



Chapter one

Introduction



What is a database?

- A database is a **collection** of shared information that exists over a long period of time, often many years.
- Database systems are designed to manage large data set in an organization.
- The data management involves both **definition** and the **manipulation** of the data.



Cont...

- Data management passes through the different levels of development along with the development in technology and services.
- The major three levels are;
 - 1. Manual Approach**
 - 2. Traditional File Based Approach**
 - 3. Database Approach**



I. Manual Approach

- In the manual approach, **data storage** and **retrieval** follows the **primitive** and **traditional** way of information handling where **cards** and **paper** are used for the purpose.
- **Files** for as **many** event and objects as the organization has are used to store information.
- Each of the files containing various kinds of information is labelled and stored in one or more **cabinets**



Cont...

- The **cabinets** could be kept in safe places for security purpose based on the sensitivity of the information contained in it.
- **Insertion** and **retrieval** is done by searching first for the right cabinet then for the right the file then the information.
- One could have an **indexing** system to facilitate access to the data



Limitations of the Manual approach

- Prone to error
- Difficult to update, retrieve, integrate
- You have the data but it is difficult to compile the information
- Limited to small size information
- Cross referencing is difficult



2. Traditional File Based Approach

- File based systems were an early attempt to computerize the manual filing system.
- This approach is the decentralized computerized data handling method.
- A collection of **application programs** perform services for the end-users.
 - In such systems, every application program that provides service to end users define and manage its own data

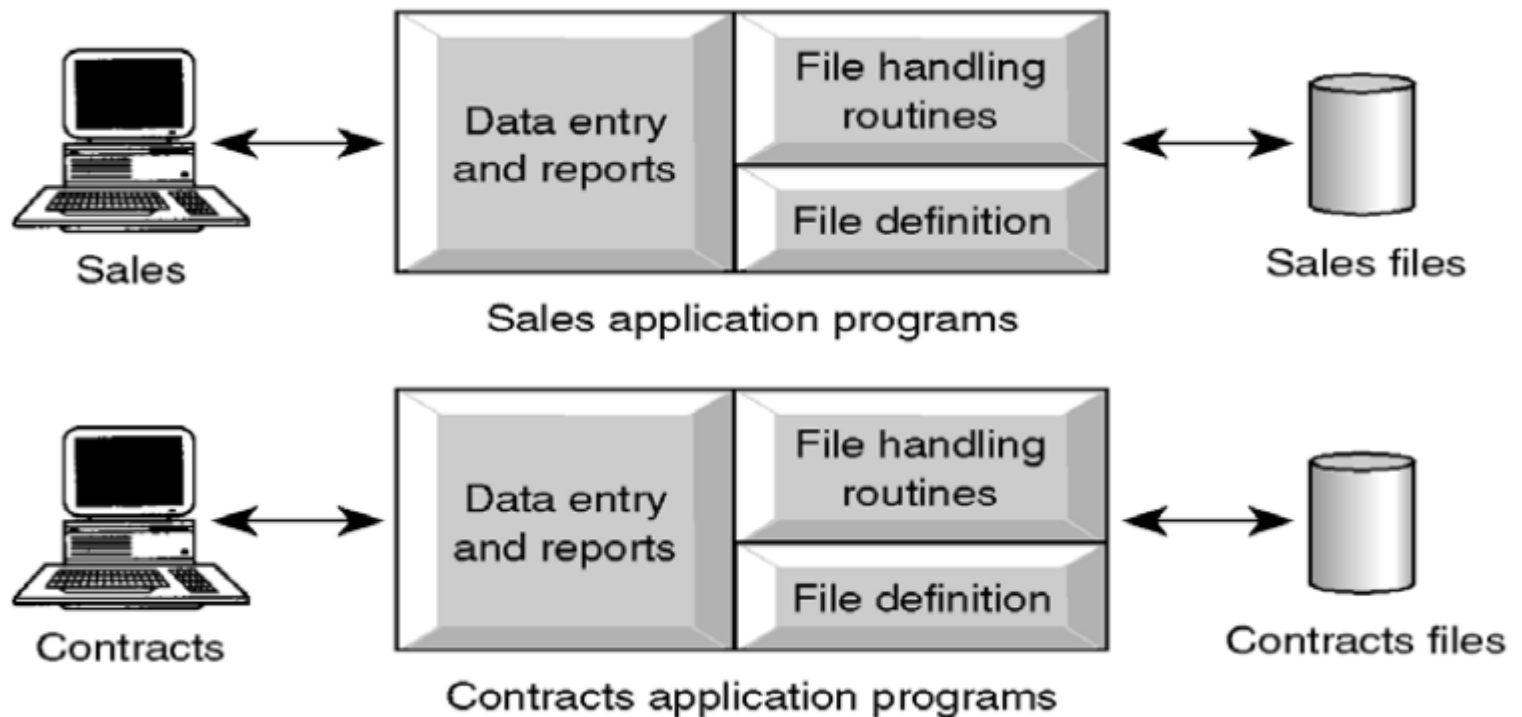


Cont...

