

## INFORMATION TECHNOLOGY FOR STATISTICS

### Topic summary

#### 1.Role of IT and computers in todays society

Information technology - is the use of computers,software,networks and other digital technologies to manage,process,store,and transmit information.

### Roles of IT

#### a)In education

Students are taught computer skills and asked to incorporate into their daily work assignment

#### b)In business

Enables more efficient operations by allowing workers to do a wider variety of work

#### c)In industries

Enables personnel management,product design and manufacturing to shipping

#### d)In government

Enables tallying of people (population)

Calculation and filling returns done via the internet (taxation)

Calculating trajectory of missiles (military)

#### e)Healthcare

Manages schedules,maintain patients records, and perform billing

#### f)marketing

Enables businesses show products online

#### g)Entertainment

There are online videos sports and internet games

### 2.Fundamentals of computer operations

It describes how a computer processes information and executes tasks

#### a)Input

Gets data into the computer

Devices used include;keyboard, sensors, and network interfaces

#### b)Processing

It manipulates,calculate, and transform input data based on instructions.

Components involved are:

\*CPU -executes instructions,performs arithmetic operations

\*Memory -temporarily holds data and instructions the CPU needs

\*ALU(Arithmetic logic unit) -part of the cpu that handles math and logical comparisons

c)Output

Presents the processed data to the user or another system.

Devices used:

\*monitor,speaker,printers -display/produce human readable output

\*network interfaces -send data to other devices

\*storage devices -save data for future use

d)Storage

Retains data and programs even when the computer is off

\*primary storage -volatile,temporary,fast access for active data

\*secondary storage -non-volatile,persistent storage for files,OS,and apps

e)Control unit

Manages and coordinates the sequence of operation

\* It directs data flow between components

\*Controls instruction execution timing and sequence

Input devices=feed data

cpu=process the data

output=presenr results

storage=holds the data

### 3.Basics of computer hardware and software

#### *Hardware*

*Refers to the physical components of a computer system*

a)CPU

*Executes instructions,performs calculations, and manages data flow*

*Key specs: core,clock speed,cache size(e.g Intel core i5, AMD Ryzen 7)*

b)Memory(RAM)

*Temporary storage for data and instructions the cpu needs while working*

c)Storage devices

*Holds data and software persistently, even when the computer is off*

\*Hard disk drive(HDD)

Spinning disks.mechanical,large capacity,slower

\*Solid state drive(SDD)

Flash memory,no moving parts,faster and more durable

\*NVMe SSD

High-speed SSDs using NVMe protocol

d)Motherboard

Main circuit board connecting all hardware components and manages data flow

Key features include:cpu sockets,RAM slots, expansion slots, USB e.t.c

e)Power supply unit

Converts AC power to DC power and supplies the power to the motherboard

f)Input devices

Sends data to the computer

g)Output devices

Presents processed data

h)Expansion cards

Adds extra functionality e.g graphics card, network card, and sound card

Software

Refers to the programs,apps and instructions that tell the hardware what to do

a)System software

Manages the hardware and provides a base for running apps

\*operating system(OS)

Core software that manages hardware, files and apps

\*device drivers

Specialized software that lets the OS communicate with specific hardware

\*utilities

Helps manage the system

b)Application software

Programs designed for users to perform specific tasks

\* Productivity software

Processors,spreadsheets,presentations

\*creative software

Photo editing,video editing,graphic design

\*web browsers

Access online content

\*games

## Entertainment software

c)Firmware

Software embedded in a hardware device

Critical for hardware initialization

#

Hardware needs software to function:the OS tells the CPU,RAM, and other hardwares how to work and manage tasks

Software relies on hardware to execute:apps run using the cpu,memory, and storage

## 4.Data files

It involves creating structured files to store data in a format that is readable and usable by software or humans.

a)CSV (comma-separated values)

A plain text values separated by commas

Good for tabular data e.g;

Name,age,course,grade

Tracy,19,financial engineering,A

### \*STRUCTURE

First line, column headers

Each subsequent line a record

Use commas to separate values;if a value contains a comma,enclose it in quotes  
(e.g “Alvin,Asingo”,19,financial engineering,A)

### \*creation

Open a text editor,type the data following the CSV structure, and save with a.csv extension

b)JSON(Javascript object notation)

