DataBase Management Systems

In early day computing, data was kept in data files and programs were written to interact with these data files in order to produce reports. This type of environment lead to a situation in which every new application required the services of an experienced programmer, who usually took a fairly long time to study, design, develop and test his application. Very often the number of programmers in an organisation were limited and computing requests of various departments were many. This resulted in a long queue of undeveloped applications and a general feeling of dissatisfaction about the wisdom of the investment in computers.

World over the computer professionals were aware of this problem and were busy researching to find ways and means of improving productivity. One obvious point, which came to the forefront was that a major portion of their programming effort involved file handling, in which lines of code for opening, closing, writing, deleting etc. had to be written for every program.

Apart from this, if computers were to deliver their full potential, it was essential that the actual user of application should become eager to accept the computer as a friend.

All these factors along with several other considerations led to the development of file management systems, in which most of the common file management programs were incorporated, thereby eliminating the need of writing separate programs for them.

This, translated into action, meant that with the help of easy and simple commands, the user is presented with a screen to create, edit, save a file or enter data into the file. This data in then retrieved for any condition specified through simple commands.

File managers are useful when dealing with single files. Complex application require the data to be organised in more than one file. As such, in order to cope with this reality, file manager's capability was extended leading to the creation of database management systems.

A database management system can be thought of as a manager of the database consisting of different types of data organised into several files.

The data model used by most available DBMS is one in which information is grouped into files. Each file contains records. Each record contains information about a particular instance of an event or an object, and all the records represent similar events or objects. The terminology used is as follows:

DATABASE

A database is a data structure that stores organized information. Most databases contain multiple tables, which may each include several different fields. For example, a company database may include tables for products, employees, and financial records. Each of these tables would have different fields that are relevant to the information stored in the table.

A database is a collection of information that is organized so that it can easily be accessed, managed, and updated. In one view, databases can be classified according to types of content: bibliographic, full-text, numeric, and images.

Entity : It is a real world object or event which is interest to us. For example, if we are interested in computing the Payroll of the company, one of the real word object of interest to us is the employee.

Attributes : These are the characteristics or properties of the entity. For example the characteristics of employee for the payroll application will be staff number, name, designation, basic pay, allowances, number of leaves, advances/loans and so on.

MORE ON FIELDS

Field names : Every field must have a name so that data in the field can be accessed using that name. Usually field names are chosen represent as closely as possible that real word items they represent. Each DBMS has its rules for naming conventions to be used.

Field types : Data types: since the user may want to treat data in different ways, DBMS allows the user to define the data types as numeric, character, date, logical, memo, etc. A user has to be very careful in selecting the data type.

DATABASE FEATURES

Covered below are some of the features which differentiate database management from ordinary data processing.

Operation modes : The interaction of the user with the database can occur through command driven mode, menu-driven mode and program driven mode. Some database systems have all the three capabilities while the others are built around only one.

The Command mode : Here the user gives commands to the DBMS interactivily. Due to this feature some technical experience is necessary and is therefore not popular with people who wish to use the database once in a while.

Menu driven mode : This mode of operation is generally used by user with small and clearly defined

applications. The advantage of this method of operation is that it gives all necessary choices on the screen and as such the user need not have much technical expertise.

Program mode : In case the users needs are very complex, then manipulation of the database has to be done through one or many application programs. In such a situation where a large amount of data processing is involved program mode is best suited.

Data Access : Database systems demands that data should be accessed easily. This includes modifying data, adding data, searching the database and generating reports. Usually these operations are easily accomplished in database systems.

Help Facility : This feature allows the user to obtain online help while interacting with the DBMS.

ADVANTAGES OF THE DBMS APPROACH

1. Data Redundancy is reduced considerably thereby solving problems of excessive memory requirements and data inconsistencies.
2. Programming effort is also reduced considerably because of the built-in capabilities of the DBMS.
3. Response time is much faster since the end-user can directly interact with the DBMS.
4. Data independence is achieved as the responsibility of knowing the physical details of data storage lies with the DBMS. The users concentrate on the logical contents of the data.
5. Most databases give the users the freedom to make database design changes very easily. In contrast, in a non DBMS environment, these changes would require considerable programming effort.
6. Significant cost savings are achieved through the use of a DBMS.
7. DBMS's offer a wide variety of security and privacy features which are essential for the success of the system.

