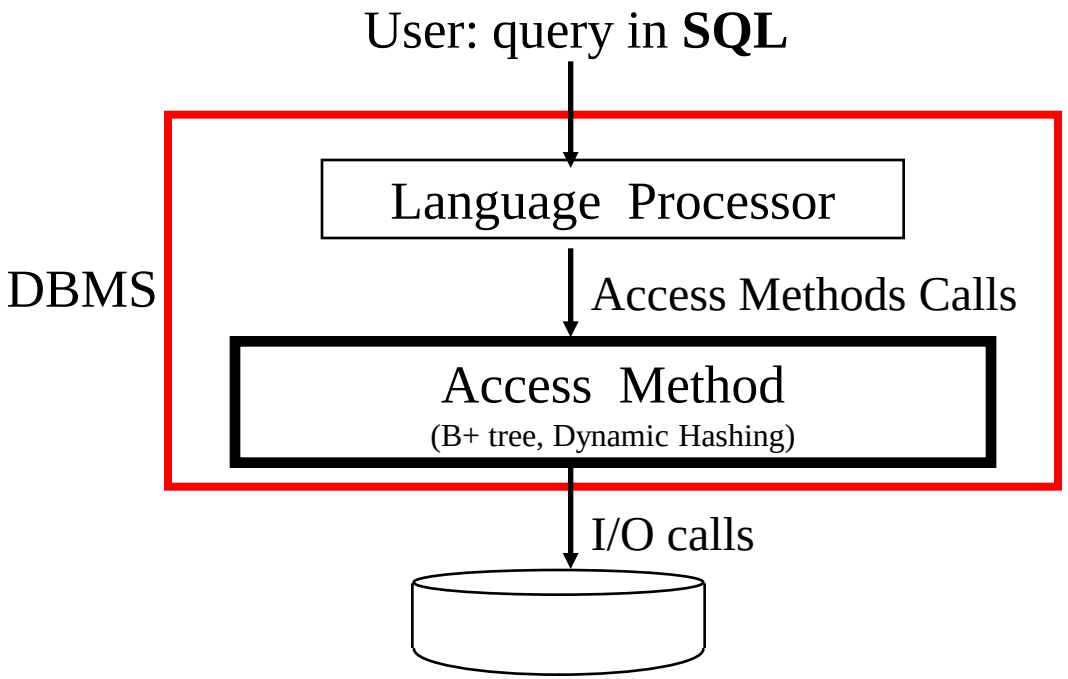
    (2) Data Manipulation Language:
    retrieve, insert, delete, update
  - Emphasize: *EASE OF USE !!*

S

S#	name		

DBMS based Information Systems:
Basic Approach - Storage/Access Method

- (3) Efficient Storage/Access Techniques:
  - implemented once rather than duplicated in all application programs.

