



Session 9

File Handling





Objectives

- ☐ Explain file organization
- ☐ Describe different types of files
- ☐ Explain the different file operations

For Aptech Centre Use Only



Introduction 1-2

□ A File:

- Is a collection of correlated data or bytes stored on a secondary storage device
- Stores information permanently which can be retrieved by specific programs

□ The collection of bytes can be interpreted as:

- Words
- Characters
- Lines
- Paragraphs and pages from a textual document
- Fields
- Records belonging to a database
- Pixels from a graphical image

Introduction 2-2

- There are two kinds of files and they are as follows:



Text
files



Binary
files

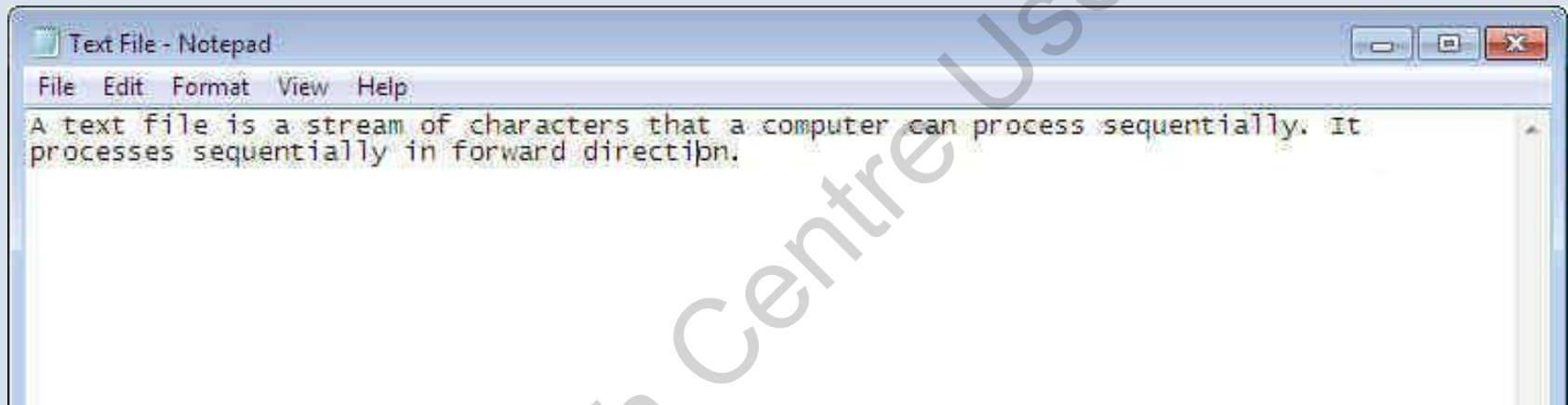


Text Files 1-2

Contains stream of characters that a computer can process sequentially in forward direction

Opened only for a single operation, such as reading, writing, or appending

Text Files 2-2





Binary Files 1-2

Are collection of bytes

Involve no special processing of data and each byte of data is transferred to or from the disk unprocessed

Are either processed sequentially or depending on the needs of the application

Enable to perform read and write operations simultaneously

Example: A database file is created and processed as a binary file

Binary Files 2-2

0000	FF	D8	FF	E1	1D	FE	45	78	69	66	00	00	49	49	2A	00
0010	08	00	00	00	09	00	0F	01	02	00	06	00	00	00	7A	00
0020	00	00	10	01	02	00	14	00	00	00	80	00	00	00	12	01
0030	03	00	01	00	00	00	01	00	00	00	1A	01	05	00	01	00
0040	00	00	A0	00	00	00	1B	01	05	00	01	00	00	00	A8	00
0050	00	00	28	01	03	00	01	00	00	00	02	00	00	00	32	01
0060	02	00	14	00	00	00	B0	00	00	00	13	02	03	00	01	00
0070	00	00	01	00	00	00	69	87	04	00	01	00	00	00	C4	00
0080	00	00	3A	06	00	00	43	61	6E	6F	6E	00	43	61	6E	6F
0090	6E	20	50	6F	77	65	72	53	68	6F	74	20	41	36	30	00
00A0	00	00	00	00	00	00	00	00	00	00	00	00	B4	00	00	00
00B0	01	00	00	00	B4	00	00	00	01	00	00	00	32	30	30	34
00C0	3A	30	36	3A	32	35	20	31	32	3A	33	30	3A	32	35	00
00D0	1F	00	9A	82	05	00	01	00	00	00	86	03	00	00	9D	82
00E0	05	00	01	00	00	00	8E	03	00	00	00	90	07	00	04	00

File Organization 1-2

- ❑ Refers to the relationship of the key of the record to the physical location of that record in the computer file
- ❑ Can either be:
 - Physical file - A physical file is a physical unit, such as magnetic tape or a disk
 - Logical file - A logical file is a complete set of records for a specific application

File Organization 2-2

- The objectives of computer based file organization are as follows:
 - Ease of file creation and maintenance
 - Efficient means of storing and retrieving information

