

Table of contents

Table of contents	1
?Tips	3
Cloud Computing	3
Cloud Deployment Models (4)	3
IT building blocks & NFA	4
Edge computing	4
C4: Networking	8
Network Topologies	8
①Physical Layer	8
②Data Link Layer	9
③Network Layer	9
④Transport Layer	10
⑤Session Layer	10
⑦Application Layer	11
4.2 Networking Virtualization	11
4.3 Networking: Availability, Performance & Security	12
Networking Availability	12
Networking Performance	13
Networking Security	14
C5: Storage	15
①Disks – command sets	15
①Disks - type (2) + Tape (2)	15
Mechanical Hard Disk (HDD)	15
Solid State Drive (SSD)	15
②Tape (2)...	15
③Controller Implementation	16
RAID (Redundant Array of Independent Disks)	16
Compression	16
Data Deduplication	17
Cloning	17
Snapshot	17
Thin Provisioning	17
④Storage Architectures (4)	17
⑤Software Defined Storage (SDS) (...Software Defined Networking)	18
5.2 Storage: Availability, Performance & Security	18
① Improve Storage Availability (4)	18
② Storage Performance (3)	19
③ Improve Storage Security(2)	19
C6: Compute	20
6.1 Compute building blocks	20
6.2 Compute Virtualization	21
6.3 Types of Compute	22
6.4 Compute: Availability, Performance & Security	23
C7: Operating systems	26

BMIS2113 Information Technology Infrastructure

7.1 Popular Operating Systems	26
7.2 Operating systems building blocks	26
7.3 Operating systems: Availability, Performance & Security	27
C8: End User Devices	28
8.1 End user devices building blocks	28
8.2 Desktop virtualization	28
8.3 End user devices: Availability, Performance & Security	29
C9: Infrastructure Management	31
9.1 Infrastructure Deployment Options	31
Infrastructure/ Cloud deployment models (4)	31
On-Premises	31
Public Cloud	31
Private Cloud	31
Hybrid Cloud	31
9.2 Infrastructure Automation	31
Configuration Management Tools	31
Orchestration Tools	32
9.3 Infrastructure Documentation	32
Infrastructure documentation tools / techniques	32

?Tips

-cloud computing

Cloud Computing	model that provide on-demand network access to a shared pool of configurable computing resources , users can easily access and use shared IT resources from anywhere, anytime, without managing the infrastructure themselves
Objective	Enables ubiquitous, convenient, on-demand network access to a shared pool of computing resources for rapid provision & release with minimal management effort / service provider interaction. Outsourcing 外包模式. To cut cost while focusing on core business
Datacenters	<ul style="list-style-type: none"> • On premises • On cloud • On hybrid mode: premises + cloud
Popular public cloud providers	<ul style="list-style-type: none"> • Amazon Web Services (AWS) • Microsoft Azure • Google Cloud Platform (GCP)

-deployment models of cloud computing

Cloud Deployment Models (4)

Public cloud: delivered by a cloud service provider, accessible through internet

Private cloud: operated solely for a single organization, managed internally

Community cloud: operated by one or more parties in communities with shared concerns,

Hybrid cloud: combine Public + (community / private) cloud; Public: Run generic services (email servers)
Community/ Private: Host specialized services (specific apps)

-explain building blocks of IT infrastructure

DNSCOE

-edge computing

Edge computing	Brings computing power and data storage closer to where it is needed such as at the edge of the network
Objective	<u>Min cloud / on-premises datacenter access</u>
Components	Routers, gateways, switches & sensors
Pros	<ul style="list-style-type: none"> • Low latency <ul style="list-style-type: none"> ◦ Data is processed close to the source (e.g., sensors, devices), reducing the delay to send data/ get response with distant cloud servers • Low bandwidth needs <ul style="list-style-type: none"> ◦ Only necessary/ summarised data sent to the cloud, minimize network usage • Real-time processing <ul style="list-style-type: none"> ◦ data is analyzed locally at the edge, responses and decisions can be made instantly <p># Most processing happens locally at the edge node (closer to the device), so only small or non-urgent data goes to the cloud.</p> <p>Cloud used only for heavy or historical tasks</p>

-calculate availability, find MTTR MTBF

- a) Mean Time Between Failures (MTBF): Uptime

b) Mean time to Repair (MTTR): Downtime

- **Availability**

- 1) **Single component:** One defect leads to downtime

$$\text{Availability} = (\text{MTBF} / (\text{MTBF} + \text{MTTR})) \times 100\%$$

