SECTION 2

# FILES.

22.3.84

## Defintion.

A file is a set of information (record) that are in some way connected with each other e.g names beginning with 'A'

Before electronic data capture was invented such files were kept on pieces of paper in filing cabinets. Today the information is stored on magnatic media.

## Types of files.

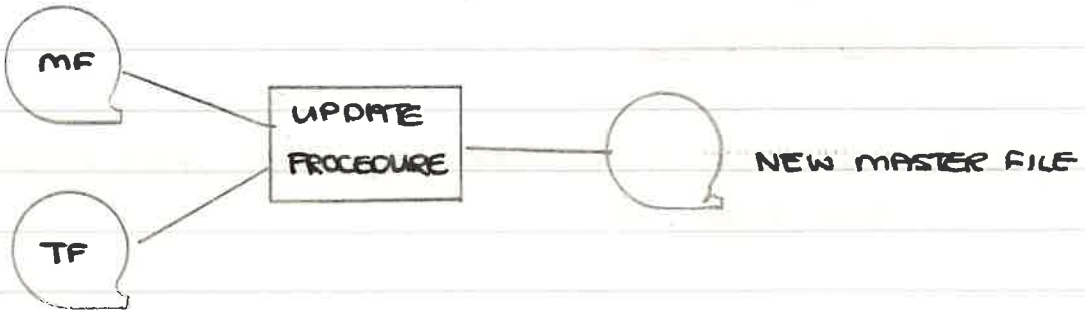
1) A master files

This is the main file updated only on a weekly basis. It contains names, addresses etc. It is kept safe in case of fire etc.

2) Transaction files

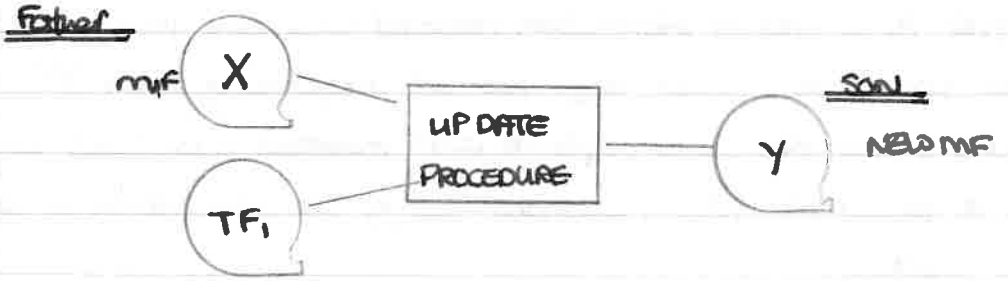
This file is updated on a daily basis. It contain's name, account number, current balance, withdrawals and deposits. It has to be updated daily because of the number of transactions made daily

## Update of files

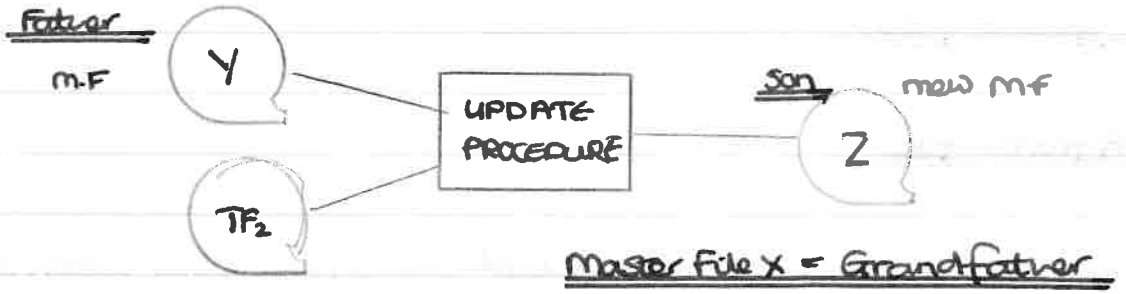


## File Security

### Week 1




### Week 2




Master file X = Grandfather

If you only had one master file it might get destroyed by fire, flood etc, and all your records would be gone. Therefore every time you create a new master file it is sensible to keep the old one as well in a seperate place, incase the new one is destroyed so you can make another new master file. It even more sensible to keep three copies, three "generations" of master file because the chances of three seperate disasters are pretty slim. on the third week the oldest or "grandfather" masterfile will be erased.