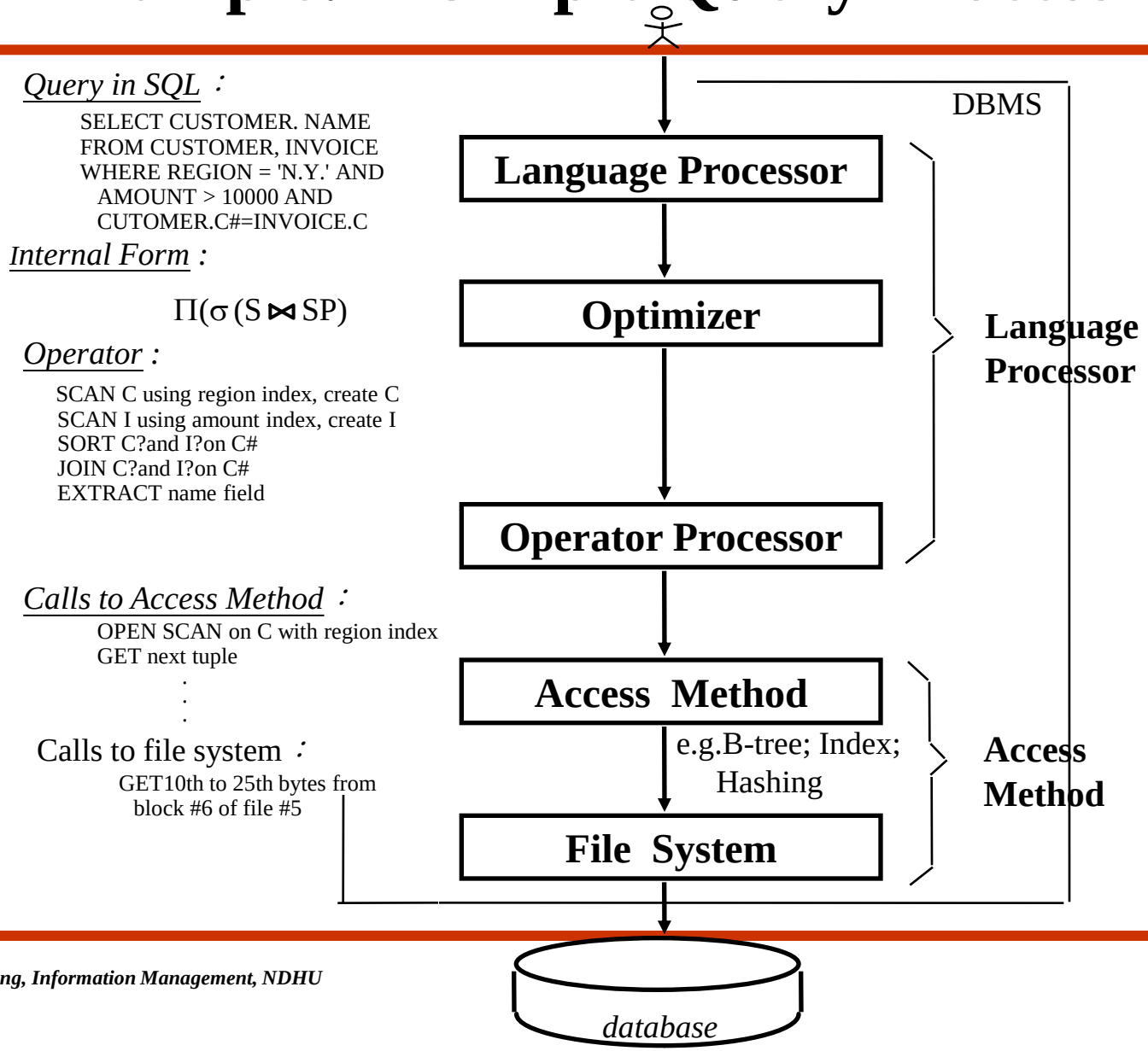


DBMS based Information Systems:

Basic Approach - Transaction Management

- (4) Provide Transaction Management:
  - Concurrency Control
  - Recovery
  - Security
  - .
    .
    .

Example: A Simple Query Processing



1.2 An Overview of a Database System

An Example

- **The Wine Cellar Database:**

Cellar:

Bin	Wine	Producer	Year	Bottle	Ready	Comments
2	Chardonnay	Buena Vista	85		83	1
3	Chardonnay	Louis Martini	81		5	84
6	Chardonnay	Chappellet	82		4	85 Thanksgiving
12	Jo. Riesling	Buena Vista	82		1	83 Late Harvest
16	Jo. Riesling	Sattui	82		1	83 Very dry
43	Cab. Sauvignon	Robt. Mondavi	77	12	87	
50	Pinot Noir	Mirassou	77		3	85 Harvest
51	Pinot Noir	Ch. St. Jean	78		2	86

- **Retrieval:**
  - DML (Data Manipulation Language):

```
SELECT   Wine, Bin, Producer
FROM     Cellar
WHERE    Ready = 85;
```

  - Result:

Wine	Bin	Producer
Chardonnay	2	Buena Vista
Chardonnay	6	Chappellet
Pinot Noir	50	Mirassou

An Example (cont.)

- **Deletion:**
  - DML: **DELETE FROM Cellar**
    WHERE Ready < 86;
  - Result:

Bin	Wine	Producer	Year	Bottle	Ready	Comments
43	Cab. Sauvignon	Robt. Mondavi	77	12	87	
51	Pinot Noir	Ch. St. Jean	78	2	86	

- **Insertion:**
  - DML: **INSERT INTO Cellar**
    VALUES (53, 'Pinot Noir', 'Franciscan', 79, 1, 86, 'for Joan');
  - Result:

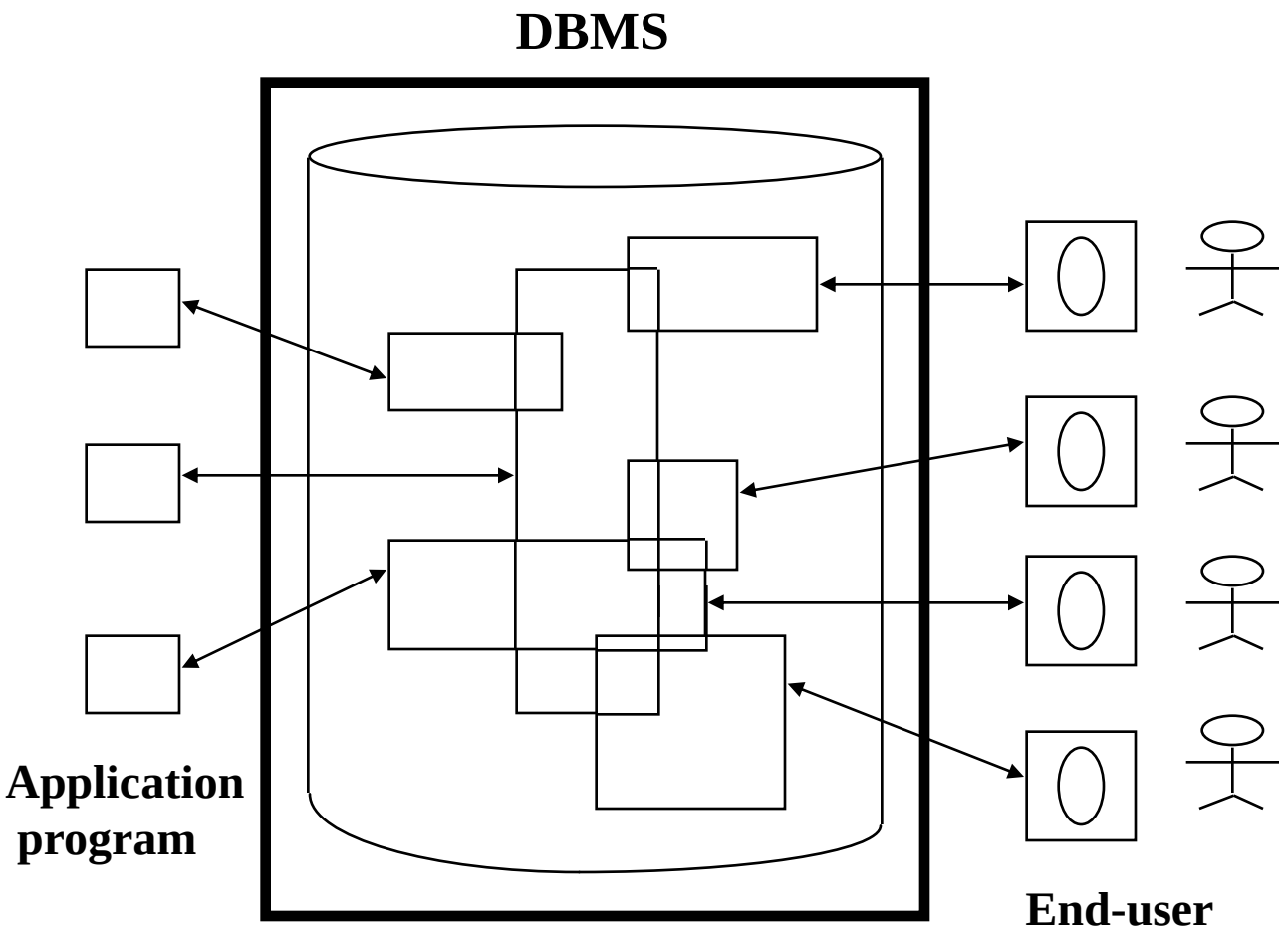
Bin	Wine	Producer	Year	Bottle	Ready	Comments
43	Cab. Sauvignon	Robt. Mondavi	77	12	87	
51	Pinot Noir	Ch. St. Jean	78	2	86	
53	*Pinot Noir*	*Franciscan*	*79*	*1*	*86*	*for Joan*

An Example (cont.)

- **Update**
  - DML: **UPDATE Cellar**
    SET Bottles = 4
    WHERE Bin = 51;
  - Result:

Bin	Wine	Producer	Year	Bottle	Ready	Comments
43	Cab. Sauvignon	Robt. Mondavi	77	12	87	
51	*Pinot Noir*	*Ch. St. Jean*	*78*	*4*	*86*	
53	Pinot Noir	Franciscan	86	1	79	for Joan

What is a Database System?



What is a Database System? (cont.)

- Major components of a database system:
  - Data: integrated and shared.
  - Hardware: disk, CPU, Main Memory, ...
  - Software: DBMS
  - Users:
    1. Application programmers
    2. End users
    3. ***Database administrator (DBA)***
       - Defining external schema
       - Defining conceptual schema
       - Defining internal schema
       - Liaison with users
       - Defining security and integrity checks
       - Defining backup and recovery procedures
       - Monitoring performance and changing requirements

Why Database ?

- Redundancy can be reduced
- Inconsistency can be avoided
- The data can be shared
- Standards can be enforced
- Security restrictions can be applied
- Integrity can be maintained
- Provision of *data independence*

objective !

Data Independence

- Application Program
  ➔ Data Structure
- Immunity of application to change in storage structure and access strategy.

Data Dependence *vs.* Data Independence

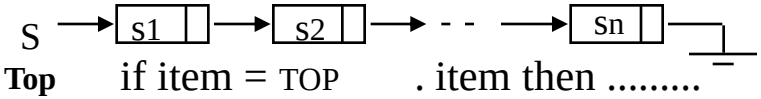
- Data Dependent

  e.g. SELECT CITY
  FROM S
  WHERE ITEM = 'X';

  - **Linked list:** TOP

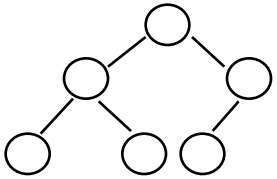
S

S#			
S_1			
S_2			
:			
Sn			



Top if item = TOP . item then .....

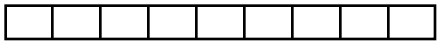
  - **Tree:**



.

if item < root.data then root := root .left .....