- 2) **Serial components:** One defect leads to downtime

$$\text{Total Availability} = \text{Availability} \times \text{Availability}$$

- 3) **Parallel components:** One defect leads to no downtime But beware of SPOFs

$$\text{Total Availability} = 1 - (1 - \text{availability}) \times (1 - \text{availability})$$

- c) Given that Server X operated for 525600 minutes in the year 2024. It was reported that the server has a total downtime of 3600 minutes. Determine the Mean Time to Before Fail (MTBF) and Mean Time To Repair (MTTR) in hours respectively. Then, calculate the availability of Server X in percentage for the year 2024. (Show your answer in 5 decimal places.) (5 marks)

$$\text{Mean Time to Before Fail} = \frac{525600}{60}$$

$$(\text{MTBF}) = 8760 \text{ hours}$$

$$\text{Mean Time To Repair (MTTR)} = \frac{3600}{60}$$

$$= 60 \text{ hours}$$

$$\text{Availability} = \frac{8760}{8760+60} \times 100\%$$

$$= \frac{8760}{8820} \times 100\%$$

$$= 99.31973\%$$

\therefore Availability of Server X in percentage for the year 2024 is 99.31973%.

-calculate PUE, state which datacenter is more energy efficient

5

a) Given that the energy usage of the datacenters is recorded as follows:

- Datacenter A uses 170,000kWh of total energy per annum with 110,000kWh dedicated to its IT equipment.
- Datacenter B uses 150,000kWh of total energy per annum with 120,000kWh dedicated to its IT equipment.

Calculate the Power Usage Effectiveness (PUE) metric for both datacenters. Which datacenter is more energy efficient? Justify your answer. (Show your answer in 1 decimal place.)

(6, 4 marks)

~~Power Usage Effectiveness (PUE) of Datacenter A~~

$$= \frac{170\,000}{110\,000}$$

~~= 1.5454545~~

~~≈ 1.5~~

~~Power Usage Effectiveness (PUE) of Datacenter B~~

$$= \frac{150\,000}{120\,000}$$

~~= 1.25~~

~~≈ 1.3~~

~~∴ Datacenter B is more energy efficient compared to Datacenter A because there is only 0.3 kWh of energy consumed by Datacenter for every 1 kWh consumed by IT equipment. In Datacenter A, there is 0.5 kWh energy consumed by Datacenter for every 1 kWh consumed by IT equipment.~~

6

4

-concepts/methods to improve network availability

- Spine and leaf topology
- Network teaming
- Spanning Tree Protocol
- Multihoming

-factors that affect network performance

- Throughput and bandwidth
- Latency
- Quality of Service

-guidelines to ensure datacenter security

- Restrict physical access to selected, qualified staff
- Use entry registration systems and maintain access logs
- Secure doors with conventional or electronic locks (with authentication)

c) Discuss TWO (2) guidelines for ensuring the security of a datacenter. (5 marks)

- Implement CCTV

to record the environment in datacenter, easier to monitor (faster react when who destroy or steal equipment, or act as evidence).

5

- Access control by biometric sensor (fingerprint, face) or access card (NFC), ensure that only authorized access to data center

-Storage media

- Mechanical Hard Disk HDD: Serial ATA
- Solid State Drive SSD: NVMe (PCIe)
- Tape: Tape Library, Virtual Tape Library (VTL):

-raid

- RAID 0 (Striping)
- RAID 1 (Mirroring)
- RAID 10 (Striping + Mirroring)
- RAID 5 (Striping + Distributed Parity)
- RAID 6 (Striping + Double Parity)

-Advantage of VM

- Allows multiple servers on one host
- Can run different OS per VM (Windows, Linux, etc.)
- Easy to migrate, back up, and clone
- Strong isolation between VMs

-performance of OS

Factors affecting OS performance:

- i) Hardware performance: Faster CPU, RAM, and storage directly enhance OS responsiveness.
- ii) Application load: Heavy workloads slow down the OS; optimization ensures smoother multitasking.
- iii) OS configuration: Inefficient settings or unnecessary services waste system resources.

To improve performance:

- i) Increase memory: Reduces paging/swapping to disk, enabling faster data access and smoother multitasking.
- ii) Decrease kernel size: Frees up RAM, shortens boot time, reduces crash risk, and lowers attack surface.

Why it improves performance: Optimized memory and kernel management maximize hardware utilization, making the system faster and more stable.

