# EDO UNIVERSITY IYAMHO
## Department of Computer Science
## CMP 255: File Organization and Management

**Introduction of Lecturer/ Lecture:**

John Temitope Ogbiti, email: Ogbiti.john@edouniversity.edu.ng Lectures: Monday, 8am – 10am, NLT1, phone: (+234) 8034883386

Office hours: Tuesday, 9AM to 3.30 PM, Office: FOS Room A010.

Grading: We will assign 10% of this class grade to home works & lecture notes, 10% for the programming projects, 10% for the mid-term test and 70% for the final exam. The Final exam is comprehensive.

**General overview of Lecture**

Logical files definition, labels (header and Tailer) Record blocking and deblocking, inter-related records gaps. Basic methods of file processing. Sequential, index sequential and random. Information storage & retrieval, information management applications, Information capture and representation, analysis & indexing, search, retrieval, information privacy; integrity, security; scalability, efficiency and effectiveness. Introduction to database system: Components of database system DBMS function, Database architecture and data independence use of database query language.

**Learning Outcome:** This course is intended to give the students a thorough knowledge of how file management techniques and tools are used in organization. To provide a detailed description of various ways organizational file are managed

## INTRODUCTION

File organization refers to the way data is stored in a file. File organization is very important because it determines the methods of access, efficiency, flexibility and storage devices to use. Basic methods of file processing, Sequential, index sequential and random. Information storage & retrieval, information management applications will be discuss.

**The Logic File System: -** is a research file system which replaces pathnames with expressions in propositional logic. It allows file metadata to be queried with a superset of the Boolean syntax commonly used in modern search engines. The actual name is the **Logic Information Systems File System**, and is abbreviated **LISFS** to avoid confusion with the log-structured file system (LFS).

### Header and Trailer Format

To place a header or trailer on the pages of a text message, you point to a block of data with a set format.

To place a header on the pages of a text message, you point to a block of data in the following format in the HEADER option:


You use the same format for trailer text, but you point to it with the TRAILER option. Here:

**LL**
is the length of the header (trailer) data, not including the four bytes of LL, P, and C characters. LL should be expressed in half word binary form.

**P**
is the page-number substitution character (see PNFLD below). Use a blank if you do not want page numbers.

**C**
is a reserved 1-byte field.

**TEXT**

is the header (trailer) text to be placed at the top (bottom) of each page of output. Use new-line characters ( X'15' ) to indicate where line breaks should occur if you want multiple lines.

**PNFLD**

is the page number field within your header (trailer) text. If you want to number the pages of your output, choose a character that does not otherwise appear in your header (trailer) text. Place this character in the positions where the page number is to appear. You can use from one to five adjacent positions, depending on how large you expect your page numbers to get (32,767 is the maximum BMS allows). Place the same character in the P field above, to tell BMS where to make the substitution. Do not use X'0C' , X'15' , X'17' , X'26' , or X'FF' for P; these values are reserved for other purposes. If you do not want page numbering, place a blank ( X'40' ) in P.

When you are building a logical message, you should repeat your HEADER and TRAILER options on each SEND TEXT command, so that they are present when the page breaks occur, and you need to specify the trailer again on the SEND PAGE command that terminates the message.

Here is an example of a COBOL definition for a header that numbers the pages, leaving room for a number up to 99.

```
EXEC CICS SEND TEXT FROM (OUTPUT-AREA)

HEADER(HEADER-TEXT) PAGING ACCUM END-EXEC.
```

where:

```
01 HEADER-TEXT
02 HEADER-LL PIC S9(4) COMP VALUE +11.
02 HEADP PIC X VALUE '@'.
02 FILLER PIC X VALUE LOW-VALUE.

02 HEADING PIC X(11) VALUE 'PAGE NO. @@'.
```

Screens built with SEND TEXT are not designed for extensive input from the terminal operator. However, you can interpret the attention identifier and read simple inputs—such as those used in the CSPG transaction to control the page display—if the field structure on the screen is suitable and the operator knows or can see what is expected. (A new field starts at each line, as well as at the first

character of the text sent with each SEND TEXT command that made up the message. The fields defined are unprotected, alphabetic and normal intensity, so that the operator can key into them.) Normally a terminal control RECEIVE is used in this situation; you can use RECEIVE MAP only if you can build a map with a field structure matching that of the screen.

**A Blocking Record:-** is data communicated from one password synchronization server (A) to every transparent password synchronization server (B, C, ...) on a network. The Blocking Record indicates that (A) is actively engaged in synchronizing passwords for a given user (U), and that transparent password synchronization servers (B, C, ...) should ignore any events that would otherwise trigger transparent password synchronization for that user (U).

## Blocks and records

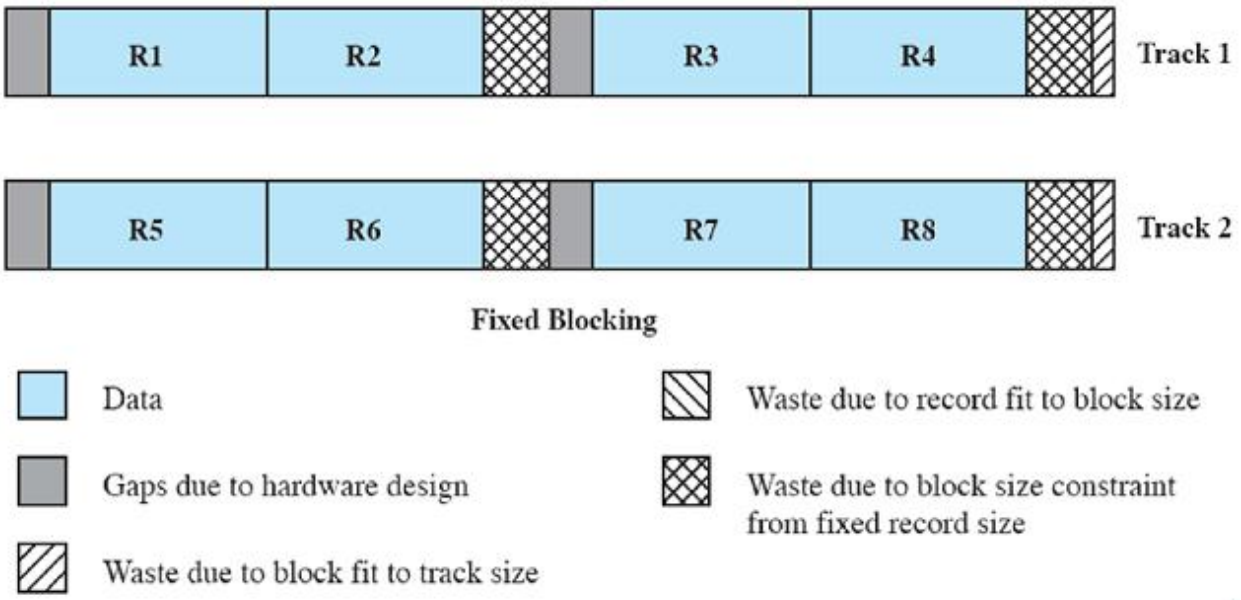
Records are the logical unit of access of a structured file, Blocks are the unit for I/O with secondary storage. For I/O to be performed, records must be organized as blocks. Three methods of blocking are common

