DATABASES

A database is an organized collection of (one or more) related data file(s). The way the database organizes data depends on the type of database, called its data model, which, may be hierarchical, network and relational models.

Benefits of databases

- Provide Data Independence
  - separating the physical (how) & logical (what) aspects of the system
- Physical data independence
  - protects the application programs from changes in the physical placement, of the files
  - the ability to modify the physical schema without changing the logical schema
- Logical data independence
  - Modify logical schema without changing application programs
- Reduce redundancy
  - reduce duplicate data items
  - some redundancy may be necessary for business or technical reasons - DBA must ensure updates are propagated (a change to one is automatically applied to the other)
- Avoid inconsistency (by reducing redundancy)
  - if it is necessary - propagate updates
- Maintain integrity - i.e. ensure the data is accurate by
  - reducing redundancy
  - implementing integrity rules, e.g. through foreign keys
- Share data
  - among existing applications
  - used in new applications
- Allow implementation of security restrictions
  - establish rules for different types of user for different types of update to database
- Enforce standards for
  - data representation - useful for migrating data between systems
  - data naming & documentation - aids data sharing & understandability
- Balance conflicting requirements
  - structure the corporate data in a way that is best for the organisation

Database Management Systems (DBMS)

DBMSs are system software that aid in organizing, controlling and using the data needed by application programs. A DBMS provides the facility to create and maintain a well-organized database. It also provides functions such as normalization to reduce data redundancy, decrease access time and establish basic security measures over sensitive data.

DBMS can control user access at the following levels:
- User and the database
- Program and the database
- Transaction and the database
- Program and data field
- User and transaction
- User and data field

Advantages of DBMS:
- Data independence for application systems
- Ease of support and flexibility in meeting changing data requirements
- Transaction processing efficiency
- Reduction of data redundancy (similar data being held at more than one point – utilizes more resources) – have one copy of the data and avail it to all users and applications
- Maximizes data consistency – users have same view of data even after an update
- Minimizes maintenance cost through data sharing
- Opportunity to enforce data/programming standards
- Opportunity to enforce data security
- Availability of stored data integrity checks
- Facilitates terminal users ad hoc access to data, especially designed query languages/application generators

Most DBMS have internal security features that interface with the operating system access control mechanism/package, unless it was implemented in a raw device. A combination of the DBMS security features and security package functions is often used to cover all required security functions. This dual security approach however introduces complexity and opportunity for security lapses.

DBMS architecture
Data elements required to define a database are called metadata. There are three types of metadata: conceptual schema metadata, external schema metadata and internal schema metadata. If any one of these elements is missing from the data definition maintained within the DBMS, the DBMS may not be adequate to meet users' needs. A data definition language (DDL) is a component used for creating the schema representation necessary for interpreting and responding to the users' requests.

Data dictionary and directory systems (DD/DS) have been developed to define and store in source and object forms all data definitions for external schemas, conceptual schemas, the internal schema and all associated mappings. The data dictionary contains an index and description of all the items stored in the database. The directory describes the location of the data and access method. Some of the benefits of using DD/DS include:

- Enhancing documentation
- Providing common validation criteria
- Facilitating programming by reducing the needs for data definition
- Standardizing programming methods

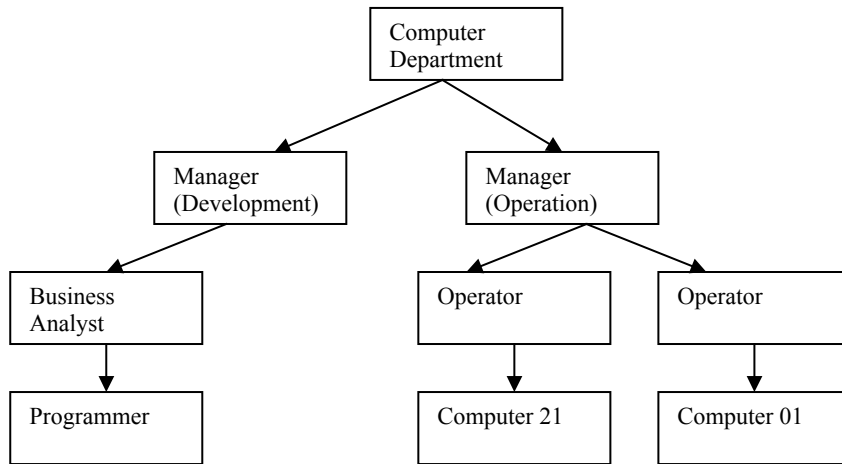## Database structure
The common database models are:
- Hierarchical database model
- Network database model
- Relational database model
- Object–oriented model

Hierarchical database model
This model allows the data to be structured in a parent/child relationship (each parent may have many children, but each child would be restricted to having only one parent). Under this model, it is difficult to express relationships when children need to

relate to more than one parent. When the data relationships are hierarchical, the database is easy to implement, modify and search.