## File organisation

1) **File** - collection of information e.g people whose surname begins with the letter 'A'

2) **Record** - A file is made up of a set of records. These contain a number of pieces of information, related data which form a single unit e.g 1 customers name, account No, address and balance - Fred adams - 1234567, 23, John St, £10.50

3. Field - each record is divided into fields. These contain one item of information e.g the first field in this record is the customers name

4 Key Field - within a file the records must be arranged in some form of logical order. In this case the key field is the customers name and the file is arranged alphabetically. This is obviously easier for the bank clerks and is in a human readable form. The file could be arranged using the account No as the key field and this is machine readable in that the computer will understand more easily a numeric order.

## Accessing Information

### Serial (Sequential) Access

Each record stored on magnetic tape can only be accessed sequentially. That is, is you want file No 21 you have to go to through files 0-20 first. This is slow. This is only true of media like magnetic tape, not discs. An application of this where serial access would not be a problem is in the payroll of a factory where all the files have to be accessed anyway

### Random (direct) access

Any record anywhere on the disc can be accessed at once Media with random access are magnetic discs, floppy discs, and magnetic drums. A drum is the fastest An application for this is an airport where seats are booked, or records in a hospital.

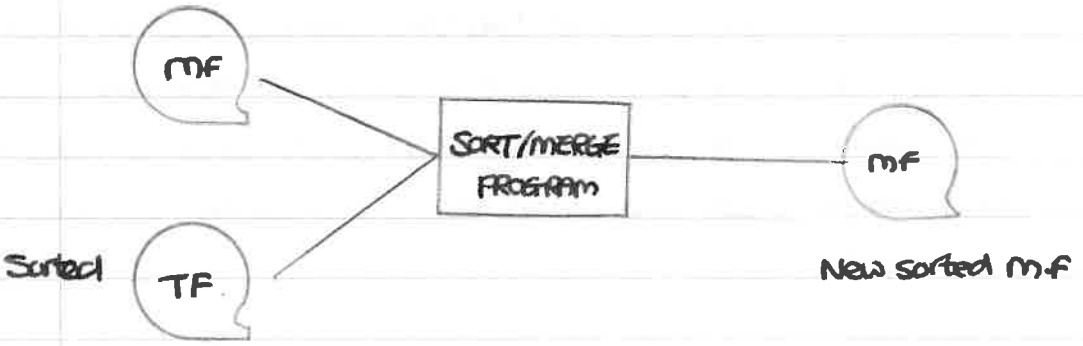
## Searching, sorting, and merging

### Searching

Searching is done by a program to search through records to find a certain field, for example at the National Blood

Transfusion service where they might need to find rare blood types for example ABRhesus negative and then find the address to send the donor a letter to tell them their blood is needed

## Sorting and Merging




At the end of the week the transaction files and ~~merging~~ master files have to be merged, and a Mr Bloggs who performs a transaction in the week will have to go next to his record on the master file so the transaction file has to be sorted before they can be merged. This then forms a new sorted master file

## Data processing!

This is defined as any process which takes information, processes it in some way and produces some expected result.

N.B

This is not confined to computer processing

## Data-capture

Data capture is gathering information in a ~~prepared~~ manual form e.g filling in forms

## Problems

If a form just asks for the date problems can arise. People write the date in different ways - 29/3/8, 29th march 84, 29th March 1984 - march 29th 1984 - The Americans even write it backwards 3/29/84.

A solution to this is having boxes

month [03] Day [29] Year [84]

## Data Preparation

### Definition

This is the process of preparing information into a machine/computer readable form, usually arranged in human readable for on specially prepared documents / questionnaire / forms.