– Fixed length blocking
–Variable length spanned blocking
– Variable-length in spanned blocking

### Fixed Blocking

• Fixed-length records are used, and an integral number of records are stored in a block
• Unused space at the end of a block is ***internal fragmentation***
• Common for sequential files with fixed length records

## Fixed Blocking

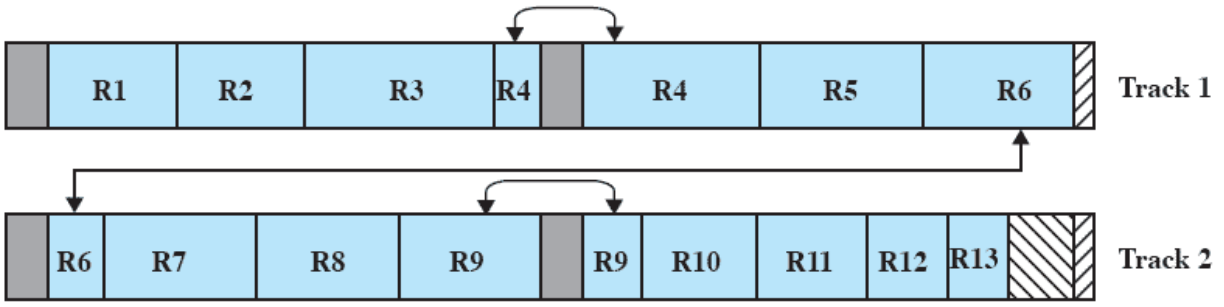



### Variable Length Spanned Blocking

- Variable-length records are used and are packed into blocks with no unused space
- Some records may span multiple blocks
  – Continuation is indicated by a pointer to the successor block
- Efficient for storage and does not limit the size of records

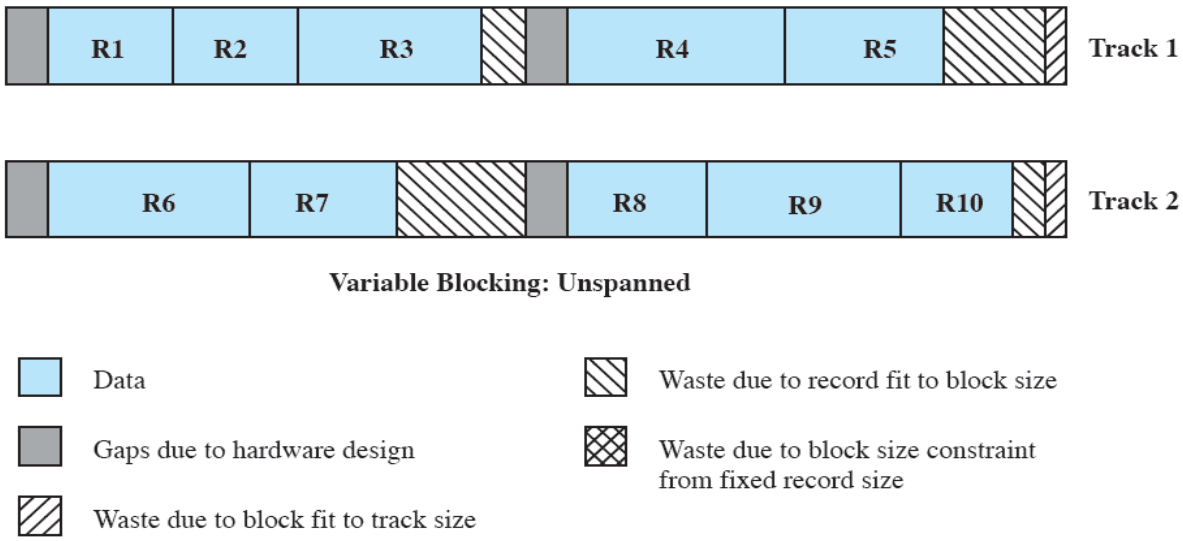
### Variable Blocking: Spanned

Difficult to implement
Records that span two blocks require two I/O operations




**Variable-length in spanned blocking:**

- Uses variable length records without spanning
- Wasted space in most blocks because of the inability to use the remainder of a block if the next record is larger than the remaining unused space
- Limits record size to the size of a block




- The basic operations that a user or application may perform on a file are performed at the record level
  - The file is viewed as having some structure that organizes the records
- File organization refers to the logical structuring of records
  - Determined by the ***way*** in which files are accessed (access method)

## File Organization and access

- The basic operations that a user or application may perform on a file are performed at the record level
  - The file is viewed as having some structure that organizes the records
- File organization refers to the logical structuring of records
  - Determined by the ***way*** in which files are accessed (access method)