A hierarchical structure has only one root. Each parent can have numerous children, but a child can have only one parent. Subordinate segments are retrieved through the parent segment. Reverse pointers are not allowed. Pointers can be set only for nodes on a lower level; they cannot be set to a node on a predetermined access path.

```
                        ┌──────────────┐
                        │   Computer   │
                        │  Department  │
                        └──────────────┘
                         ╱            ╲
              ┌──────────────┐    ┌──────────────┐
              │   Manager    │    │   Manager    │
              │ (Development)│    │  (Operation) │
              └──────────────┘    └──────────────┘
                    │               ╱          ╲
              ┌──────────┐   ┌──────────┐   ┌──────────┐
              │ Business │   │ Operator │   │ Operator │
              │ Analyst  │   └──────────┘   └──────────┘
              └──────────┘        │              │
                    │        ┌──────────┐   ┌──────────┐
              ┌──────────┐   │Computer 21│  │Computer 01│
              │Programmer│   └──────────┘   └──────────┘
              └──────────┘
```

Network Database Model

The model allows children to relate to more than one parent. A disadvantage to the network model is that such structure can be extremely complex and difficult to comprehend, modify or reconstruct in case of failure. The network structure is effective in stable environments where the complex interdependencies of the requirements have been clearly defined.

The network structure is more flexible, yet more complex, than the hierarchical structure. Data records are related through logical entities called sets. Within a network, any data element can be connected to any item. Because networks allow reverse pointers, an item can be an owner and a member of the same set of data. Members are grouped together to form records, and records are linked together to form a set. A set can have only one owner record but several member records.
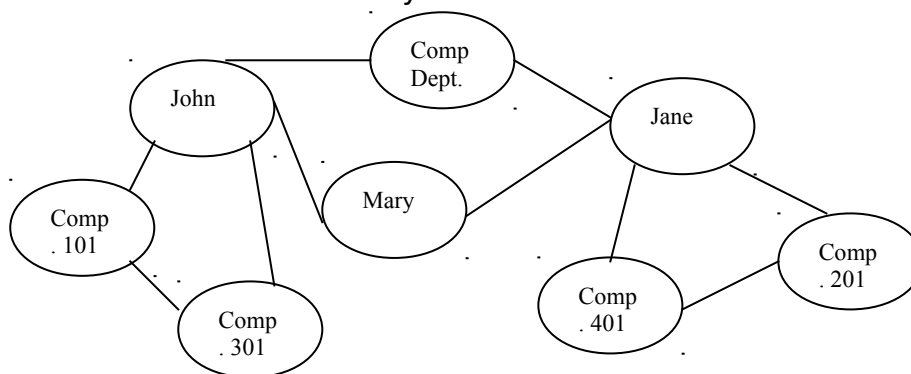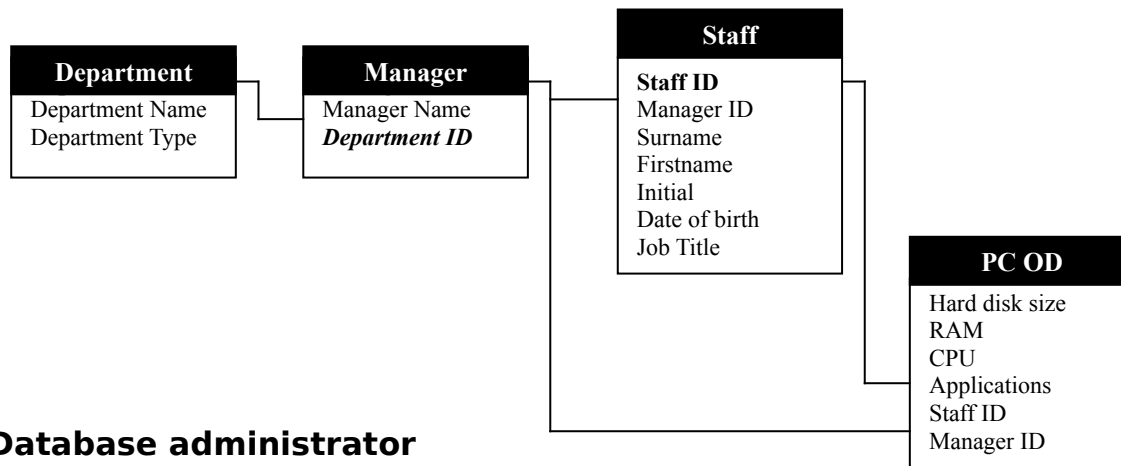