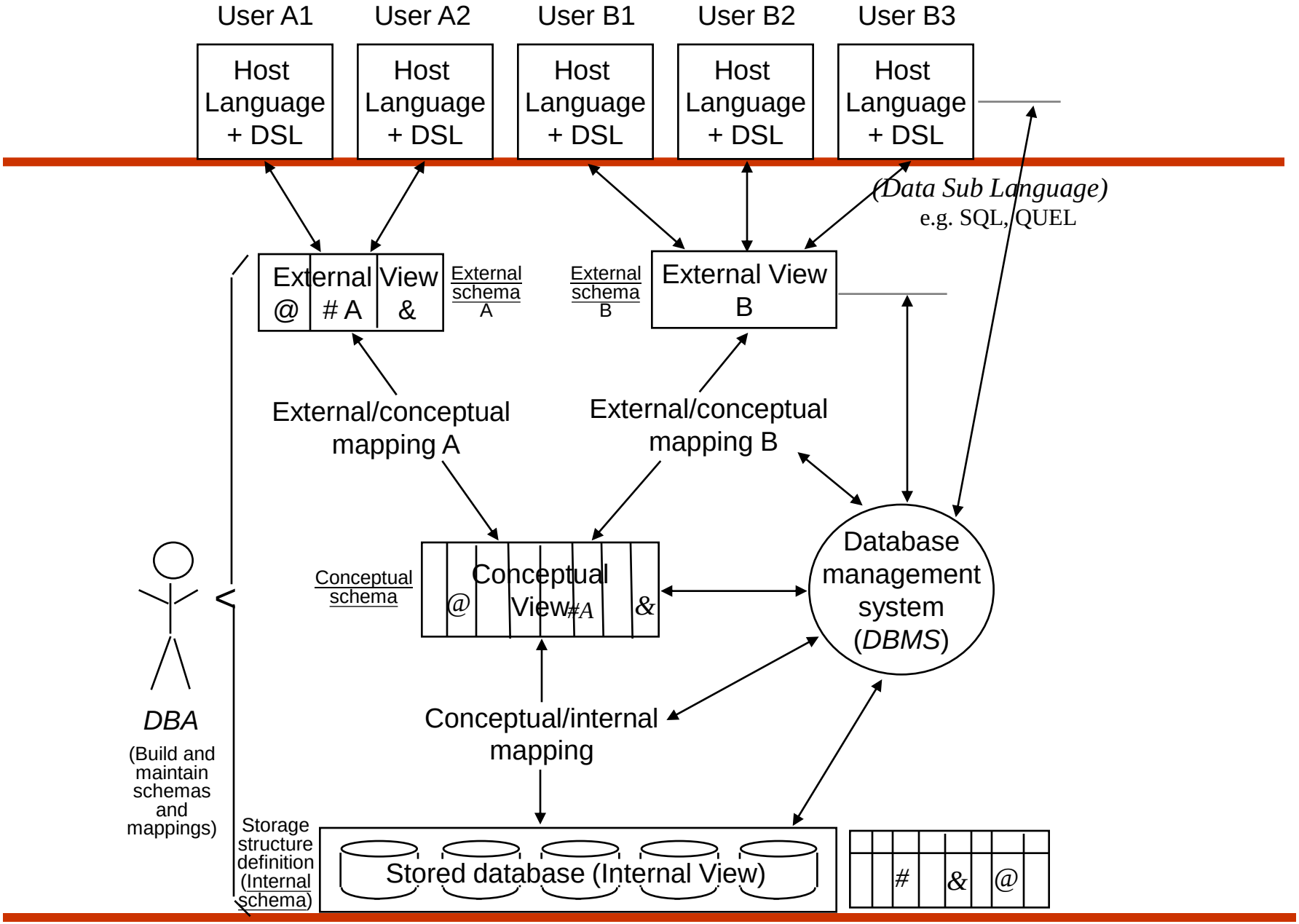
  - **Array:**



if A[I] = item then .....

  - ***Storage structure changed*** ➔ ***program changed***

1.3 An Architecture for a Database System


User A1
User A2
User B1
User B2
User B3
Host Language + DSL
Host Language + DSL
Host Language + DSL
Host Language + DSL
Host Language + DSL
(Data Sub Language)
e.g. SQL, QUEL
External View @ # A &
External schema A
External schema B
External View B
External/conceptual mapping A
External/conceptual mapping B
Conceptual schema
Conceptual View @ #A &
Database management system (DBMS)
DBA
(Build and maintain schemas and mappings)
Conceptual/internal mapping
Storage structure definition (Internal schema)
Stored database (Internal View)
& @

An Example of the Three Levels

- **Internal level:**

```
STORED_EMP    length = 18
    PREFIX        TYPE = BYTE(6), OFFSET = 0, INDEX = EMPX
    EMP#          TYPE = BYTE(6), OFFSET=0,
    DEPT#         TYPE = BYTE(4),  OFFSET = 12
    PAY           TYPE = FULLWORD, OFFSET = 16
```

- **Conceptual level:**

```
EMPLOYEE
    EMPLOYEE_NUMBER          CHARACTER(6)
    DEPARTMENT_NUMBER        CHARACTER(4)
    SALARY                   NUMERIC(5)
```