- Such systems have **number of programs** for **each** of the different applications in the organization.
- Since every application defines and manages its own data, the system is subjected to serious **data duplication problem**.

Cont...

- File, in traditional file based approach, is a collection of records which contains logically related data.





Limitations of the Traditional File Based approach

- Separation or **Isolation** of **Data**: Available information in one application may not be known.
- Limited data sharing
- Lengthy development and maintenance time
- Duplication or redundancy of data
- Data dependency on the application
- Incompatible file formats between different applications and programs creating inconsistency.
- Fixed query processing which is defined during application development



Cont...

- The limitations for the traditional file based data handling approach arise from two basic reasons.
 1. **Definition** of the data is **embedded** in the **application program** which makes it difficult to modify the database definition easily.
 2. **No control over the access and manipulation** of the data beyond that imposed by the application programs.

Cont...

- The most significant problem experienced by the traditional file based approach of data handling is the “**update anomalies**”. 3 types
 1. **Modification Anomalies:** a problem experienced when one or more data value is modified on one application program but not on others containing the same data set.
 2. **Deletion Anomalies:** a problem encountered where one record set is deleted from one application but remain untouched in other application programs.
 3. **Insertion Anomalies:** a problem encountered where one can not decide whether the data to be inserted is valid and consistent with other similar data set.



3. Database Approach

- Following a famous paper written by **Ted Codd** in 1970, database systems changed significantly.
- Codd proposed that database systems should present the user with a view of data organized as tables called **relations**.
- Behind the scenes, there might be a **complex data structure** that allowed rapid response to a variety of queries.



Cont...

Thus in Database Approach:

- Database is just a computerized record keeping system or a kind of electronic filing cabinet.
- Database is a repository for collection of **computerized data files**.
- Database is a shared collection of logically related data designed to meet the information needs of an organization.



Cont...

- Database is a collection of logically related data where these logically related data comprises **entities, attributes, relationships, and business rules** of an organization's information.
- In addition to containing data required by an organization, database also contains a description of the data which called as “**Metadata**” or “**Data Dictionary**” or “**Systems Catalogue**” or “**Data about Data**”.

Cont...

- Since a database contains information about the data (metadata), it is called a **self descriptive** collection on integrated records.
- The purpose of a database is to **store information** and to **allow** users to **retrieve** and update that information on demand.
- Database is designed once and used simultaneously by many users.
- Unlike the traditional file based approach in database approach there is **program data independence**.