FILE STRUCTURES

Some of the commonly used files structures are discussed below :

1. Sequential files : The sequential file is distinguished by the following characteristics :

a. In this type of file the records should be accessed in same order in which they were written. In other words, there is no method by which a particular record can be accessed by skipping other records.

b. If the file has to be updated and a record has to be modified, then the file has to be rewritten to another file.

Sequential files are suited for applications which require all file records to be processed at the same time. The most common application is the Payroll system of any organisation, where salaries have to be computed once a month for all the employees.

2. Random Access Files : Many applications require the records in a file to be accessed in a random manner rather than sequentially. The rail reservation system requires that the booking position should be available in a few seconds. Random access files have this characteristic. Two of the most commonly used random access files are indexed file and index-sequential files.

Consider the example of an employee file in which records of each employee comprising of staff number, name, designation, basic pay are stored. If we wish to find the record of particular employee, then an identifier for that record must exist. This identifier is called the record key. For instance in our example the staff number can be used as an identifier to locate the record of a particular employee. Usually a file must have at least one unique key whose value is taken as the name of the record. This is called a *primary key* of the file.

It is also possible that one key may not uniquely identify the record. For example, two persons can have the same name as well as the same fathers name. In such a situation, the name, fathers name and the address may be jointly taken as the identifier key. A key thus formed by joining two or more fields is known as a *concatenated key*. Since this combination uniquely identifies the record it may be used as a primary key.

All other fields which do not uniquely identify the record are termed as *secondary keys*.

Indexed Files : While using databases one of the method to obtain random access is to create one or more indexes on that file. Since the records of the file are scattered on the disk, an index should be created which will contain a table of values in which the first column contains record key values and the second column contains disk addresses. The keys are arranged sequentially in their column.

Suppose we have an indexed file called EMPLOYEE which is indexed on field called staffno. Since the file is indexed on the primary key, the index will be called primary key index.

Adding records to the file Employee : When the file is first opened, the operating system will allocate a block of disk space to it. As records are added to the file, they are written to this block and on filling up of this block a new block is allocated. Whenever a record is written an entry is made in the staffno index which comprises of the key value and the block number into which the record is written (known as the relative address of the record).

A file directory which gives the physical address corresponding to the relative address is also created. Whenever a search is made for the record which contains a specific value for staffno, the index table is read first and the physical address is then cross referenced.

Indexed Sequential Files : In this type of file records are physically stored on the disk in groups. Records are stored sequentially within each group.

CLB-108

When each group corresponds to a physical subdivision of the disk such as a track, the file type is known as ISAM (Indexed Sequential Access Method). In addition to this, with each track, a separate primary key index also maintained. In this arrangement, the ISAM file can be accessed in two ways. i.e., entire file can be read sequentially or alternatively records may be accessed via the index: the index is entered by key value, the address is located within the index and the record is retrieved.

These files are used typically when :

(a) Random access is needed to the records using the primary key.

(b) Frequent access to the file in primary key sequence is also needed.

Query : A database query is a piece of code (a query) that is sent to a database in order to get information back from the database. It is used as the way of retriving the information from database. A database "query" is basically a "question" that you ask the database. The result of the query is the information that is returned by the database management system. Queries are usually constructed using SQL (structured query language) which resembles a high-level programming language.

The term 'query' means to search, to question, or to find. When you query a database, your searching for information in the database. Different query languages exist for different type of databases. MS Access uses SQL, which stands for Structured Query Language. MS Access contains Tables, Forms, and Queries. The Forms are used to enter or display the data, the Tables are where the data is saved, and the queries are used to search for specific data.

An example of an query can be this:

SELECT id, customer_name, day_of_order, product, quantity

FROM orders;

The query searches the id, customer name, day, product, quantity from the order table.

LOGICAL DATABASE DESIGN

In designing for a database it is important to develop a conceptual model of the involved data which states the relationship between the data items. The purpose of this exercise is to group the data into a number of tables in order to reduce data redundancy and avoid addition, deletion and modification anomalies.