### Data preparation department hardware

Key-to-disc encoder
Teletypewriter (paper tape)
Paper tape reader
Card Punch
Punched card reader
Verifier (punched cards)
Mark reader
Mark sensor

### Verification

**Defn** This is the process which checks that data has been entered correctly into the system. One of its uses is to verify whether or not a typist has made any mistakes entering the data.

A card punch verifier is a machine which checks that the information of the punched cards is correct. First of all one person punches the cards and they are then fed into the verifier. Here there is another person who types the information in again and if it is the same as on the card then as it is correct. If they are not the same there is an error and the card is taken out and a new one is

made and put in its place. This is a long, tedious job.
This is an off-line process because it takes place independantly of the computer

## Data Validation

Defn: An on-line process undertaken by the computer to ensure that the data which has been entered is valid
e.g that there are no letter appearing where there should be figures and vice versa.

The punch room staff should not have any technical because they would be likely to change data if it wasn't what they thought it should be.

The computer checks inputs. If it is invalid then there will be an error.

e.g If the number required is between 0 and 10 and someone inputs 25 there will be an "INVALID INPUT" error
This will also happen if a letter is entered instead of a number

## Check Sums, check digits

Numerical data is added up. If this gives the correct result then the data is correct.

DATA 1, 2, 3 1+2+3 = 6 - check sum

DATA 1, 2, 3, 6 checksum digit.

## Batch total, hash total

This is when a whole batch of data has to be checked, the batch total is the total of the check digits.

## Modulus Eleven

This is a method of checking input data using a system of weightings. e.g

number = 4 7 8 4 3 ?
6 5 4 3 2 1

$2\times3 + 4\times3 + 8\times4 + 7\times5 + 6\times4$
$= 6 + 12 + 32 + 35 + 24$
$= 109$

$11 \overline{)109}$ r 10 (9)

$11 - 10 = 1$

47843 1

(If the number is 10, use T. eg if the remainder = 1)

## Merging

At the end of a week at a bank for example, the transaction file and master file have to be merged to form a new master file. The process is a lot easier if both files are sorted on the same key field. The two files are compared, and all relevant information is transferred into a third file, the new master file.

## Example Question

For the two files A + B the first item of each record is to be taken as the key by which they are sequenced.

File A

| | | |
|---|---|---|
| 256023 | A.F. SMITH | £234.56 |
| 403214 | J.P. JONES | £156.25 |
| 207888 | L.C JACKSON | £2478.00 |

File B

864512 P.R. TAYLOR £105.23
956421 A. FREEMAN £325.20
125642 S. ARBER £1025.60

a) Sort the two files into correct sequence
b) Merge them. Give algorithm.

a) Sort

File A

207888, L.C. JACKSON. £2478.00.
256023. A.F. SMITH. £234.56.
403214. J.P. JONES. £156.25.

File B

125642. S. ARBER. £1025.60.
864512. P.R. TAYLOR. £105.23.
956421. A. FREEMAN. £325.20.

b) Merge Algorithm

125642 S. ARBER £1025.60
207888 L.C. JACKON £2478.00
256023 A.F.SMITH £234.56
403214 J.P. JONES £156.25
864512 P.R.TAYLOR £105.23
956421 A.FREEMAN £325.20

1. Check first numbers from both files, put lower into 3rd file
2. Check lowest numbers from each file, compare them, and put the lowest into the third file.

3. Repeat until files A+B are empty:

Computers merge by using bubble sort

## Searching

### Linear Search

A search through a whole file, checking a certain field against the information you want to know, e.g. to find who lives in a certain area etc. This is very slow.

### Binary Search.

Sequentially narrowing down the field of search by approximately half until only one is left:

| Question | Answer | Field of search |
|---|---|---|
| | | A → Z (26 letters) |
| A→M | NO | N → Z |
| N→T | YES | N → T |
| N→Q | YES | N → Q |
| N,O | NO | P → Q |
| P | NO | Q |

Using this ~~techniue~~ method, one entry out of a million can be identified after 20 questions requiring a YES/NO answer. Only a very simple condition program is therefore required