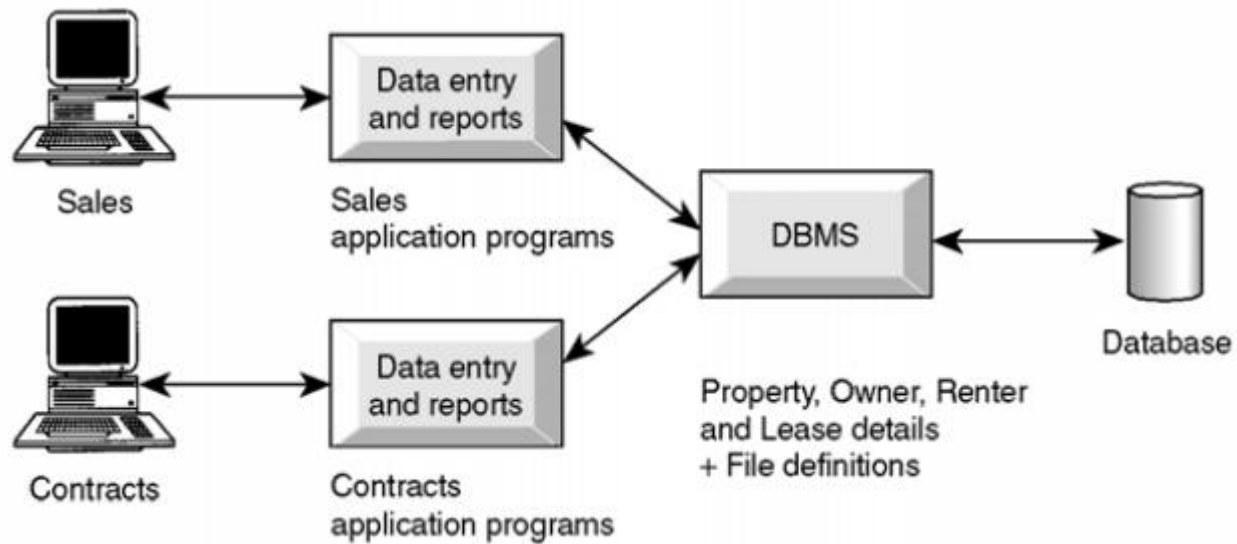
Benefits of the database approach

- Data can be shared
- Improved accessibility of data
- Redundancy can be reduced
- Quality data can be maintained:
- Inconsistency can be avoided:
- Transaction support can be provided:
- Speed
- Less labour
- Centralized information control:

Cont...

- Integrity can be maintained
- Security majors can be enforced
- Improved decision support
- Standards can be enforced
- Compactness
- Speed

Cont...





Limitations and risk of Database Approach

- Introduction of **new professional** and specialized personnel.
- Complexity in **designing** and managing data
- Extra cost and risk during conversion from the old to the new system
- High cost incurred to develop and maintain
- Complex **backup** and **recover** services from the users perspective
- Reduced **performance** due to centralization
- High impact on the system when failure occur



Database Management System (DBMS)

- DBMS is a Software package used for providing EFFICIENT, CONVENIENT and SAFE MULTI-USER storage of and access to MASSIVE amounts of PERSISTENT data.
- A DBMS also provides a systematic method for **creating, updating, storing, retrieving** data in a database.
- DBMS also provides the service of controlling data access, enforcing data integrity, managing concurrency control, and recovery.

Cont...

- DBMS should at least have the following services to provide to the user.
 - **Data storage, retrieval and update in the database**
 - **A user accessible catalogue**
 - **Transaction support service**
 - **Concurrency Control Services**
 - **Recovery Services**
 - **Authorization Services (Security)**
 - **Support for Data Communication**



Cont...

- **Integrity Services:**
- Services to promote **data independency** between the data and the application
- **Utility services:** sets of utility service facilities like
 - Importing data
 - Statistical analysis support
 - Index reorganization
 - Garbage collection



DBMS and Components of DBMS Environment

- **Data Definition Language (DDL):**
 - Language used to define each data element required by the organization.
 - Commands for setting up schema of database
 - Common **DDL** statements are CREATE, ALTER, and DROP
- **Data Manipulation Language (DML):**
 - Language used by end-users and programmers to **store, retrieve**, and access the data e.g. SQL
 - Also called "query language"
- **Data Dictionary:** tool used to store and organize information about the data

Cont...

The DBMS environment has five components.

1. **Hardware:**
2. **Software:**
3. **Data:** This is the most important component to the user of the database.
 - There are two types of data in a database approach that is **Operational** and **Metadata**.
4. **Procedure:** this is the rules and regulations on how to design and use a database.
5. **People:** people in the organization responsible to designing, implement, manage, administer and use of the database.

Database Development Life Cycle

The major steps in database design are;

1. **Planning:** that is identifying information gap in an organization and propose a database solution to solve the problem.
2. **Analysis:** that concentrates more on fact finding about the problem or the opportunity.
3. **Design:** in database designing more emphasis is given to this phase.
 - The phase is further divided into three sub-phases.

Cont...

- A. Conceptual Design:** concise description of the **data, data type, relationship** between data and constraints on the data.
 - There is no implementation or physical detail consideration.
 - Used to elicit and structure all information requirements
- B. Logical Design:** a higher level conceptual abstraction with selected specific data model to implement the data structure.
 - It is particular DBMS independent and with no other physical considerations.
- C. Physical Design:** physical implementation of the upper level design of the database with respect to **internal storage** and file structure of the database for the selected DBMS.



Cont...

4. Implementation: the testing and deployment of the designed database for use.

5. Operation and Support: administering and maintaining the operation of the database system and providing support to users.



Roles in Database Design and Use

I. DataBase Administrator (DBA)

- Responsible to oversee, control and manage the database resources
- Authorizing access to the database
- Coordinating and monitoring the use of the database
- Responsible for determining and acquiring hardware and software resources
- Accountable for problems like poor security, poor performance of the system
- Involves in all steps of database development

Cont...