ER analysis : A method called entity-relationship analysis has been used for obtaining the conceptual model of the data, which will finally help us in obtaining our relational database. In order to carry out ER analysis it is necessary to understand and use the following features :

1. Entities : These are the real word objects in an application.

2. Attributes : These are the properties of importance of the entities and the relationships.

3. Relationships : These connect different and represent dependencies between them.

In order to understand these terms let us take an example. If a supplier supplies an item to a company, then the supplier is an entity. The item supplied is also an entity. The item and supplier are related with each other are in the sense that the supplier supplies an item. Supplying is thus the verb which specifies the relationship between item and supplier. A collection of similar entities is known as an entity set. Therefore each member of the entity set is described by its attributes.

Relationships Cardinality and Participation : A relationship cardinality is the number of relationships in which an entity can appear. This cardinality is usually shown on E-R-diagram. An entity can have —

(i) One Relationship

(ii) Fixed number of relationships

(iii) Variable number of relationships

RELATIONAL DATABASE SYSTEMS

Usually, a conceptual design cannot be directly implemented on a particular database system. Rather, the design must first be translated into a physical database design or physical model, that confirms to the features of the particular DBMS being used. The reason for this is that each DBMS has its own particular set of data structures for storing and maintaining information and there is often no direct relationship between these structures and those of a particular conceptual model. Thus, a conceptual design, which is built up from entity-types, relationships, constraints and so on, must be recast into forms that can be directly implemented on the DBMS. In other words, they must be recast into a suitable physical database design.

THE RELATIONAL DATABASE MODEL

Basic Characteristics : Relational databases have certain characteristics that distinguish them from other types of databases. These are :

- A relational database consists of one or more two-dimensional tables of data values, with very simple rules defining the construction of a tables of data values, with very simple rules defining the construction of a table.

- Relationships between two or more tables is based on a common field, taken as primary key in one table and foreign key in the other tables.

Table : The basic structure of a relational design is the table, known as a relation. Each row of a table, is known as a table. Each column of a table contains some data corresponding to a particular row. The terms attributes and columns are used interchangeably.

Rules for Table Construction

- Each row must be unique. This means that no two row should have the same data for every column.

○ Each entry in a table in any particular column corresponding to any row should be a single entry.

○ The order in which the rows are projected is immaterial.

○ Each column should have a name which should closely reflects the entry made in it.

File : File is a collection of related records, such as the payroll or supply inventory records. Normally, all records within the file are in the same format. When processing data, we think in terms of data files. For example, to process a parts inventory, you would need the master parts inventory file and the file that contains up-to-date information on each part that has been issued. The master parts inventory file would have a record for every part in the inventory. The update file, parts issued file, would have a record for each part issued. You would use a program to read the records on the parts issued file and update the matching records on the master parts inventory

file. Depending on whether the data is stored on magnetic tape or disk or in internal storage, the program would use different methods to access storage to obtain the data. In the next section you'll learn about storage access methods.

Field : Field is an item of information in a data record. One or more related characters that are treated as a unit of information..A field (also referred to as a data item) may be alphabetic, numeric, or alphanumeric, and may be either fixed or variable in length. For example, your social security number (SSN) is of a fixed length; that is, it's always positions in length. Names are variable length because they may be from 2 to 25 positions in length.

Record : Record is a group of related fields, all pertaining to the same subject; a person, a thing, or an event. For example, your payroll record (LES statement) might include fields for your name, amount paid, taxes withheld, earned leave, and any allotments you might have. On the other hand, a supply inventory record might consist of fields containing stock number, the name of the item, its unit price, the quantity on hand, and its bin location.

Attributes and Domains : An attribute of a table represents a particular characteristic of the set of facts represented by the table. A related concept that is very useful is the domain of an attribute, which is defined as all of the possible values that may be assumed by the attribute.

Keys (Constraints) : Constraints are used to limit the type of data that can go into a table. Constraints can be specified when a table is created (with the CREATE TABLE statement) or after the table is created (with the ALTER TABLE statement).Following are the type of constraints:

○ NULL constraint   ○ UNIQUE

○ PRIMARY KEY   ○ FOREIGN KEY

NULL constraint : The NOT NULL constraint enforces a column to NOT accept NULL values. The NOT NULL constraint enforces a field to always contain a value. This means that you cannot insert a new record, or update a record without adding a value to this field

Unique Key : The unique constraint uniquely identifies each record in a database table. the unique and primary key constraints both provide a guarantee for uniqueness for a column or set of.columns. a primary key constraint automatically has a unique constraint defined on it.