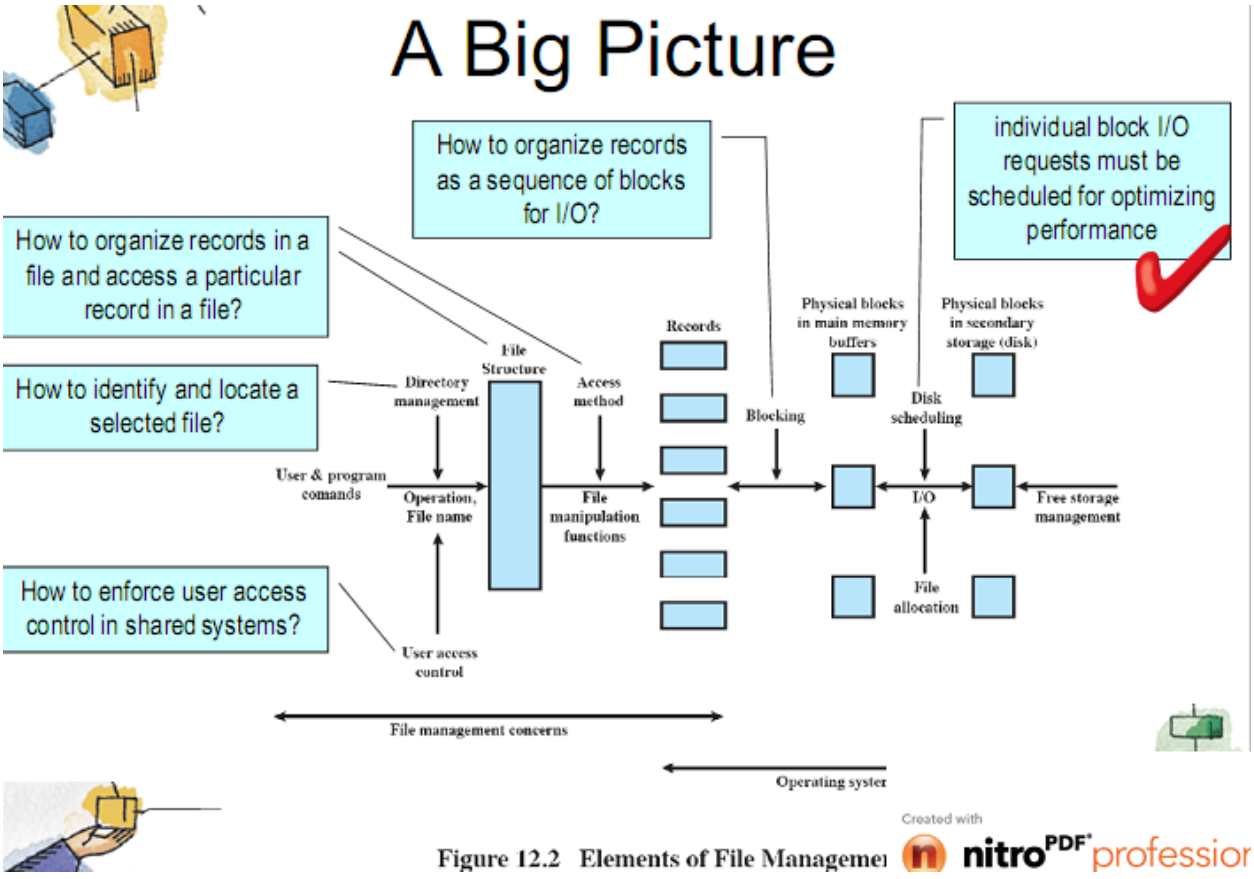



Figure 12.2 Elements of File Managemei



## Criteria for File Organization

Important criteria include:

- Short access time
- Ease of update
- Economy of storage
- Simple maintenance
- Reliability

Priority will differ depending on the use

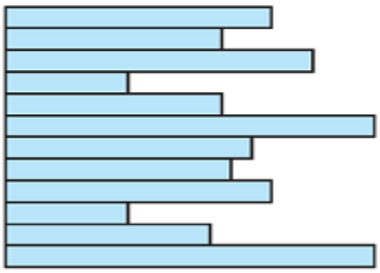
- For batch mode file processing, rapid access for retrieval of a single record is of minimal concern

These criteria may conflict

- Use of indexes (conflict with economy of storage) can be a primary means of increasing the speed of access to data

**The Pile**

- Data are collected in the order they arrive
  - No structure
- Purpose is to accumulate a mass of data and save it
- Records may have different fields
  - field should be self-describing (field name + value)
  - field length should be known (delimiters, subfield or default for a field type)



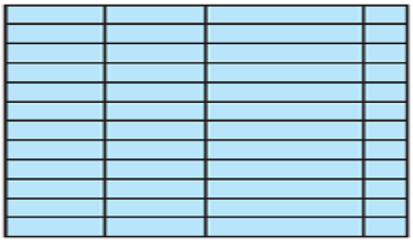
Variable-length records
Variable set of fields
Chronological order

- Record access is by exhaustive search
- Used when data are collected and stored prior to processing or data are not easy to organize
- Uses space well when data vary in size and structure
- Adequate for exhaustive searches
- Easy to update
- Unsuitable for most applications

**The Sequential File**

- Fixed format used for records
- Records are of the same length

  - same number of fixed-length fields in a particular order
- Only the values of fields need to be stored
- Field name and length are attributes of the file structure



Fixed-length records
Fixed set of fields in fixed order
Sequential order based on key field

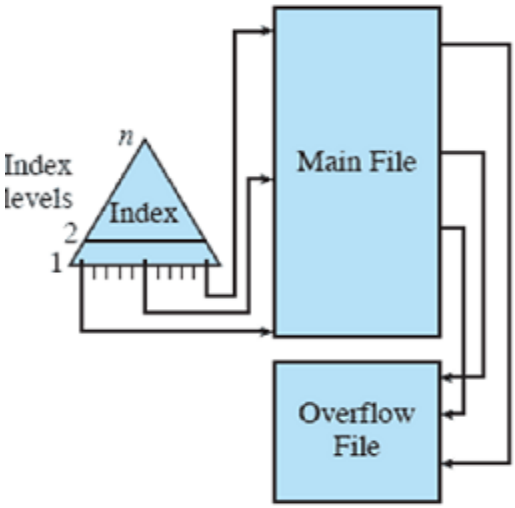
- Key field
  - Uniquely identifies the record
  - Records are stored in key sequence
- Optimal for batch applications if they involve the processing of all the records
- Easily stored on tape and disk
- Poor performance for interactive applications
  - considerable processing and delay due to the sequential search of the file for a key match