- We can have further classifications of this role in big organizations having huge amount of data and user requirement.
 - I. **Data Administrator (DA):** is responsible on management of data resources.
 - Involves in database **planning, development, maintenance** of standards policies and procedures at the **conceptual** and **logical** design phases.
 - II. **DataBase Administrator (DBA):** is more technically oriented role.
 - Responsible for the **physical realization** of the database.
 - Involves in **physical design, implementation, security and integrity** control of the database.



Cont...

2. DataBase Designer (DBD)

- Identifies the data to be stored and choose the appropriate structures to represent and store the data.
- Should understand the user requirement and should choose how the user views the database.
- Involve on the design phase before the implementation of the database system.
- We have two distinctions of database designers, one involving in the **logical** and **conceptual** design and another involving in physical design.



Cont...

A. Logical and Conceptual DBD

- Identifies **data** (entity, attributes and relationship) relevant to the organization
- Identifies **constraints** on each data
- Understand **data** and **business rules** in the organization
- Sees the database independent of any data model at conceptual level and consider one specific data model at logical design phase.

Cont...

B. Physical DBD

- Take logical design specification as input and decide how it should be physically realized.
- Map the logical data model on the specified DBMS with respect to tables and integrity constraints. (DBMS dependent designing)
- Select specific storage structure and access path to the database
- Design security measures required on the database



Cont...

3. Application Programmer and Systems Analyst

- System analyst determines the user requirement and how the user wants to view the database.
- The application programmer implements these specifications as programs; code, test, debug, document and maintain the application program.
- Determines the interface on how to retrieve, insert, update and delete data in the database.

Cont...

4. End Users

1. Naïve Users:

- Sizable proportion of users
- Unaware of the DBMS
- Only access the database based on their access level and demand
- Use standard and pre-specified types of queries.

2. Sophisticated Users

- Are users familiar with the structure of the Database and facilities of the DBMS.
- Have complex requirements
- Have higher level queries
- Are most of the time engineers, scientists, business analysts, etc



Cont...

3. Casual Users

- Users who access the database occasionally.
- Need different information from the database each time.
- Use sophisticated database queries to satisfy their needs.
- Are most of the time middle to high level managers.



Actors On the Scene:

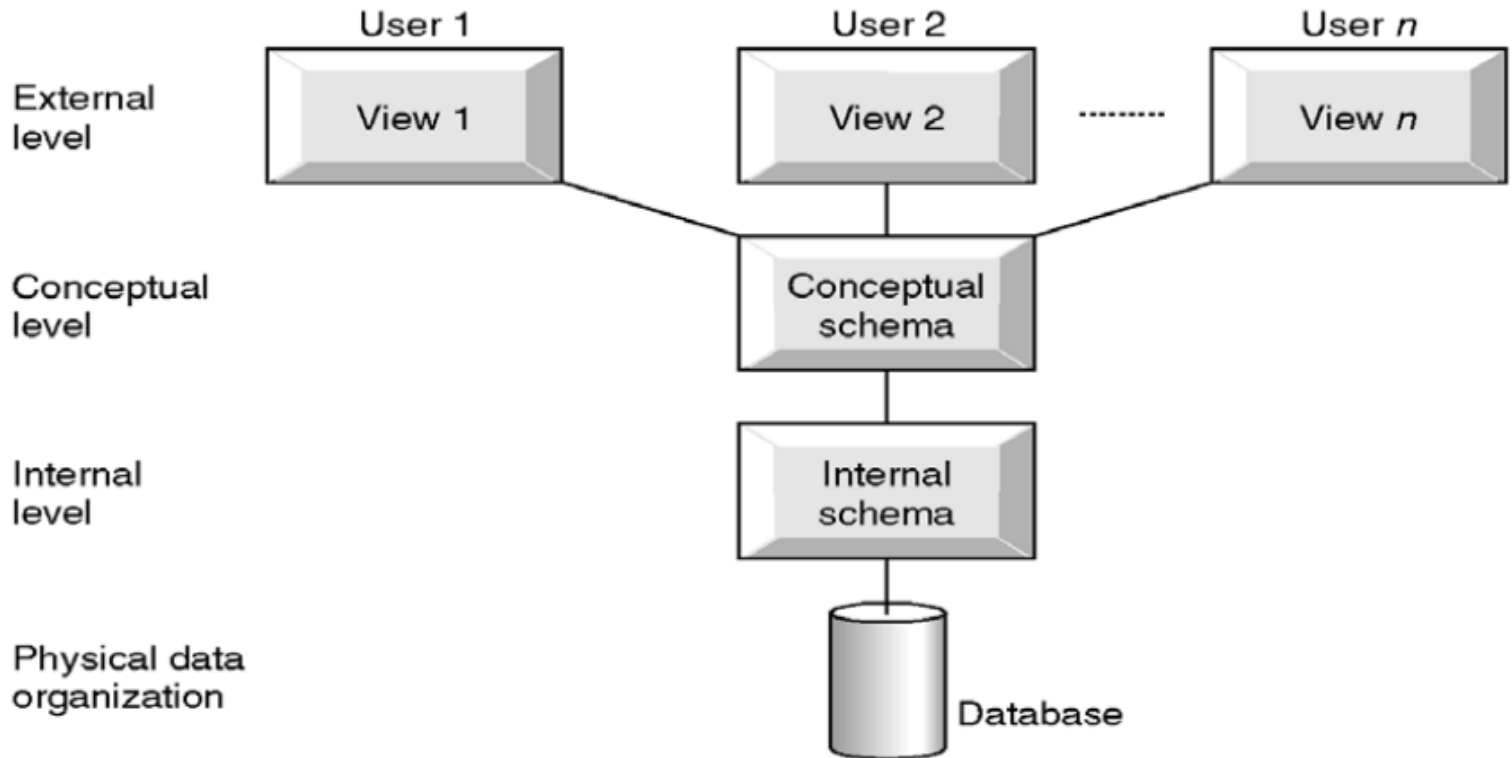
- **Data Administrator**
- **Database Administrator**
- **Database Designer**
- **End Users**