File Organization Methods 1-2

- ❑ The three file organization methods are as follows:

Sequential access

Direct access

Indexed sequential access

File Organization Methods 2-2

- ❑ The selection of a particular method depends on:
 - Type of application
 - Method of processing
 - Size of the file
 - File inquiry capabilities
 - File volatility
 - The response time

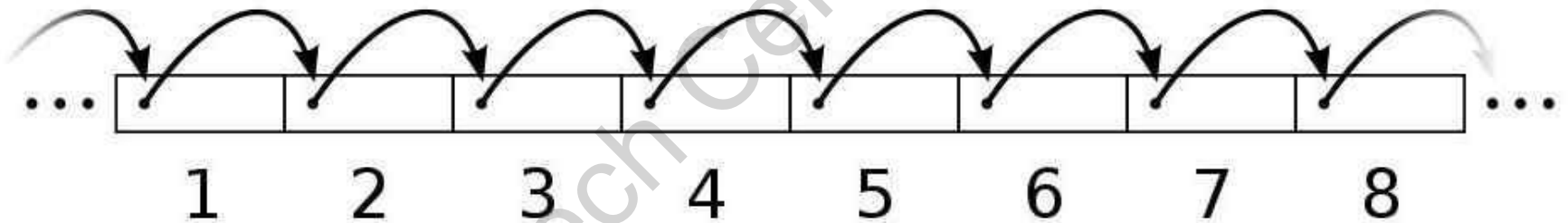
Sequential Access 1-4

□ Sequential access files:

- Records are arranged in the ascending, descending, or chronological order of a key field
- Storage location is not identified, since the records are ordered by a key field
- Are normally created and stored on magnetic tape using batch processing method
- Are used in applications like payroll management where each file in the record is processed

Sequential Access 2-4

Sequential access



Sequential Access 3-4

□ Advantages of sequential access are as follows:

It is simple to understand

It is easier to maintain and organize

In sequential access loading a record requires only the record key

A relatively inexpensive I/O media and devices can be used

It is easy to reconstruct the files

The proportion of file records to be processed is high

Sequential Access 4-4

❑ Disadvantages of sequential access are as follows:

To get specific information, the entire file must be processed

It stores very low activity rate

Transactions are required to be placed and stored in a sequence prior to processing

Data redundancy increases, as same data can be stored at different places with different keys

Random enquiries are impossible to handle

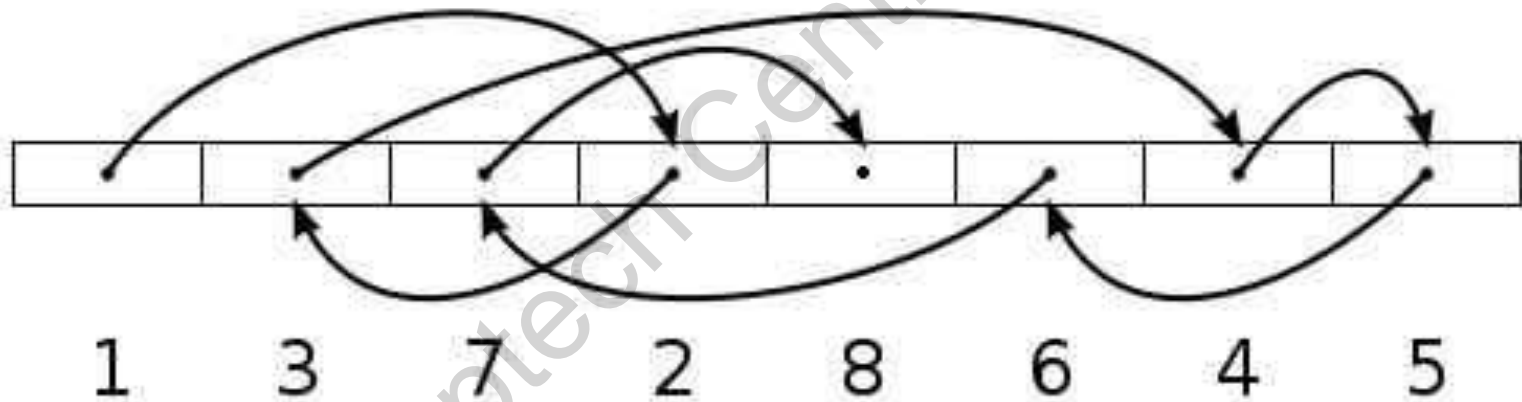


Direct Access 1-4

- ☐ The files are stored in direct access storage devices, using an identifying key
- ☐ The identifying key connects to its actual storage position in the file
- ☐ The computer can directly locate the key to identify the specific record without searching through the whole record
- ☐ In the online system, the response and updation are fast

Direct Access 2-4

Direct access





Direct Access 3-4

❑ Advantages of direct access are as follows:

Records can be immediately accessed for updation

Several files can be simultaneously updated during transaction processing

Transaction need not be sorted

Existing records can be amended or modified

Very easy to handle random enquiries

Most suitable for interactive online applications



Direct Access 4-4

☐ Disadvantages of direct access are as follows:

Data may be accidentally erased or over written unless special precautions are taken

Risk of loss of accuracy and breach of security. Special backup and reconstruction procedures must be established

Less efficient use of storage space

Expensive hardware and software are required

High complexity in programming

File updation is more difficult when compared to that of sequential method

Indexed Sequential Access 1-3

- ☐ The records are stored sequentially on a direct access device and the data is accessible randomly and sequentially
- ☐ It covers the positive features of both sequential and direct access files
- ☐ This type of file organization is appropriate for batch processing and online processing
- ☐ Indexing permits access to selected records without searching the entire file

Indexed Sequential Access 2-3

❑ Advantages of indexed sequential access are as follows:

Permits efficient and economic use of sequential processing technique when the activity rate is high

Permits quick access to records, in an efficient way

Indexed Sequential Access 3-3

- ❑ Disadvantages of indexed sequential access are as follows:

Slow retrieval, when compared to other methods

Does not use the storage space efficiently

Hardware and software used are relatively expensive

Different File Operations 1-5

□ Operations that can be performed on files are as follows:

- Creating a file
- Writing a file
- Reading a file
- Repositioning within a file
- Deleting a file
- Truncating a file

Different File Operations 2-5

❑ Creating a file

- Space in the file system must be found for the file
- An entry for the new file must be made in the directory

❑ Writing a file

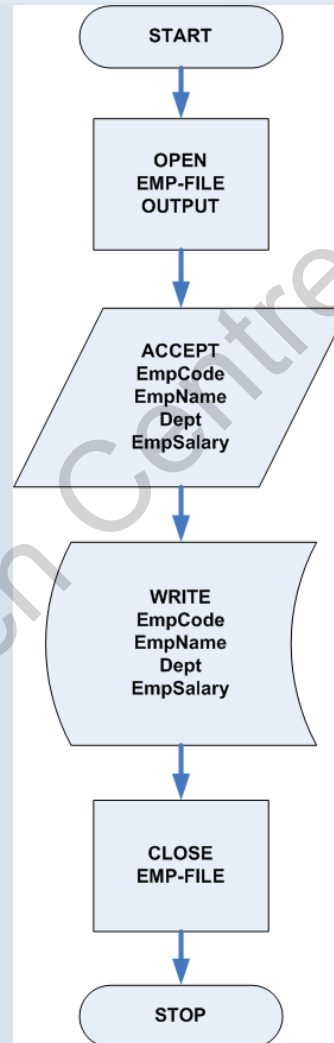
- To write a file, a system call is made specifying both the name and the information to be written to the file
- The system must keep a write pointer to the location in the file where the next write is to take place

Different File Operations 3-5

- ❑ Pseudocode shows a scenario to create employee record file and to write a record into it

```
START
OPEN EMP-FILE FOR OUTPUT
DISPLAY "Enter Employee code:"
ACCEPT EmpCode
DISPLAY "Enter Employee Name:"
ACCEPT EmpName
DISPLAY "Enter Employee Department:"
ACCEPT Dept
DISPLAY "Enter Employee Salary:"
ACCEPT EmpSalary
WRITE EmpCode, EmpName, Dept, EmpSalary
CLOSE EMP-FILE
STOP
```

Different File Operations 4-5



Different File Operations 5-5

☐ Repositioning within a file

- The directory is searched for the appropriate entry, and the current file position pointer is repositioned to a given value
- Repositioning within a file need not involve any actual I/O

☐ Deleting a file

- To delete a file, search the directory for the named file

☐ Truncating a file

- The user can erase the contents of a file, but keep its attributes
- Rather than forcing the user to delete the file, and then recreate it, this function allows all attributes to remain unchanged

Sequential Files and Control Break Logic

□ Sequential files

- Provides a straightforward way to read and write files
- Easy portability to other programming languages and computers
- Used as the common denominator of data processing

□ Sequential files with control break:

- A control break occurs when there is a change in the value of a single key on which a file is sorted for some added processing
- Control break processing is used when writing report programs manually



Summary 1-2

- ☐ A file is a collection of correlated data or collection of bytes stored on a secondary storage device, such as hard disk drive, a pen drive, and so on
- ☐ A text file is a stream of characters that a computer can process sequentially
- ☐ A programming language does not place constructs on the file, and it is read, or written, in a method selected by the programmer
- ☐ File organization refers to the relationship of the key of the record to the physical location of that record in the computer file



Summary 2-2

- ❑ There are three types of file organization methods, they are as follows:
 - Sequential
 - Relative
 - Indexed
- ❑ In sequential access searching the records are arranged in the ascending, descending, or chronological order of a key field which can be numeric or both
- ❑ In a computer program, a control break occurs when there is a change in the value of a single key on which a file is sorted for some added processing