Introduction to Database Development Strategies
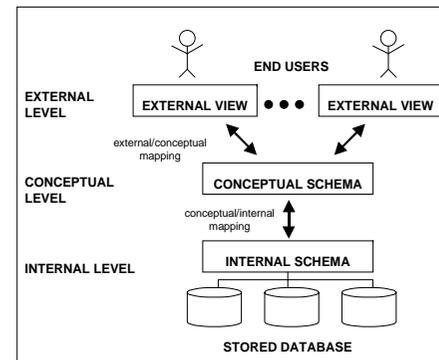

Database Approach

- Three important characteristics:
  - Program-data independence,
  - Support of multiple user views, and
  - Use of catalogue to store the database description (schema)

The Three-Schema Architecture

- The goal of the three-schema architecture is to separate the user applications and the physical database.
- In this architecture, schemas can be defined at the following three levels:
  - The Internal Level
  - The Conceptual Level
  - The External or View Level

The Three-Schema Architecture

END USERS

EXTERNAL LEVEL — EXTERNAL VIEW ● ● ● EXTERNAL VIEW

external/conceptual mapping

CONCEPTUAL LEVEL — CONCEPTUAL SCHEMA

conceptual/internal mapping

INTERNAL LEVEL — INTERNAL SCHEMA

STORED DATABASE

Internal Level

- Has an internal schema
- Describe the **Physical Storage Structure** of the database.
- The internal schema uses a physical data model describe the complete details of data storage and access paths for the database.

Conceptual Level

- Has a conceptual schema
- Describe the **Structure** of the whole database for a community of users.
- It is a global description of the database
- Hides the details of physical storage structures and concentrates on describing entities, data types, relationships and constraints.

External or View Level

- Includes a number of external schemas or user views.
- Each external schema describes the **Database View** of one group of database users.
- Each view typically describes the database view of one group of database users group is interested in and hides the rest of the database from that user group.

The Three-Schema Architecture

- Most DBMS do not separate the three levels completely.
  - For example, many DBMS will include some physical level details in the conceptual schema.
  - However, other DBMS fit in the general framework of the three-schema architecture.
  - In most DBMS that support user views, external schemas are specified in same data model at the conceptual level.

Mapping

- Three schemas are only descriptions of data
- The only data that actually exists is at the physical level.
- Each user group refers only to its own external schema.
- The DBMS must transform specified on an external schema for processing on the stored database.
- If the request is a database retrieval, the database extracted from the stored database must be reformatted to match the user's external view before it is presented to the user.
- The process of transforming requests and results between levels are called **Mapping**.

Data Independence

- The three-schema architecture can be used to explain the concept of data independence, which can be defined as the capacity to change the schema at one level of a database system without having to change the schema at the next higher level.
- There are two types of data independence :
  - Logical Data Independence
  - Physical Data Independence

Logical Data Independence

- The capacity to change the conceptual schema without having to change external schemas or application programs.
  - For example, we may change the conceptual schema to expand the database by removing a record type or data item.
  - The external schemas that refer only to the remaining data should not be affected.
  - Only the view definitions and the mapping need to be changed in a DBMS that supports logical data independence.

Physical Data Independence

- The capacity to change the internal schema without having to change the conceptual (or external) schemas.
  - For example, when create additional access structures to improve the performance of retrieval or update.
  - If the same data as before remains in the database, we should not have to change the conceptual schema.

Comparison of Data Independence

| Level of data independence	Examples of changes
Logical Data item format	Data item, type, length, or representation
Data item usage	How a data item is derived, used, edited, or protected
Logical record structure	How data items are grouped into logical records
Logical data structure	Overall logical structure or conceptual model
Physical Physical data organization	How the data are organized into stored records
Access method	What search techniques and access strategies are used
Physical data Location Storage device	Where data are located on storage devices Characteristics of the physical storage devices used

13

Advantages of DBMS

- The mapping information is stored in the repository.
- The DBMS refers to this mapping information to accomplish mapping between inter-levels.
- Data independence is accomplished because when we changed the schema at some level, we can leave the schema at the next higher level unchanged and change only the mapping between the two levels.
- Application programs referring to the higher-level schema need not to be changed.

DB212 © Peter Lo 2004

14

Database Development Process
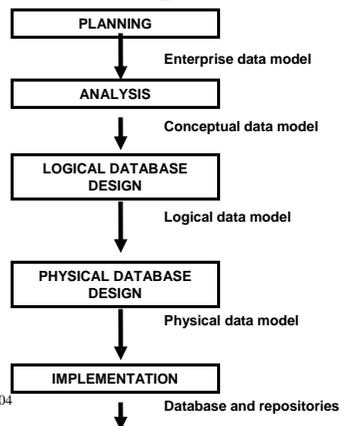
DB212 © Peter Lo 2004

15

Planning

- To align information technology with the business strategies of an organization.
- Achieve a competitive advantage when they are able to develop sound information systems plans
- Critical success factors and problems areas have to be identified
- At the end of planning phase, an enterprise model is developed.
  - ◆ Breaking the functions of an organization down into progressively lower levels of details
  - ◆ Identifying the entity types and the relationship between them (a draft ERD)

DB212 © Peter Lo 2004

16

Analysis

- Develop detailed specifications for the information systems required to support the organization.
  - Study of the current business situation;
  - Determination of the new system requirements.
- At the end of this phase, a conceptual data model (a detailed ERD) will be built.
- This conceptual data model includes the relevant entities, relationships and attributes as well as the business rules and constraints that define how the data are used.

Design

- Transform the conceptual data model to an implementation model that a particular DBMS can process with performance that is acceptable to all users throughout the organization.
- This database design includes **Logical Database Design** and **Physical Database Design**.

Logical Database Design

- The process of mapping conceptual data model to structures that are specific to the target DBMS.
  - E.g. If the target environment is a relational DBMS, then the conceptual data models (ERD) are mapped to normalized relations

Physical Database Design

- The process of mapping the database structures from logical design into physical storage structures such as files and tables, indexes, access methods.
  - Provide adequate performances for user applications in terms of response times, throughout rates etc.

Implementation

- Involves a series of steps leading to operational information systems that includes
  - Creating database definitions
  - Creating program code
  - Testing the systems
  - Developing operational procedures and documentation
  - Training and loading the databases

Review Questions

- Give the definitions for the following terms:
  - Conceptual schema
  - Internal schema
  - External-conceptual mapping
- Explain what is data dependency.
- What is the difference between logical data independence and physical data independence? Which is easier to accomplish?
- Discuss briefly what are the output of each database development phase.