**Indexed Sequential File**

An index is added to support random access

- An index record contains a key field and a pointer into the main file
- The index is a sequential file
- For searching

- Search the index to find the highest key value that is equal to or precedes the desired key value
- Search continues in the main file at the location indicated by the pointer



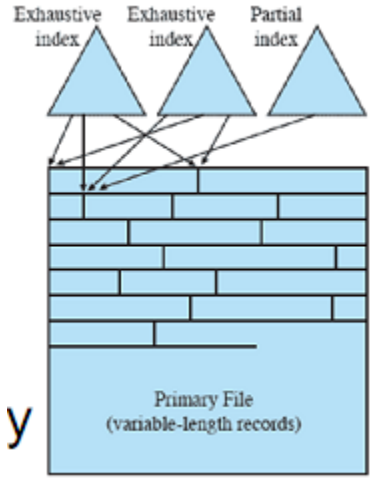

**Indexed Sequential File Example:**

- Consider searching a particular key value in a sequential file with 1 million records
  - without index
    - requires on average one-half million record accesses
  - with an index containing 1000 entries with the keys in the index evenly distributed over the main file
    - requires on average 500 accesses to the index file + 500 accesses to the main file

- An overflow file is added
  - A new record is added to the overflow file and is located by following a pointer from its predecessor record
  - The indexed sequential file is occasionally merged with the overflow file in batch mode
- Greatly reduces the time required to access a single record, without sacrificing the sequential nature.

- Records are accessed only through their indexes
  - no restriction on the placement of records
  - allows variable-length records
- Uses multiple indexes for different key fields
  - An exhaustive index contains one entry for every record in the main file

- A partial index contains entries to records where the field of interest exists




- When a new record is added to the main file, all of the index files must be updated.
- Used mostly in applications where
  - timeliness of information is critical and
  - data are rarely processed exhaustively
  - examples: airline reservation systems and inventory control systems

**Information Storage and Retrieval System (ISAR)**

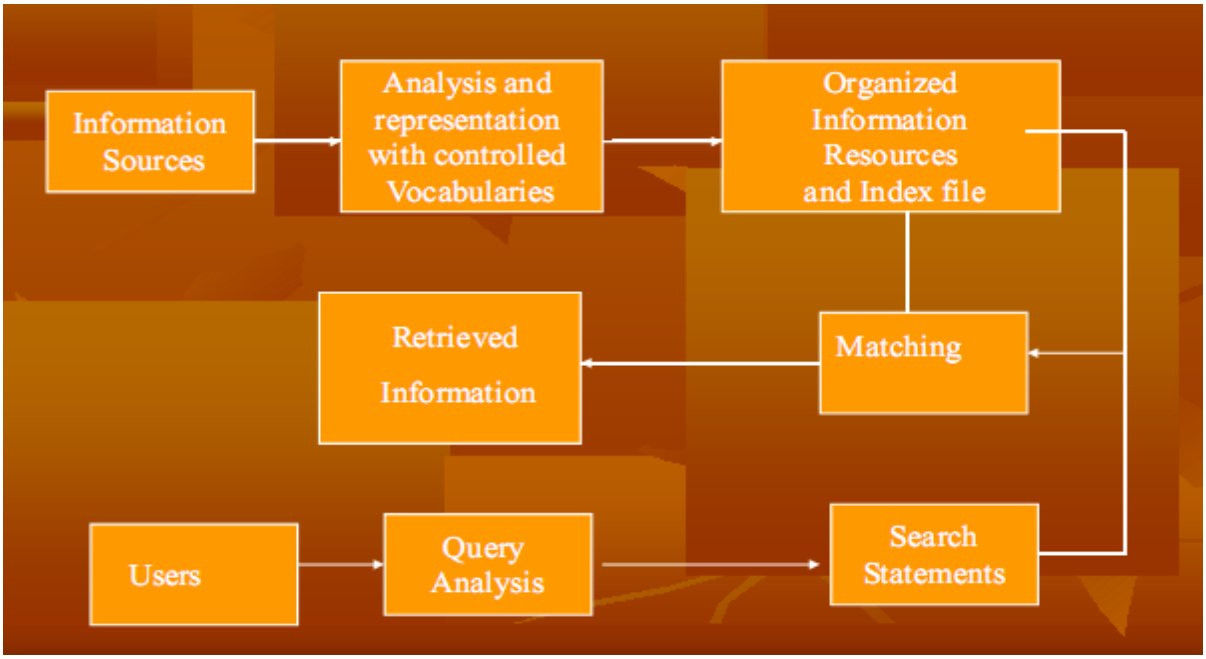
**Information retrieval** is the science of searching for information in documents, searching for documents themselves, searching for metadata which describe documents, or searching within databases, whether relational searching within databases, stand stand-alone databases or hypertext networked alone databases such as the Internet or World Wide. Data bases such as the Internet for text, sound, images, Web or data.

**Purpose of ISAR**

- A writer presents a set of ideas in a document using a set of concept.
- ISAR system serves to match the writer's ideas expressed in user's requirement or demand those.

- ISAR system serves as a bridge between the generator of information and the users of that information.

**Broad outline of an Information and Retrieval Systems (IRS)**




**Information system** is an organized combination of people, hardware, software, communications Networks and data resources that collects, transforms, and disseminates information in an organization.

**Types Of Management Information System**

- **Transaction processing systems:** These systems process a large volume of routine, recurring transactions.
- **Operations information systems:** These systems gather comprehensive data, organize it and summarize it in a form that is useful for managers.
- **Decision support systems:** These systems help mangers with the necessary information to make intelligent decisions.

- **Expert systems:** They are meant to mimic humans in making decisions in a specific field.

**Management Information Systems**

Management Information Systems (MIS) are information systems, typically computer-based, that are used within an organization. WorldNet describes an information system as "a system consisting of the network of all communication channels used within an organization". A management information system may also be defined as "a system that collects and processes data (information) and provides it to managers at all levels who use it for decision making, planning, program implementation, and control." An information system is comprised of all the components that collect, manipulate, and disseminate data or information. It usually includes hardware, software, people, communications systems such as telephone lines, and the data itself. The activities involved include inputting data, processing of data into information, storage of data and information, and the production of outputs such as management reports.