Note that you can have many unique constraints per table, but only one primary key constraint per table.

Primary key : Primary key means main key. the primary key constraint uniquely identifies each record in a database table. Primary keys must contain unique values. a primary key column cannot contain null values and also does not allow duplicate values.

FOREIGN KEY : The foreign key constraint is used to prevent actions that would destroy links between tables. The foreign key constraint also prevents that invalid data form being inserted into the foreign key column, because it has to be one of the values contained in the table it points to.

Functional Dependency : An attribute B is functionally dependent on attribute A if every value of A uniquely determines the value of B. In this case, attribute A is called the determinant of B.

Table Normalisation : The tables should be properly designed. If the tables are not pre planned in design, the organisation may face a lot of problems in terms of money and time. After the poorly database design, if the tables are implemented, it will be a costly affair to eliminate the problems since this will involve people time and large data storage. To help eliminate these problems, a systematic technique has been developed for translating a conceptual design into a set of well-designed tables. This technique is known as normalisation.

A Field

Address Books Table

Name	Address	Designation	Tel No
Vajira M.	Akuressa Rd Matara	Sales Manager	041-48245
Wimalaratne C.	Matara Rd, Hambantota	Network Engineer	047-72453
Shamini K.	Colombo-04	Clerk	01-906754
Gamage T.W.	Fort, Galle		09-20456

A Record

Address Books Table

Name	Address	Designation	Tel No
Vajira M.	Akuressa Rd, Matara	Sales Manager	041-48245
Wimalaratne C.	Matara Rd, Hambantota	Network Engineer	047-72453
Shamini K.	Colombo-04	Clerk	01-906754
Gamage T.W.	Fort, Galle		09-20456

OBJECTIVE QUESTIONS

1. Information in a database is typically stored in this, which consists of vertical columns and horizontal rows ?
   (1) Field   (2) Record
   (3) Table   (4) File
   (5) None of these
   SBI PO Tier-2 Exam, 13.10.2008

2. A _____ is an organised collection of data about a single entity.
   (1) file
   (2) library
   (3) database
   (4) dictionary
   (5) None of these
   SBI PO Tier-2 Exam, 13.10.2008

3. The smallest unit of information about a record in a database is called a
   (1) Cell
   (2) Field
   (3) Record
   (4) Query
   (5) None of these
   Allahabad Bank Clerk Exam, 31.08.2008

4. In database, a field is a _____.
   (1) label
   (2) table of information
   (3) group of related records
   (4) category of information
   (5) None of these
   SBI PO Tier-I Exam, 27.07.2008

5. _____ means that the data contained in a database is accurate and reliable.
   (1) Data redundancy
   (2) Data integrity
   (3) Data reliability
   (4) Data consistency
   (5) None of these
   SBI PO Tier-I Exam, 27.07.2008

6. Arranging of data in a logical sequence is known as
   (1) Classifying
   (2) Searching
   (3) Sorting
   (4) Reproducing
   (5) None of these
   SBI PO Tier-2 Exam, 27.07.2008

7. Data structure includes
   (1) Array
   (2) Execution statement
   (3) Interaction
   (4) Both (2) and (3)
   (5) List
   SBI Clerk Exam, 10.02.2008

8. The dbase III plus is mostly used for
   (1) Office automation
   (2) Database management problems
   (3) Scientific problems
   (4) Calculations only
   (5) None of these

9. The basic building block of a database that contains related records is
   (1) Query
   (2) Form
   (3) Table
   (4) Edited Page
   (5) Report

10. Which out of the following is not a DBMS software ?
    (1) dBASE
    (2) FOXPRO
    (3) ORACLE
    (4) SyBase
    (5) Data base 2000

11. Array is —
    (1) linear data structure
    (2) non-linear data structure
    (3) complex data structure
    (4) simple data structure
    (5) None of these
    SBI Clerk Exam, 25.11.2007

12. A computer checks the _____ of user names and passwords for a match before granting access.
    (1) website   (2) network
    (3) backup file
    (4) database
    (5) None of these
    SBI Clerk Exam, 19.08.2007

13. In a relational database, this is a data structure that organizes the information about a single topic into rows and columns.
    (1) block   (2) record
    (3) tuple   (4) table
    (5) Command line interface