-security policies

(OS)

Security

To improve OS security:

- **Patching: Fixes vulnerabilities, bugs, and design flaws** 缺陷 to avoid potential attack.
- **Hardening: Disables unnecessary** services, users, and protocols to reduce attack surface.
 - step by step process of configuring an operating system to protect it against security threats
 - used to instantiate new operating systems. Ensure security is optimal and is consistent in all deployment
- Virus Scanning: Detects and removes malicious software that can harm data or performance.
- Host-Based Firewalls: Filters incoming/outgoing traffic to block unauthorized access.
 - Most operating systems, including Windows, Linux, and UNIX, provide a built-in host-based firewall

- Limiting User Accounts: Minimizes risk from privileged misuse; enforces principle of least privilege.
- Hashed Passwords: Protects credentials by preventing recovery of original passwords.

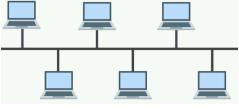
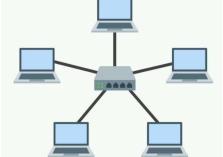
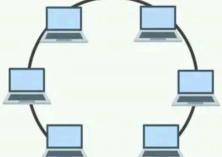
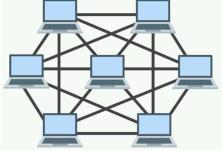
Why it improves security: Each measure reduces the likelihood of unauthorized access, malware infection, and privilege abuse, ensuring system integrity and reliability.

(end-devices)

- **Mobile Device Management (MDM)**
 - to monitor organization data on personal devices.
 - It can enforce password strength, device encryption, and remote wipe in case of device loss or theft.
- **Network Access Control and segmentation**
 - require devices to **meet security compliance standards (updated OS, antivirus, encryption)** before connecting to internal systems.
 - **Provide DMZ or VLAN for personal devices to connect,** instead of the core banking network.

C4: Networking

Network Topologies

	Bus	Star	Ring	Mesh
				
Connection by	Single cable / bus	Central hub / switch	A circular fashion	By connected to multiple devices
Data transmission	Along the bus, to intended recipient	Through the central point	In 1 direction around the ring	Routed through different paths
Pros	Simple & inexpensive to implement	Easy to manage and troubleshoot	Simple & efficient	Fault-tolerant
Cons	Prone to performance issues	Central hub as a SPOF	Disruptive to add / remove devices	Expensive to implement

1 Physical Layer

- physical components that carry the data

Cables

Twisted pair

- Unshielded Twisted Pair (UTP)
 - Used in **Ethernet** networks and for **telephone** systems
- Shielded Twisted Pair (STP) 
 - Used in factories, datacenters, or areas with **high electrical noise**

Fiber Optic

- Multimode
 - Used in LAN, datacenter, short distance
- Single-mode 
 - Used in telecommunication, long distance

Coaxial: For TV & Internet connection, outside Malaysia

Patch Panel

- Provides a **centralized location to organize and manage network cables**
- Ease changes, **Protect connections**, Provide clear reference for labelling & documentation

Cabling (2)

- Vertical Cabling x Horizontal Cabling

Vertical Cabling	Horizontal Cabling
<ul style="list-style-type: none"> connects different floors or buildings within a network infrastructure Connection: Patch panels to data center Long distance 	<ul style="list-style-type: none"> Connects devices within a floor Endpoints in the walls to patch panels Up to 90 meters
<ul style="list-style-type: none"> Enhance network reliability: fault isolation, increase uptime, disaster recovery Improve network performance: Load balancing, reduce latency Simplify network management: Centralized management, ease troubleshooting 	

Leased lines

T or E lease line

- Definition: Not standard industry terms for leased lines

Synchronous Optical Network (SONET)

- Definition: Transmit large amounts of data using fiber optic

Synchronous Digital Hierarchy (SDH)

- Definition: The international standard equivalent to SONET
- Used in: Telecommunications to transmit multiple digital data stream over a fiber optic

Dark Fiber

Dark Fiber	Regular Fiber (Lit Fiber)
Not actively used by the provider Installed but unused fiber optic cables	Actively used by the provider to deliver services
Fully controlled by the customer(lessee)	Controlled and managed by the provider
Leassee then installs their own equipment to "light up" the fiber and create their own private network	Provider supplies and manages equipment
Private, point-to-point network <ul style="list-style-type: none"> • High bandwidth, Low latency • Security, Scalability, Reliability 	Shared infrastructure managed by ISP

Internet Access

- DIA
- Broadband
- WiFi

2 Data Link Layer

- Ensure reliable & error-free data transfer between devices
- Network

	PAN	LAN	MAN	WAN
Distance	Within a few meters	Home, office building or school	A city or a large campus	Large geographical area
Used in	Connecting personal devices within a few meters	Setting up a network in a home office/ campus for fast local access	Linking multiple LANs across a city or large institution	Connecting networks across cities, countries or globally
Protocols	Bluetooth (L2CAP), Zigbee, Infrared	Ethernet (CSMA/CD), Wi-Fi (CSMA/CA)	Metro Ethernet, DQDB	PPP, HDLC, Frame Relay