As an area of study it is commonly referred to as information technology management. The study of information systems is usually a commerce and business administration discipline, and frequently involves software engineering, but also distinguishes itself by concentrating on the integration of computer systems with the aims of the organization. The area of study should not be confused with [Computer Science] which is more theoretical in nature and deals mainly with software creation, or [Computer Engineering], which focuses more on the design of computer hardware.

## The concept of management information systems

The management information system is not something distinct and separate from other information systems. It provides a general framework based on the information systems that are compatible with each other. Over time it became clear that it is very difficult to implement the concept of a fully integrated system.

## HOW DOES AN INFORMATION SYSTEM WORK?

Information systems perform three major activities. First, the sources of information within the organization or outside the organization receive as input. The work done on the received information allows the system to produce the desired information. Finally, the user-generated information intended for a manager or an employee puts.

For example: A computerized information system that is doing savings accounts in bank. Bank clerks enter data and then processing and storage, shall be printed and delivered to the customer. The customer can also receive reports from the operating account. (Mentzas, 1995)

## WHY THE INFORMATION SYSTEMS ARE USED?

Usually, there are two approaches in the use of information systems. These two approaches are both originated from the use of computers and IT in the organization's activities. Proponents of these approaches seek to exploit ways of using information systems, computer and communication technology.

First Approach: This approach is focused exclusively on the inherent capabilities of computer and communication technology and how it can be used to improve efficiency. So there is emphasis on improving system performance

through process efficiency and reliability of performance, not the specific use of the capabilities of computer and communications technology.

Second approach: This approach focuses on the organization's strengths and opportunities and assessing how the use of information technology the situation. This approach takes into account the enabling technology. In other words, Instead of having a technology-driven approach, Followed by the development of information technology systems. This approach is better because the computer and communications technology is important but how the computer can assist the organization to be important.

There are generally three types of systems:

1 - Operating Systems

2 - Information Systems

3 - Systems Management

Information systems that In this section, we will describe them, Interface between operating systems and management who actually made it through the operation of the organization's managers can be notified.

**Types of Information Systems**:

• Transaction Processing Systems (TPS)

• Management Information System (MIS)

• Decision Support System (DSS)

• Executive Support Systems (EIS)

• Strategic Information System (SIS)

• Accounting Information System (AIS)

The following brief description of each of the systems we mentioned.

**Transaction Processing Systems (TPS)**

Processing system has the task of the organization's events that are recorded. The MIS, records and stores information about the queue and Chiefs organization collects, processes and provides them in the form of documents, reports, or information systems and decision support information.

**Management Information System (MIS)**

MIS is the most popular information systems.

MIS receives internal data from the operation processing system and summarized in the form of meaningful and useful as management reporting. To be used when performing administrative tasks such as controlling and decision making.

The aim of Management Information Systems is to enhance the presentation and reduce speculation in resolving problems at different organizational levels.

**Decision Support System (DSS)**

Decision Support System is defined as a computer-based system to be used by a manager or group of managers at any organizational level decision process for problem solving

DSS, helps decision makers with putting together human judgment and computerized information to solve .A DSS can help the decision maker, but they are never replaced.

**Executive Support Systems (EIS)**

Which is a special type of DSS can help the decision-making high levels .Accurate picture of the performance of the system and shows a summary of the activities of competitors. The system is easy to work with, because they provide information in a way that can be easily downloaded (as a graphic and charts).

In summary, the purpose of the EIS is to support the management of the

information supplied in accordance with operational managers.

**Accounting Information System (AIS)**

AIS System describes the data that collect organization's activities and convert the data into information. Gives the available information to internal and external organization users. The AIS is a system that provides information for various units.

**(SIS) Strategic Information System**

SIS is one of the most important information systems applications, organization, management, politics, the military, business and strategic information is provided for supplying to achieve its strategic objectives. Mainly in terms of strategic information systems are complex, challenging, unpredictable, and chaotic critical applications

**The benefits of using information systems**

Three kinds of information systems have the potential to provide benefits to the organization (Young, 1983)

(1) Improve Productivity

(2) Improve the Effectiveness

(3) Competitive Advantage

**Productivity improvement:**

Productivity improvement occurs when the work can be done more or less the same source. In Organizations, improve productivity happen when improve work processes occur. For example, For example, when an employee may be replaced with new methods of processing 25 Orders to 50 Orders, His labour productivity is 2 fold. Information systems can be faster, easier and more effective to do things that are improving productivity in this way.

**Effectiveness:**

Effectiveness of ability of an individual or an organization to do things that should be done. Manager predicts the conditions that may be problematic terms before it causes problems arise; effective manager who joined the badly problems that had prevented them. Information systems, which provide information to help managers evaluate the situation and choose the best options and there by improve the effectiveness.

**Competitive advantage:**

The organization that has improved the efficiency and effectiveness of the use of information systems to be able to have the potential to transform the way organizations compete. (Young, 1983)

**Success factors in the implementation of top level manager's information systems:**

high manager's information systems is management information systems, which makes information easy to management decision-making at the highest levels. These systems are designed based software that provides easy access to large amounts of complex data. Through its software, analysis data in a quick manner and then display. With quick access to internal and external information systems, senior managers are prepared to make strategic decisions.