Workers Behind the Scene

- **DBMS designers and implementers:** who design and implement different DBMS software.
- **Tool Developers:** experts who develop software packages that facilitates database system designing and use.
- **Operators and Maintenance Personnel:** system administrators who are responsible for actually running and maintaining the hardware and software of the database system and the information technology facilities.

ANSI-SPARC Architecture and Database Design Phases



ANSI-SPARC Three-level Architecture



External Level

- Users' view of the database.
- Describes that part of database that is relevant to a particular user.
- Different users have their own customized view of the database independent of other users.



Conceptual Level

- Community view of the database.
- Describes what data is stored in database and relationships among the data.



Internal Level

- Physical representation of the database on the computer.
- Describes how the data is stored in the database.

Cont...

External view 1

Sno	FName	LName	Age	Salary
-----	-------	-------	-----	--------

External view 2

Staff_No	LName	Bno
----------	-------	-----

Conceptual level

Staff_No	FName	LName	DOB	Salary	Branch_No
----------	-------	-------	-----	--------	-----------

Internal level

```
struct STAFF {  
    int Staff_No;  
    int Branch_No;  
    char FName [15];  
    char LName [15];  
    struct date Date_of_Birth;  
    float Salary;  
    struct STAFF *next;           /* pointer to next Staff record */  
};  
index Staff_No; index Branch_No; /* define indexes for staff */
```



Defines DBMS schemas at three levels

- **Internal schema** at the internal level to describe physical storage structures and access paths.
 - Typically uses a physical data model.
- **Conceptual schema** at the conceptual level to describe the structure and constraints for the whole database for a community of users.
 - Uses a conceptual or an implementation data model.
- **External schemas** at the external level to describe the various user views.
 - Usually uses the same data model as the conceptual level.

The meaning of logical and physical data independence

Logical Data Independence is

- The capacity to change the **conceptual** schema without having to change the **external** schemas and their application programs.
- Conceptual schema changes e.g. addition/removal of entities should not require changes to external schema or rewrites of application programs.
- Example, two users A and B select the same fields 'student name' and 'student ID' then user A adds a new column to the field i.e. 'student age' then this does not affect the external view of user B but the internal patterns of both the users will be changed



Cont...