- Common Protocols Used

	Ethernet	WiFi	PPP
Technology	Wired	Wireless	Direct connection between two devices
Used for	Connecting devices in a LAN	Connecting to the Internet wirelessly	VPNs, dial-up & Internet connection

3 Network Layer

- Define the route the data is sent to the recipient
- Implementations: IPv4, IPv6, Routing and addressing

IP Protocol

A unique numerical label assigned to each device connected to a network

Server can have multiple IP addresses, both public and private

- Load balancing

- **Hosting multiple websites**
- **Providing different services on the same server**

	Public IP Address	Private IP Address
Visible to	Entire network	Within a local server within a private network
Used for	<ul style="list-style-type: none"> • Web server • Email server • Public-facing applications 	<ul style="list-style-type: none"> • Internal communication

Addressing

	Static IP Address	Dynamic IP Address
Nature	Constant consistently reachable at the same address	Change periodically
Ideal for	Public servers	Internal servers

Routing

- Routers compile routing tables to make IP packet forwarding decisions
- Routing in the context of a server, covers:
 - Network Routing
 - Direct data packets are from the server to the other network devices
 - Application-Level Routing
 - Handle incoming requests within the server itself: load balancing, Content Delivery Networks (CDNs), application-specific routing

4 Transport Layer

- Maintain flow control
- Provide error checking
- Recovery of data between network devices

Protocols Used

	Transmission Control Protocol (TCP)	User Datagram Protocol (UDP)
Function	Provides reliable delivery of a stream of data between applications	Reduced latency over reliability by sending data without checking if the data arrived
Used in	FTP Web browsing Email	Live stream Online games VoIP
Used for	Large DNS queries	Small DNS queries

5 Session Layer

- Provides mechanisms for opening, closing and managing a session between end-user application processes

Virtual Private Network (VPN)

- Uses a **public network to interconnect private sites in a secure way**, a.k.a. VPN tunnel
- VPNs use **strong encryption and strong user authentication**. Using the Internet for transmitting sensitive data is considered safe
- often used for **remote access** to the LAN by users outside of the organization's premises
- Most common VPN communication protocol standards:

	Point-to-Point Tunneling Protocol (PPTP)	Layer 2 Tunneling Protocol (L2TP)	IPsec
Uses	For individual client to server	For individual client to server	For network-to-network

	connections	connections	connectivity. IPsec is built into IPv6 standard and is implemented as an add-on to IPv4
Security	Least	No encryption and relies on IPsec for encryption and authentication	Provides encryption and authentication for network traffic
Used in	Legacy system or when speed is prioritized over security	Paired with IPsec for a more secure VPN connection	VPNs and other secure network applications

6 Presentation Layer

- Takes the data provided by the application layer and converts it into a standard format that the other layers can understand

	Secure Socket Layer (SSL)	Transport Layer Security (TLS)
Nature	SSL is considered insecure and should not be used	TLS is securing WWW traffic carried by HTTP to form HTTPS

7 Application Layer

- Interacts with the OS or application

Protocol	Definition	Function
Domain Name System (DNS)	DNS is a distributed database that links IP addresses with domain names DNS was not designed with security in mind Updates to DNS records are done in non-encrypted clear text Authorization is based on IP addresses only	Translates human-readable domain names (like google.com) into machine-readable addresses
DNS Security Extensions (DNSSEC)	Provides origin authentication of DNS data for data integrity	Verifies the authenticity & integrity of DNS data
IP Address Management (IPAM)	IPAM systems are appliances that can be used to plan, track and manage IP addresses in a network IPAM systems integrate DNS, DHCP and IP address administration in one high available redundant set of appliances	Plans, tracks and manages the IP address space within a network
Network Time Protocol (NTP)	NTP ensures all infrastructure components use the same time in their real-time clocks Particularly important for: Log file analysis, clustering software, Kerberos authentication	Synchronization clocks of computers in a network to a common time source
Post Office Protocol (POP)	Used by email clients to retrieve email messages from a mail server	Retrieves emails from a mail server
Simple Mail Transfer Protocol (SMTP)	Used to send email messages from a mail client to a mail server or between mail servers	Sends emails from a mail client to a mail server or between mail server
Multipurpose Internet Mail Extensions (MIME)	Enables SMTP to support file attachments in email messages	Transmits non-text data (like images, audio, video) in email
File Transfer Protocol (FTP)	A protocol for transferring files between computers	Transfers files between computers over a network
Hypertext Transfer Protocol (HTTP)	Defines how messages are formatted and transmitted, and how web servers and browsers should respond to various commands	Transfers hypertext (HTML, CSS, JavaScript, images, etc)
Hypertext Transfer Protocol Secure (HTTPS)	Used when browsing the web with a web browser	Protected data during transfer