**Key factors for success**

The factors that are considered as key success factors in implementing an information system are as follows:

1- Manager who is familiar with the information system

2- Project Manager

3- Enterprise Information Systems IS appropriate box

4- The right technology

5- Data Management

6- Communicate clearly with business goals

7- Management of organizational strength

8- Manage the development and expansion of the system

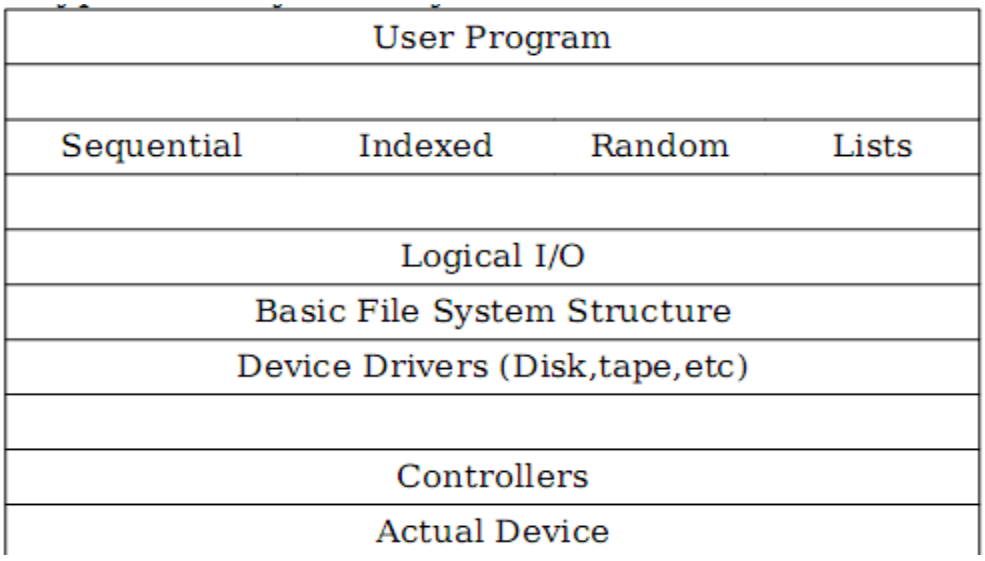
9- Evolutionary development

10- Definition of informational needs of users

## Functional Components of a DBMS

A typical DBMS has the following six basic components:

1. **File Manager** manages the allocation space on disk storage and the data structures used to represent info stored on other media. In most applications (99.9%) the file is the central element. All applications are designed with the specific goal: generation and use of information. A typical file system layered architecture is the following (see also CPS510).




**2. Buffer Manager** among other tasks, it transfers blocks between disk (or other devices) and Main Memory (MM). A DMA (Direct Memory Access) is a form of I/O that controls the exchange of blocks between MM and a device. When a processor receives a request for a transfer of a block, it sends it to the DMA which transfers the block uninterrupted.

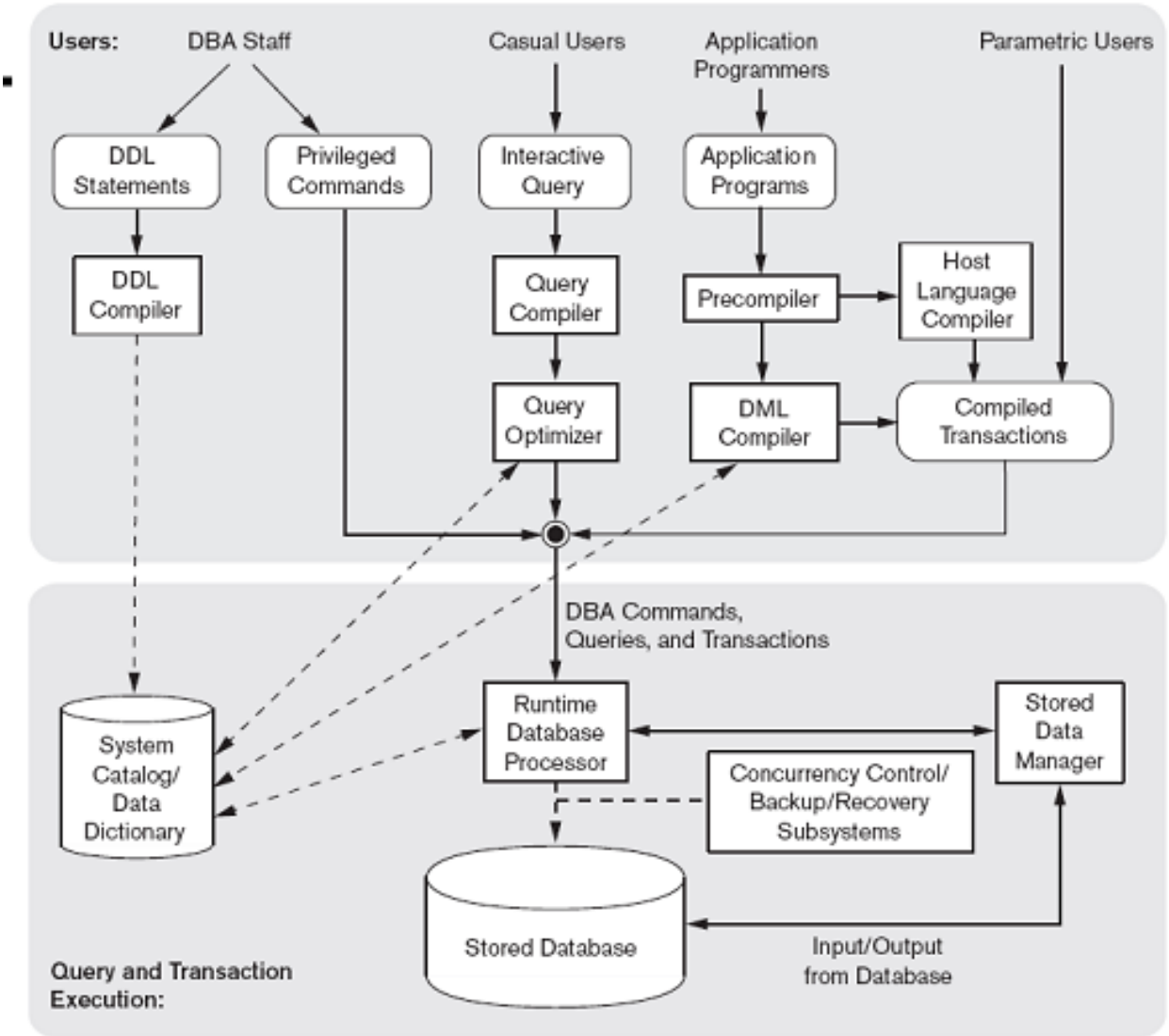
**3. Query Parser** translates statements in a query language, whether embedded or not, into a lower level language. (See RL language example from CPS510). This parser is also a strategy selector: i.e., finding the best and most efficient way (faster?) of executing the query.