Physical Data Independence

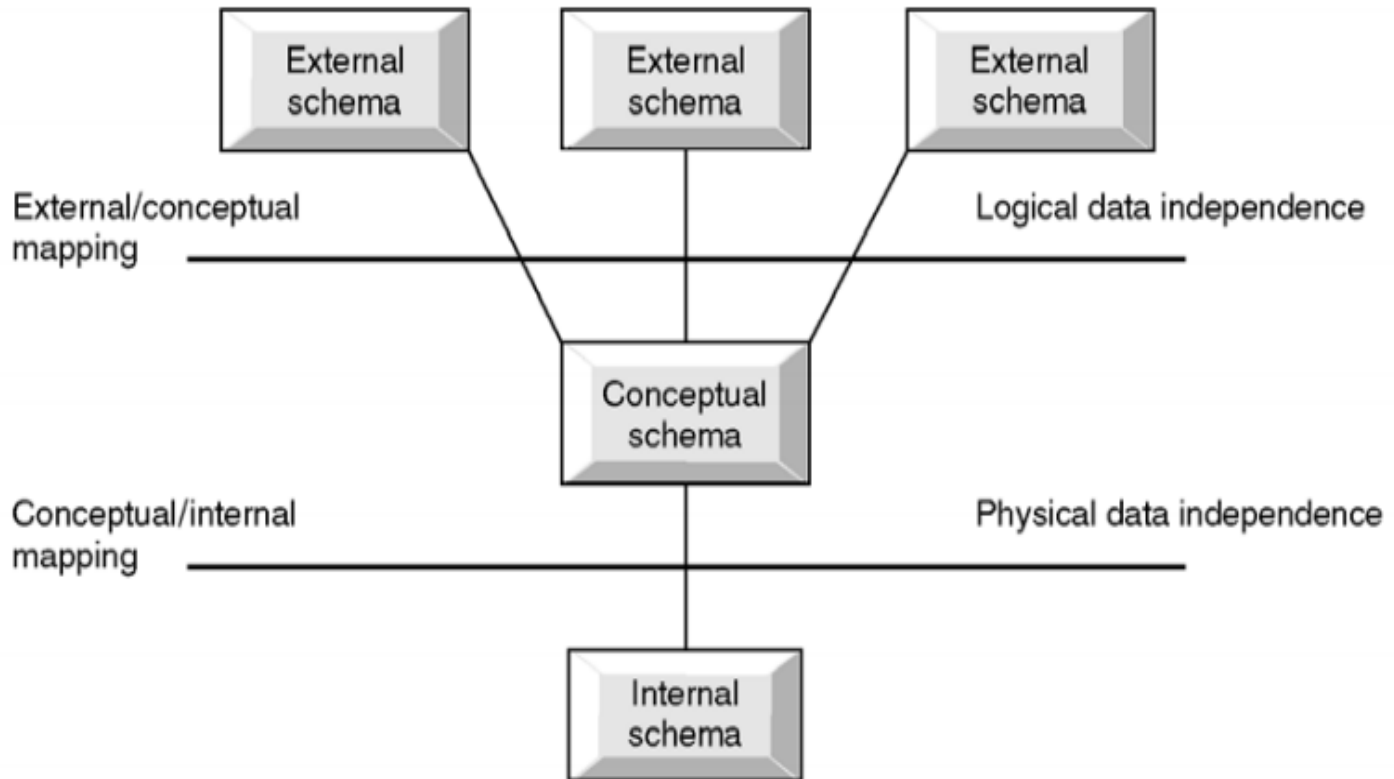
- The ability to modify the physical schema without changing the logical schema
- Applications depend on the logical schema
- In general, the interfaces between the various levels and components should be well defined so that changes in some parts do not seriously influence others.
- The capacity to change the internal schema without having to change the conceptual schema



Cont...

- Changes in physical schema may be
 - Using new storage devices
 - Using different data structures
 - Modifying index

Cont...



Data Independence and the ANSI-SPARC Three-level Architecture



Data Model

- **Data Model:** a set of concepts to describe the structure of a database, and certain constraints that the database should obey.
- A data model is a description of the way that data is stored in a database.
- Data model helps to understand the relationship between **entities** and to **create** the most effective structure to hold data



Cont...

- Data Model is a collection of tools or concepts for describing
 - Data
 - Data relationships
 - Data semantics
 - Data constraints
- The main purpose of Data Model is to represent the data in an understandable way.



Database Languages

Data Definition Language (DDL)

- Allows DBA or user to describe and name entities, attributes and relationships required for the application.
- Specification notation for defining the database schema

Data Manipulation Language (DML)

- Provides basic data manipulation operations on data held in the database.
- Language for accessing and manipulating the data organized by the appropriate data model
- DML also known as query language



Database Languages

- Procedural DML: user specifies what data is required and how to get the data.
- Non-Procedural DML: user specifies what data is required but not how it is to be retrieved

SQL is the most widely used non-procedural language query language



Cont...

- Fourth Generation Language (4GL)
 - Query Languages
 - Forms Generators
 - Report Generators
 - Graphics Generators
 - Application Generators