Key values pairs,structured in objects

Good for hierarchical data

### \*STRUCTURE

Data is in{} with key value pair

Multiple values are in array separated by commas

It requires double quotes around keys and string values

No commas after the last element in an object

```
[  
 {  
   "Name": "otieno frank".  
   "Age": "19",  
   "Course": "financial engineering",  
   "grade": "A"  
 },  
 ]
```

After creation save with a .json extension

c) Plain textfile(TXT)

Simple line by line text

Has one record per line separated by commas

After creation save with a.txt extension

E.g

Name: Mary Kelcy, Age:19, Course:Financial engineering

\*\*constructing data files

a) use the same structure for all records ,pick a delimiter and stick to it

b) include headers to describe fields

c) wrap fields with commas or quotes in double quotes.use proper escaping for special chars

d) validate JSON/XML with tools to check syntax

e) save text files with UTF-8 encoding for international characters

## 5. Means of disk storage

Disk storage refers to data storage technologies that use disks or platters to store digital data

### a) Hard Disk Drive

Uses spinning magnetic disks and mechanical read/write heads

Data is stored magnetically on the platters; the read head moves across the spinning disk to access data

\*\* characteristics

\* capacity-high(500gb-16tb)

\* speed-slower than SSDs

\* cost-generally cheaper per GB than SSDs for large capacities

\* durability-susceptible to mechanical failure, shock, and wear over time

\*\* use cases; bulk storage/often as secondary storage

b)Solid State Drive(SSD)

Uses flash memory

Data is stored in interconnected flash memory chips

\*\*characteristics

\*capacity -medium to high(128GB-8TB)

\*speed -much faster than HDDs

\*cost -more expensive per GB than HDDs

\*durability -resistant to physical shock, faster and more durable than HDDs

c)Hybrid Disk(SSHD-solid state hybrid drive)

Combines a traditional HDD with a small amount of SSD cache

Frequently accessed data is stored on the SSD part for faster access bulk data goes to the HDD part

\*\*characteristics

Capacity-like HDDs with a small SSD cache

Speed-faster than pure HDDs for frequently accessed files

Cost-cheaper than pure SSD but more expensive than HDDs

Balancing cost and speed good for laptops needing moderate performance without paying for full SSD

d)Optical Disks(CD,DVD,Blu-ray)

Use laser read optical media, data is stored as tiny pits and lands on the disk surface

A laser reads the reflection pattern of the disc surface to retrieve data

\*\*characteristics

Capacity -CD 700MB

-DVD4.7GB-9GB

-Blu-ray25GB

Speed -much slower than HDDs/SSDs

Cost - cheap per disk

Durability-prone to scratches, degradation over time

Software distribution movie/video distribution

e)External/portable disk drive

Usually HDDs or SSDs in an external enclosure, connected via USB or thunderbolt

\*\*characteristics

Portability-designed for mobility; often rugged or compact

capacity/speed-can be HDD or SSD

connection-USB3.0,USB-C Thunderbolt

Used for transferring data between computers

## 6.Data files and file management

Data file are computer files that store data in a structured or unstructured format

Types of data files

a)structured data files

Data organized with a defined scheme

b)semi-structured data files

Data in mixed structure and unstructured elements

c)unstructured data files

No pre-defined structure

Common data file formats

CSV(students.csv)

JSON(data.json)

Texts(log.txt)

File management

Involves organizing,storing,retrieving,and manipulating data files on a computers storage systems

a)file organisation

Files are stored in directories and subdirectories

Use meaningful names

Some OSs allow tagging files for better searchability

b)file operation

Make a new file

View the file

Modify the file content

Duplicate a file

c)file access control

controls who can read ,write ,or execute files

d)file compression and archiving

Saves space or bundle multiple files

e)file backup and versioning

Protects against data loss

f)file integrity and encryption

Verifies file authenticity

protects sensitive data

## BINARY COMPUTER STORAGE FOMAT

Bits and Bytes

Bit-the smallest unit

Byte-a group of 8 bits

A computer stores its data in bits

A word is a group of bytes

word=4bytes

1024bytes=1 KB

1024KB=1MB

1024MB=1GB

1024GB=1TB

1024TB=1PB

1024PB=1EB

1024EB=1ZB

1024ZB=1YB