**4. Authorization and Integrity Manager** Checks for the authority of the users to access and modify info, as well as integrity constraints (keys, etc.).

**5. Recovery Manager** ensures that the database is and remains in a consistent (sound) state after any kind of failure.

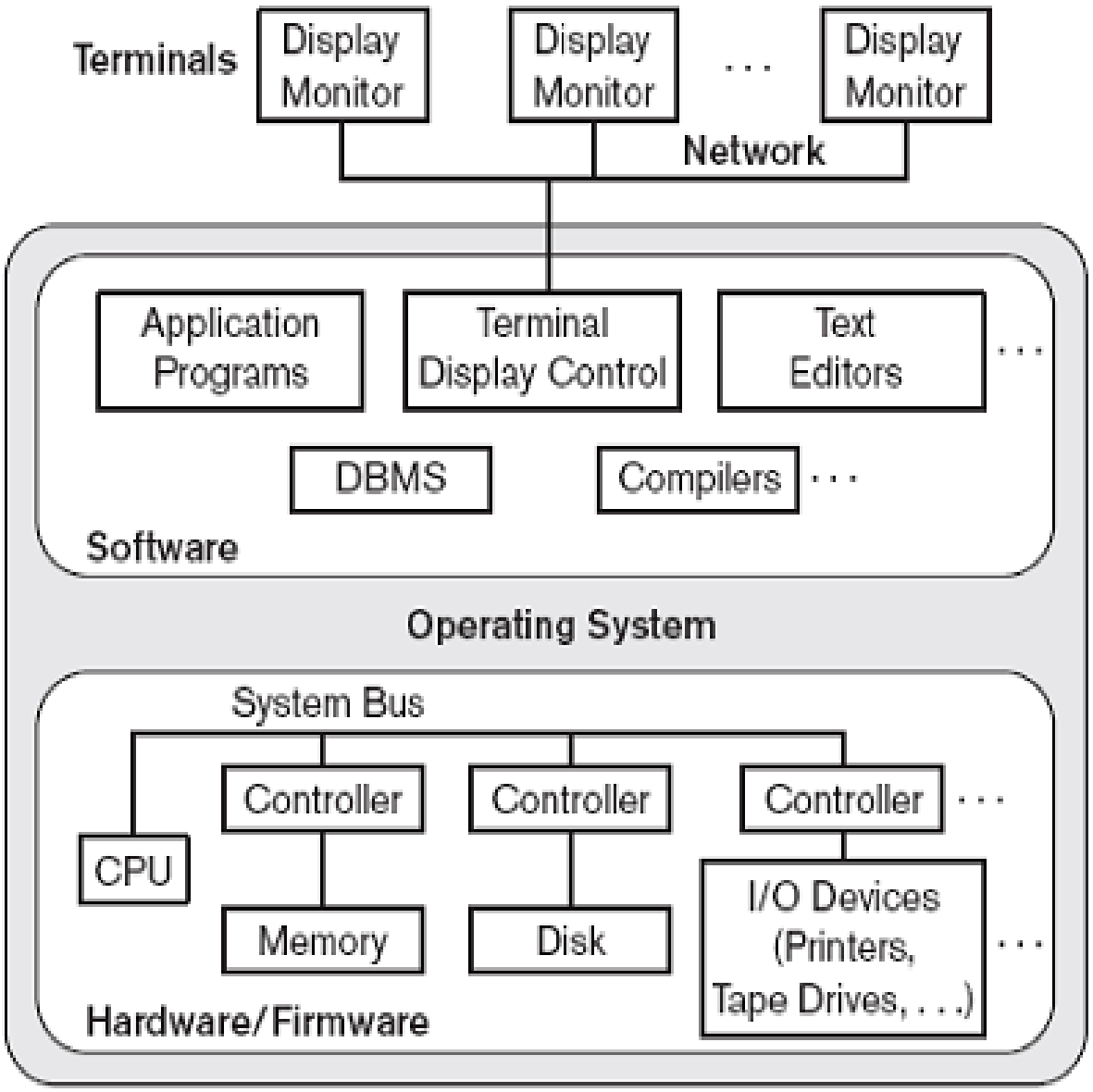
**6. Concurrency Controller** enforces Mutual Exclusion by ensuring that concurrent interactions with the data base proceed without conflict (deadlocks, etc.).

**Component model of a DBMS there Interactions**



**CENTRALIZED DBMS ARCHITECTURE**

All DBMS functionality, application program execution, and user interface processing carried out on one machine



# The Information Security Triad: Confidentiality, Integrity, Availability (CIA)

### Confidentiality

When protecting information, we want to be able to restrict access to those who are allowed to see it; everyone else should be disallowed from learning anything about its contents. This is the essence of confidentiality. For example, federal law requires that universities restrict access to private student information. The university must be sure that only those who are authorized have access to view the grade records.


## Integrity

Integrity is the assurance that the information being accessed has not been altered and truly represents what is intended. Just as a person with integrity means what he or she says and can be trusted to consistently represent the truth, information integrity means information truly represents its intended meaning. Information can lose its integrity through malicious intent, such as when someone who is not authorized makes a change to intentionally misrepresent something. An example of

this would be when a hacker is hired to go into the university's system and change a grade.

## Availability

Information availability is the third part of the CIA triad. *Availability* means that information can be accessed and modified by anyone authorized to do so in an appropriate timeframe. Depending on the type of information, *appropriate timeframe* can mean different things. For example, a stock trader needs information to be available immediately, while a sales person may be happy to get sales numbers for the day in a report the next morning. Companies such as Amazon.com will require their servers to be available twenty-four hours a day, seven days a week. Other companies may not suffer if their web servers are down for a few minutes once in a while.

### Tools for Information Security

In order to ensure the confidentiality, integrity, and availability of information, organizations can choose from a variety of tools. Each of these tools can be utilized as part of an overall information-security policy, which will be discussed in the next section.

## Authentication

The most common way to identify someone is through their physical appearance, but how do we identify someone sitting behind a computer screen or at the ATM? Tools for authentication are used to ensure that the person accessing the information is, indeed, who they present themselves to be.