Relational Database Model

The model is independent from the physical implementation of the data structure. The relational database organization has many advantages over the hierarchical and network database models. They are:

- Easier for users to understand and implement in a physical database system
- Easier to convert from other database structures
- Projection and joint operations (referencing groups of related data elements not stored together) are easier to implement and creation of new relations for

applications is easier to do.
- Access control over sensitive data is easy to implement
- Faster in data search
- Easier to modify than hierarchical or network structures

Relational database technology separates data from the application and uses a simplified data model. Based on set theory and relational calculations, a relational database models information in a table structure with columns and rows. Columns, called domains or attributes, correspond to fields. Rows or tuples are equal to records in a conventional file structure. Relational databases use normalization rules to minimize the amount of information needed in tables to satisfy users' structured and unstructured queries to the database.

Department	Manager	Staff	PC OD
Department Name Department Type	Manager Name ***Department ID***	**Staff ID** Manager ID Surname Firstname Initial Date of birth Job Title	Hard disk size RAM CPU Applications Staff ID Manager ID

Database administrator
Coordinates the activities of the database system. Duties include:
- Schema definition
- Storage structure and access method definition
- Schema and physical organisation modification
- Granting user authority to access the database
- Specifying integrity constraints
- Acting as liaison with users
- Monitoring performance and responding to changes in requirements
- Security definitions

Database Security, Integrity and Control
Security is the protection of data from accidental or deliberate threats, which might cause unauthorized modification, disclosure or destruction of data and the protection of the information system from the degradation or non-availability of service. Data integrity in the context of security is when data are the same as in source documents and have not been accidentally or intentionally altered, destroyed or disclosed. Security in database systems is important because:

- Large volumes of data are concentrated into files that are physically very small
- The processing capabilities of a computer are extensive, and enormous quantities of data are processed without human intervention.
- Easy to lose data in a database from equipment malfunction, corrupt files, loss during copying of files and data files are susceptible to theft, floods etc.
- Unauthorized people can gain access to data files and read classified data on files
- Information on a computer file can be changed without leaving any physical trace of change

- Database systems are critical in competitive advantage to an organization

Some of the controls that can be put in place include:
1) Administrative controls – controls by non-computer based measures. They include:
    a. Personnel controls e.g. selection of personnel and division of responsibilities
    b. Secure positioning of equipment
    c. Physical access controls
    d. Building controls
    e. Contingency plans
2) PC controls
    a. Keyboard lock
    b. Password
    c. Locking disks
    d. Training
    e. Virus scanning
    f. Policies and procedures on software copying
3) Database controls – a number of controls have been embedded into DBMS, these include:
    a. Authorization – granting of privileges and ownership, authentication
    b. Provision of different views for different categories of users
    c. Backup and recovery procedures
    d. Checkpoints – the point of synchronization between database and transaction log files. All buffers are force written to storage.
    e. Integrity checks e.g. relationships, lookup tables, validations
    f. Encryption – coding of data by special algorithm that renders them unreadable without decryption
    g. Journaling – maintaining log files of all changes made
    h. Database repair
4) Development controls – when a database is being developed, there should be controls over the design, development and testing e.g.
    a. Testing
    b. Formal technical review
    c. Control over changes
    d. Controls over file conversion
5) Document standards – standards are required for documentation such as:
    a. Requirement specification
    b. Program specification
    c. Operations manual
    d. User manual
6) Legal issues
    a. Escrow agreements – legal contracts concerning software
    b. Maintenance agreements
    c. Copyrights
    d. Licenses
    e. Privacy
7) Other controls including
    a. Hardware controls such as device interlocks which prevent input or output of data from being interrupted or terminated, once begun
    b. Data communication controls e.g. error detection and correction.

Database recovery is the process of restoring the database to a correct state in the event of a failure.

Some of the techniques include:
1) Backups
2) Mirroring – two complete copies of the database are maintained online on different stable storage devices.

3) Restart procedures – no transactions are accepted until the database has been repaired
4) Undo/redo – undoing and redoing a transaction after failure.

A distributed database system exists where logically related data is physically distributed between a number of separate processors linked by a communication network.

A multidatabase system is a distributed system designed to integrate data and provide access to a collection of pre-existing local databases managed by heterogeneous database systems such as oracle.

QUESTION ONE

(a) Describe different type of database models (10 Marks)
(b)
      (i) What is a Database Management System (DBMS)?   (2 Marks)
      (ii) What are the advantages of using a DBMS?   (8 Marks)
                                    (Total: 20 marks)

QUESTION TWO
(a) What factors would you consider when purchasing a software application for your department?   (6 Marks)