

Week 3 Lecture Note

DATA STRUCTURES

ORGANISATION OF DATA

Characters, facts, records, files and databases form an organisation of data. The basic building block of data is a character. The character consists of upper and lower-case letter, numeric digits or symbols. Upper and lower-case letters are Aa, Bb, Cc,... Zz. Numeric digits are 0, 1, 2,...,9. Symbols involve commas (,) quotation mark (?) plus (+) division (/) and so on. Upper and lower-case letters are called alphabetic character. Numeric digits are called numeric character. Symbol is called special characters. A combination of the three types is called alphanumeric characters (#2B, N2.50K). A computer can accept both alphanumeric and numeric and store them in memory.

Characters are put together to form a fact. A fact is also called a field. A fact or field is a number, an item, word, name or a combination of characters. Facts are put together to form a record. A record is a related items of data in a file. An employee record in a company would be a collection of facts about one employee. Their facts would include the employee's name, address, department, phone number, position, pay rate, earning made to date etc.

Records are combined together to make a file. A collection of related records is a file. E.g. A collection of all employee records for one company would be an employee file. Files are combined together to make a database. The heart of most computer processing is data. An organisation uses data as raw materials to be stored in database. Once the data have been processed, they are called information.

TYPES OF DATA

There are two types of data - Numeric data and alphanumeric data. Numeric data is expressed in numbers e.g. age is 35; date of birth is 1970. Numeric data contain only numeric characters or numbers.

Alphanumeric data is composed of combination of letters, numbers or special punctuation character e.g.

Name = Abeokuta

Address = 17, Ibadan Road

Date = 26th October, 2000

STRUCTURE OF THE DATA

The structure of the data is the composition of records into files for generating information. Let us take an example of long-distance telephone call, the following items of data are recorded:

- Telephone number of the person to whom the call is to be billed
- Telephone number of the person receiving the call
- Duration of the call in minutes
- Time that call is placed
- Type of call e.g. person-to-person or station-to-station.

These data need processing for generating bill information.

Description – Long-distance telephone call data		
Field Names	Types of data	Number of Character
Phone no to be billed	Numeric	10
Phone no of call receiver	Numeric	10
Duration of call	Numeric	4
Time call is placed	Numeric	4
Type of call	Alphanumeric	1
Total Character		29

In database system, the structure will be like this:

Phone No to be billed

10 characters

Phone No of call receiver	10 characters
Duration of call	4 characters
Time call is placed	4 characters
Type of call	1 character

Let assume that we have 2,000 calls for the month.

The records will look as this

Record	Phone to be billed	Phone no of call receiver	Duration of call	Time call is placed	Types of call
1	221	134	1 min	2 p.m.	P
2	411	820	1 min	3 p.m.	S
.					
.					
2000	238	918	4 min	5 p.m.	S

Each record has five fields of 29 characters. With 2,000 records, the file requires 2000 x 29 characters on external disk device. To process these long-distance telephone call data in a bill, the records have to be identified through key record. The record key in this case may be the telephone number of the person to whom the call is to be billed. This information may be kept in a computer storage device and named Telephone.doc as a file name.

FILE CONCEPTS

BIT:- This stands for BINARY DIGIT. It is the unit used in binary representation.

BYTE:- A sequence of bits operated upon as a unit and usually shorter than a computer word. A byte is a character and is equal to 8 bits (which means 3 bytes are three characters and are equal to 24 bits).

CHARACTER:- This is a number or a letter or a symbol e.g. L, 4, +, etc.

WORD:- A group of characters occupying one storage location in a computer. It is treated by the computer circuits as an entity, by the control unit as an instruction, and by the arithmetic units as a quantity.

FIELD:- This is a combination of related characters.

RECORD:- This is a combination of related fields. More precisely, a collection of related items of data (fields) located as a unit.

FILE:- This is a collection of related records. An organized, named collection of records treated as a unit, or the storage device on which these records are kept.

DATABASE:- A non redundant collection of interrelated data file items.

TYPES OF FILES

There are four basic types of files.

- Transaction file:- This a temporary file that represents sales orders for a day in selling merchandise business.
- Master file:- This is a file that is fairly permanent in nature e.g. A file that contains records of employees for a company.
- Reference file:- This is a file that is semi permanent. An example is a file containing price list.
- Historical file:- This is a permanent file.

FILE PROCESSING

This refers to the various activities that can be carried out on the record of a file. Some of these processing activities are:-

- Sorting:- This is the arrangement of the records of a file in ascending or descending order.
- Merging:- This is the combination of two or more files into a single file.
- Validation:- This is a programming technique of carrying out logical check on the data being captured on the computer.

- Referencing:- This is the accessing of a particular record of a file on order to ascertain its contents.
- File maintenance:- This refers to the addition of new records or deletion of obsolete records or modification of existing records.
- Updating:- This is the process of making the master file to reflect the most current situation.
- File enquiry or interrogation:- This is very similar to referencing. It is ascertaining the content of the file for decision-making purposes.
- Searching:- This is the process of looking through a set of record on the file with a view to making use of those record that have similar characteristics.

FILE ORGANISATION

There are four methods by which file can be organized on a magnetic disk.

- Serial:- In this method, records are written into the tape one after the other without any regard to the order of the record key.
- Sequential:- In this method, records are first sorted according to a particular order of the record keys before being written on the disk one after the other.
- Indexed sequential:- In this method, an index is created on the disk indicating the address or location of each record as it is being written into the disk.
- Random:- In this method, records are written into the disk any how. But a mathematical formula is put in place by programming technique in such a way that the formula yields the location or address of a particular record whenever as key field is substituted.

OTHER FILE CHARACTERISTICS/ADDRESS CONCEPTS

- Volatility:- This is the frequency with which records are added to a file or deleted from it.
- Size:- This is the amount of data stored in a file.
- Growth:- This refers to increase in the size of a file as records are being added.
- Cylinder:- This is the major sub-division of a disk.
- Track:- Each cylinder is made up of a certain number of tracks.
- Sector:- This is the smallest addressable part of a disk, it is also called Block.