## Access Control

Once a user has been authenticated, the next step is to ensure that they can only access the information resources that are appropriate. This is done through the use of access control. Access control determines which users are authorized to read, modify, add, and/or delete information. Several different access control models exist. Here we will discuss two: the access control list (ACL) and role-based access control (RBAC).

**Sidebar: Password Security**

So why is using just a simple user ID/password not considered a secure method of authentication? It turns out that this single-factor authentication is extremely easy to compromise. Good password policies must be put in place in order to ensure that passwords cannot be compromised. Below are some of the more common policies that organizations should put in place.

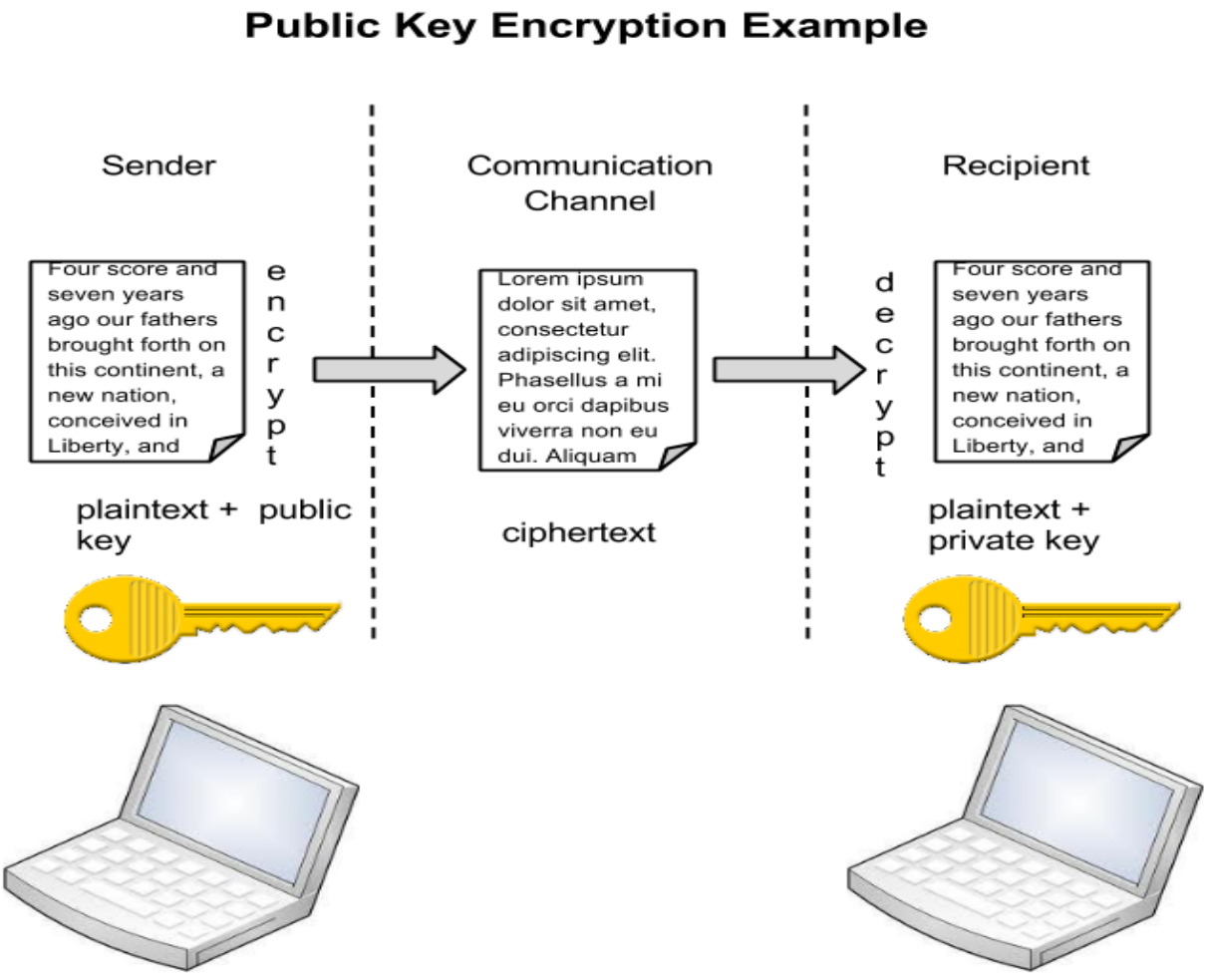
## Backups

Another essential tool for information security is a comprehensive backup plan for the entire organization. Not only should the data on the corporate servers be backed up, but individual computers used throughout the organization should also be backed up. A good backup plan should consist of several components.

## Encryption

Many times, an organization needs to transmit information over the Internet or transfer it on external media such as a CD or flash drive. In these cases, even with proper authentication and access control, it is possible for an unauthorized person to get access to the data. Encryption is a process of encoding data upon its

transmission or storage so that only authorized individuals can read it. This encoding is accomplished by a computer program, which encodes the plain text that needs to be transmitted; then the recipient receives the cipher text and decodes it (decryption). In order for this to work, the sender and receiver need to agree on the method of encoding so that both parties can communicate properly. Both parties share the encryption key, enabling them to encode and decode each other's messages. This is called symmetric key encryption. This type of encryption is problematic because the key is available in two different places.




**Public key encryption (click for larger diagram)**

## Summary and Conclusions

File organization is very important because it determines the methods of access, efficiency, flexibility and storage devices to use. The management information system is not something distinct and separate from other information systems. It provides a general framework based on the information systems that are compatible with each other. Over time it became clear that it is very difficult to implement the concept of a fully integrated system.

## Interactions and Questions

1. (a) Discuss Information Retrieval and The Logic File System

   (b) With the aid of a diagram illustrate how an Information and Retrieval Systems (IRS) operates
2. (a) Illustrate with aid of diagram a Centralized Database Management System architecture

   (b) Explain the term Management Information System
3. Discuss in details five (5) tools used for information Security

## Further Reading

**Textbook**: The recommended textbook for the class are:
Title: Theory book -File Organization and Management
Authors: IkeMag – away Gaami
Publisher:
ISBN-0471-69466-5
Year: 2005

Online materials