- **External level:**
  - (PL /I)

```
DCL  1  EMP
        2    EMP#     CHAR(6)
        2    SAL      FIXED BIN(31)
```

  - (COBOL)

```
01   EMPC
     02   EMPNO    PIC X(4)
     02   DEPTNO   PIC X(4)
```

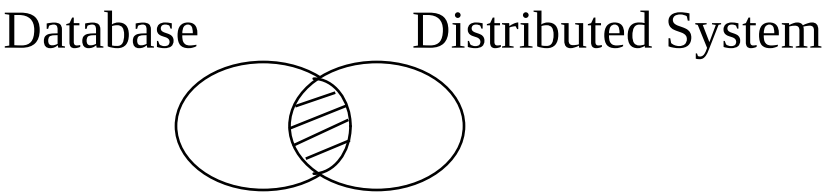
Functions of the DBMS

- Data Definition Language (DDL)
- Data Manipulation Language (DML)
- Data Security and Integrity
- Data Recovery and Concurrency
- Data Dictionary
- Performance

1.4 Database Technology Trends

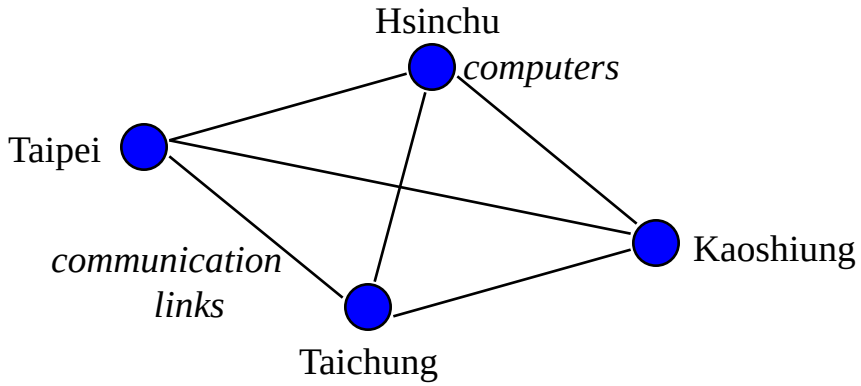
	1960s to Mid-1970s	1970s to Mid-1980s	Late 1980s	Future
Data Model	Network Hierarchical	Relational	Semantic Object-oriented Logic	Merging data models, knowledge representation, and programming languages
Database Hardware	Mainframes	Mainframes Minis PCs	Faster PCs Workstations Database machines	Parallel processing Optical memories
User Interface	None Forms	Query languages - SQL, QUEL	Graphics Menus Query-by-forms	Natural language Speech input
Program Interface	Procedural	Embedded query language	4GL Logic programming	Integrated database and programming language
Presentation and display processing	Reports Processing data	Report generators Information and transaction processing	Business graphics Image output Knowledge processing	Generalized display managers Distributed knowledge processing

Distributed Databases



- Distributed database is a database that is not stored in its entirety at a single physical location, but rather is spread across a network of computer.

< e.g.>



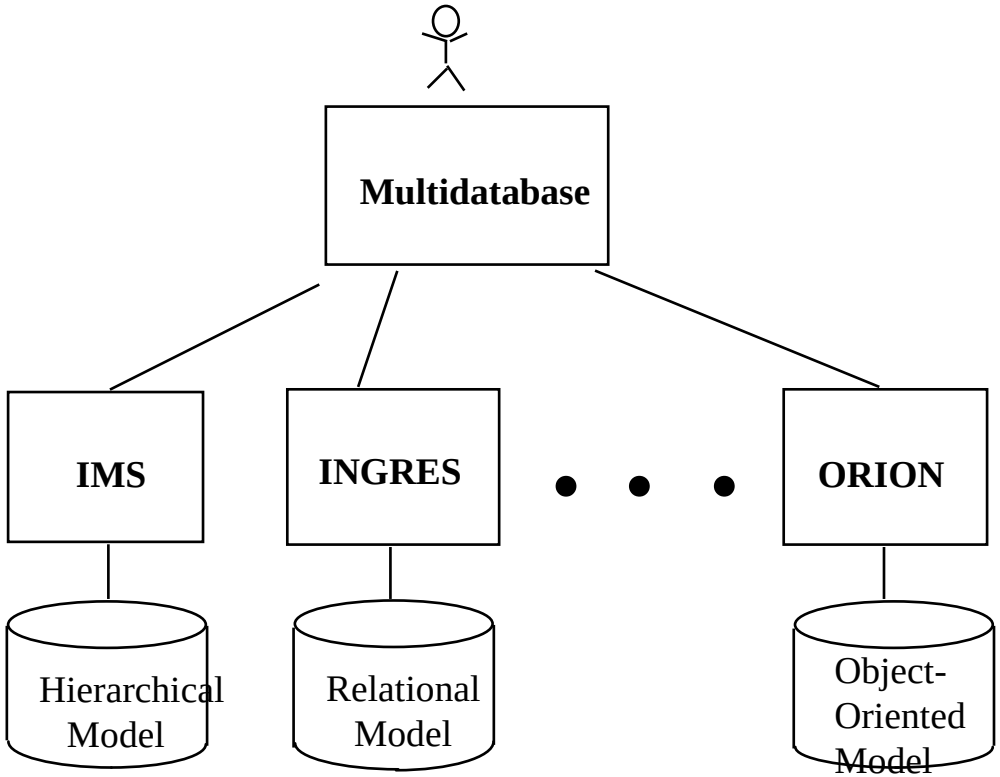
Distributed Databases (cont.)