14. The default extension of Access database file is
    (1) db   (2) adb
    (3) dba   (4) mdb
    (5) None of these
    Allahabad Clerk Exam, 31.08.2008

15. In Access, this can be created, selected and manipulated as a unit
    (1) property   (2) file
    (3) object
    (4) operator
    (5) None of these

16. A storage device where the access time is depended upon the location of the data is
    (1) Random access
    (2) Serial access
    (3) Sequential access
    (4) Transaction access
    (5) None of these

17. In a database _____ fields store numbers used to perform calculation.
    (1) Next   (2) Key
    (3) Alphanumeric
    (4) Numeric
    (5) None of these
    SBI Clerk Exam, First Sitting, 19.07.2007

18. Which of the following file organization is most efficient for a file with a high degree of file activity?
    (1) Sequential
    (2) ISAM
    (3) VSAM
    (4) B-Tree Index
    (5) None of these ,

19. Which of the following types of tables constraints will prevent the entry of duplicate rows ?
    (1) Primary key
    (2) Unique
    (3) Null
    (4) Foreign key
    (5) None of these

SBI Bank Clerk Exam, 06.01.2008

20. Which of the following is a reason for creating a list instead of creating a database?
    (1) to prevent data inconsistency
    (2) to prevent data redundancy
    (3) to make it easier to share data in a central location
    (4) to organize data for simple tasks
    (5) None of these

21. _____ means that the data contained in a database is accurate and reliable.
    (1) Data redundancy
    (2) Data integrity
    (3) Data reliability
    (4) Data consistency
    (5) None of these

22. A picture would be stored in a(n)_____ field in a database.
    (1) text   (2) memo
    (3) object   (4) hyperlink
    (5) None of these

23. In a database, a field is a:
    (1) Label.
    (2) Table of information.
    (3) Group of related records.
    (4) Category of information
    (5) None of these

24. A group of related fields in a database is called a(n):
    (1) Record.   (2) Object.
    (3) Memo.   (4) Table
    (5) None of these

25. A group of related records in a database is called a(n):
    (1) Column.   (2) Object.
    (3) Key.   (4) Table.
    (5) None of these

26. All of the following are types of databases, EXCEPT:
    (1) Relational.
    (2) Relation-oriented.
    (3) Object-oriented.
    (4) Multidimensional
    (5) None of these

27. In a relational database, tables are logically linked to each other by a:
    (1) key.
    (2) hyperlink.
    (3) field type.
    (4) field size
    (5) None of these

28. Which of the following would NOT ordinarily be considered unstructured data?
    (1) text
    (2) video
    (3) audio clips
    (4) MP3 files
    (5) None of these

29. In addition to keying data directly into the database, data entry can be done from a(n):
    (1) Input form.
    (2) Table.
    (3) Field.
    (4) Data dictionary
    (5) None of these

30. A (n) _____ enables users to have the database select and display records that match a certain criteria.
    (1) Query
    (2) Sort
    (3) Operation
    (4) Join
    (5) None of these

31. The primary key must be a(n) _____ field.
    (1) numeric
    (2) object
    (3) unique
    (4) related
    (5) None of these

32. A _____ is a large-scale electronic repository of data that is time-variant.
    (1) data warehouse
    (2) data mine
    (3) knowledge-based system
    (4) decision support system
    (5) None of these

33. Which of the following are objects in an access database?
    (1) table, form
    (2) queries, reports
    (3) macros, modules
    (4) All of the above
    (5) None of these

ANSWERS

1. (3)	2. (3)	3. (2)	4. (2)
5. (2)	6. (3)	7. (1)	8. (2)
9. (5)	10. (5)	11. (1)	12. (4)
13. (4)	14. (4)	15. (3)	16. (2)
17. (4)	18. (1)	19. (1)	20. (4)
21. (1)	22. (1)	23. (4)	24. (1)
25. (4)	26. (2)	27. (1)	28. (1)
29. (1)	30. (1)	31. (3)	32. (1)
33. (4)			

EXPLANATIONS

3. (2) Field is one or more related characters that are treated as a unit of information. A field (also referred to as a data item) may be alphabetic, numeric, or alphanumeric, and may be either fixed or variable in length.