- Advantages:
  - efficiency of local processing
  - data sharing
- Disadvantages:
  - communication overhead
  - implementation difficulties
- Reference:

  S. Ceri and G. Pelagatti

  "Distributed Databases: principles and systems"

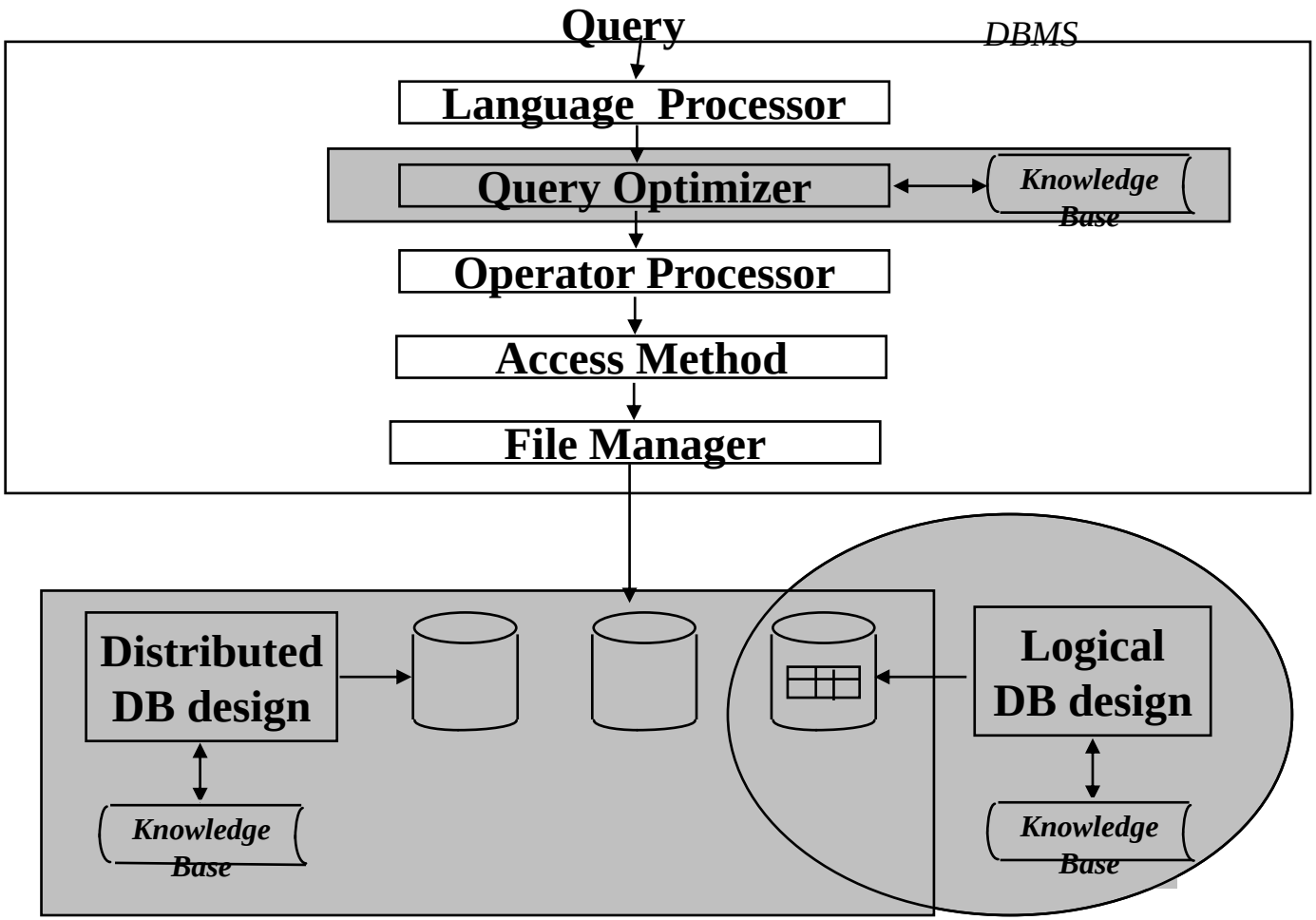
Multi-Database/Heterogeneous Database



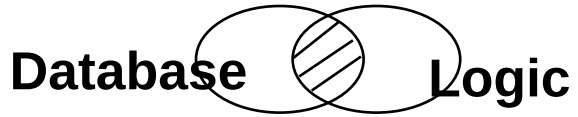
- **semantic inconsistency**
- **data incompleteness**
- **global schema**

DB + AI


Database
AI


Query
DBMS
Language Processor
Query Optimizer
Knowledge Base
Operator Processor
Access Method
File Manager
Distributed DB design
Knowledge Base
Logical DB design
Knowledge Base

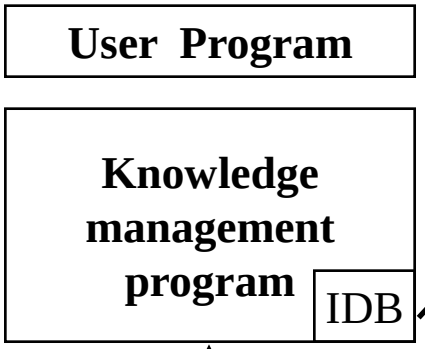
KBMS



- A Combined Model :

  Logic Programming + Relational DB

- Three layers :



Query :

? :- ancestor (taro, Y)

? :- grandfather (?, c)

IDB:

ancestor(X,Y):- parent(X,Y)

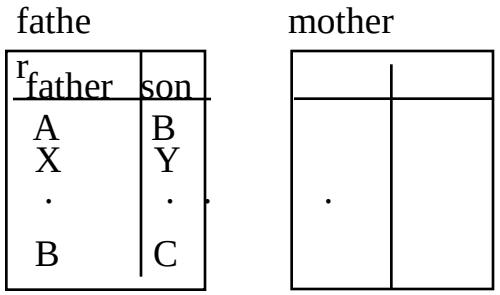
ancestor(X,Y) :- parent(X,Z), ancestor(Z,Y)

parent(X,Y):-edb(father(X,Y))

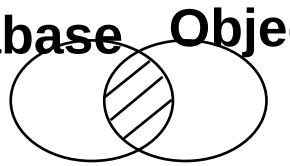
parent(X,Y):-edb(mother(X,Y))

grandfather(X,Z):- father(X,Y) ^ father(Y,Z)

EDB:



OODB



- A typical Document : MEMO [Woelk86, SIGMOD]

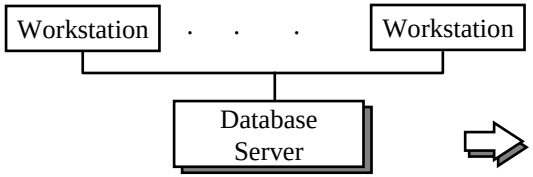
MCC

To: W. Kim
From: D. Woolk
Date: September 18, 1992
Subject: Workstations

speaker voice message associated

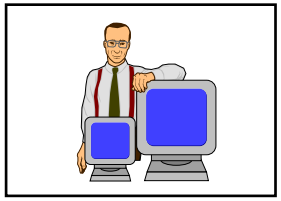
In the computer center of National Chiao-Tung University, there are a lot of workstations. There are HP RS serials, SUNs, Apollo, and so on. The students in NCTU learn to use workstation since they are fresh-men. The configuration of the workstations follows:

text



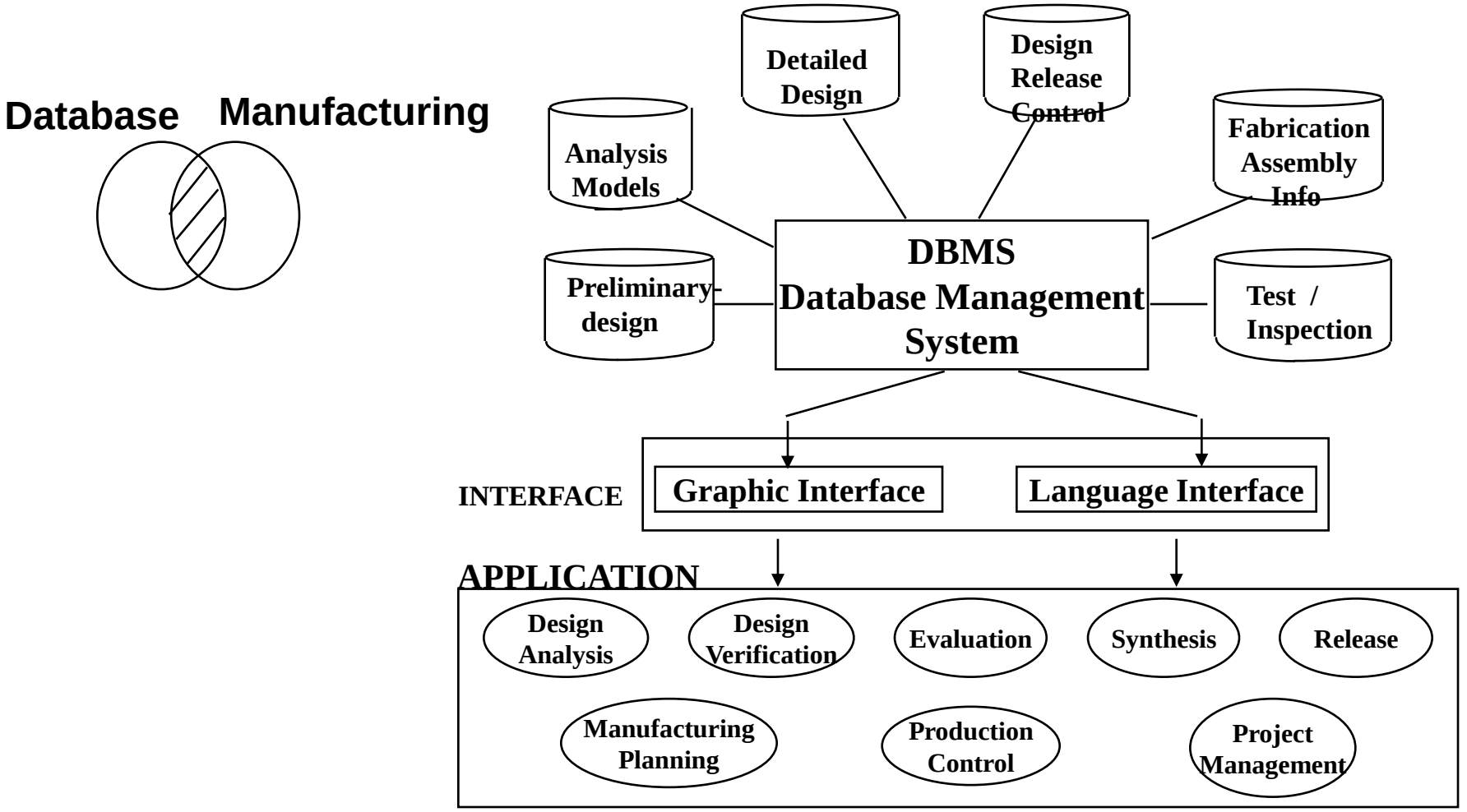
graphics

In the course introduction to Computer Science? students do their homework's on workstations.

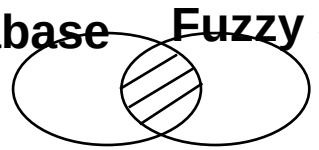


image

Use of a Database Management System in Design and Application



Fuzzy Database



- Fuzzy Query

```
<e.g.>      SELECT      STUDENT.NAME
            FROM        STUDENT
            WHERE       SEX = M
                        AND  HEIGH = TALLER
                        AND  WEIGH = SLIMMER
```

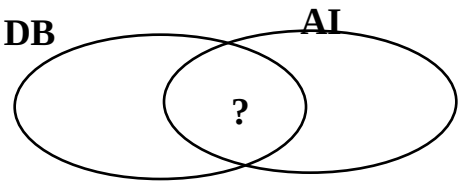
STUDENT:

NAME	SEX	HEIGHT	WEIGHT	IQ
Mary	F	158	55	High
Linda	F	165	55	Medium
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

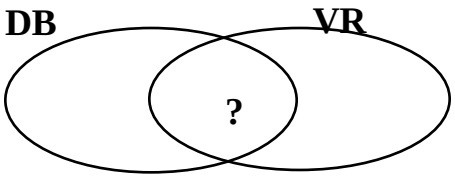
```
<e.g.>          SELECT    STUDENT.NAME
                FROM      STUDENT
                WHERE     IQ >= 130
```

More?

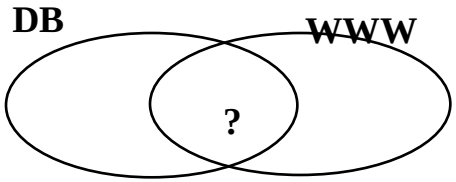
1993



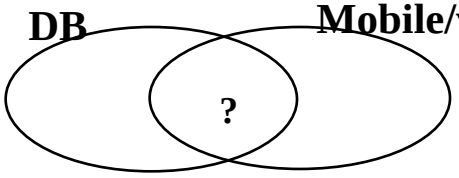
1995



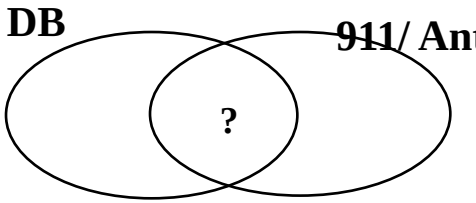
1997



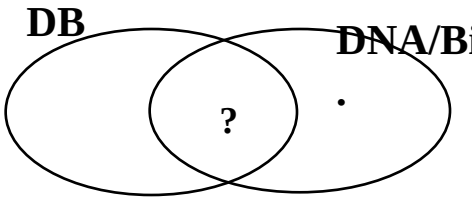
1999



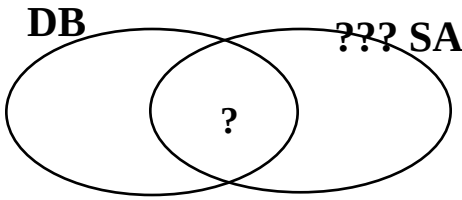
2001



2003



2004



2006