4.2 Networking Virtualization

Feature	traditional VLANs VLAN (Virtual Local Area Network) Group devices by function (eg. HR, finance) logically Isolate traffic between groups to reduce unauthorized access. Reduce broadcast domains and control network congestion. Layer: Operates at Layer 2 (Data Link layer).	VXLAN
Scale	Up to 4,094 VLAN IDs	Supports ~16 million VXLAN IDs
Layer	Operates at Layer 2	Uses Layer 3 encapsulation
Encapsulation	No encapsulation	Encapsulates Layer 2 frames in UDP
Use Case	Enterprise LANs	Large-scale cloud/data center networks
Port	—	Uses UDP port 4789
Multi-Tenant Support	Limited	Excellent for multi-tenant environments

Approaches	Used for / as	How
Virtual LAN (VLAN)	Logical grouping (for network segmentation)	Logically divides a single physical network into multiple broadcast domains
Virtual Extensible LAN (VXLAN)	Network virtualization technology	Uses encapsulation to create virtual networks that can span across physical networks, allowing for greater scalability and flexibility compared to VLANs
Virtual Routing and Forwarding (VRF)	Network routing technology (for segmentation)	Hosts multiple independent routing tables in a single physical router
Virtual Network Interface Controllers (VNIC)	Software-based representation of a network interface	Enables virtual machines to connect to the network and communicate with other virtual machines or physical devices
Virtual Switch (VS)	Software-based equivalents of physical network switches	Manages network traffic within a virtualized environment, providing functionalities like VLAN tagging, traffic shaping and connection to physical networks
Software Defined Networking (SDN)...	Software-based controllers	Manages network traffic and resources, enabling dynamic and programmable network configurations Abstract control plane(network decision-making) from data plane(traffic forwarding))
Network Function Virtualization (NFV)	Network architecture (virtualize network function)	Replaces traditional dedicated network hardware appliances with virtualized software instances running on commodity servers

4.3 Networking: Availability, Performance & Security

Networking Availability

Layered network topology	Spine-Leaf network topology
Traditional design that divides the network into three layers (Core → Distribution → Access), each with specific roles.	Modern data center design with two layers (Spine ↔ Leaf, Every leaf connects to all spines.) to improve scalability and performance.
Limited - adding devices increases complexity and potential bottlenecks at upper layers.	<ul style="list-style-type: none"> Highly scalable - just add more spine or leaf switches without major redesign.

<p>May face congestion at core/distribution layers under heavy server-to-server traffic. achieved through multiple links between layers but can be complex to configure.</p>	<ul style="list-style-type: none"> • multiple equal-cost paths - low latency and load balancing • Built-in redundancy - failure of one spine has minimal impact.
Network Teaming	<p>Provides a virtual network connections using multiple physical cables Bonds physical NICs to form a logical network team:</p> <ul style="list-style-type: none"> • Sends traffic to the team's destination to all NICs in the team • Allows a single NIC, cable or switch to be unavailable without interrupting traffic
Spanning Tree Protocol (STP)	<ul style="list-style-type: none"> • Guarantees only one path is active between two network endpoints • Redundant paths are automatically activated when the active path not available • Ensures no loops are created when redundant paths are available in the network • It is not using half of the network links, since it blocks redundant paths
Multihoming	<p>Connecting a network to two different Internet Service Providers (ISPs)</p> <ul style="list-style-type: none"> • Improve availability • If one ISP fails or experiences slowdowns, the network can automatically switch to the other, maintaining continuous internet access and eliminating single points of failure. • It is not always guaranteed that multiple network paths actually run on a different set of cables. Cables are used by multiple IS providers.

Networking Performance

Factors affect the speed of a connection:

1. Throughput & Bandwidth (Actual data transferred x Maximum data capacity).
 - Throughput cannot exceed bandwidth. When data exceeds capacity, packets are queued or dropped.
 - Higher bandwidth and optimized throughput improve data transfer speed and user experience.
2. Latency
 - One-Way Latency - Time from sender → receiver.
 - Round-Trip Latency - Time from sender → receiver → sender.
3. **QoS - Quality of service**
 - Assigns priorities to ensure critical data (e.g., VoIP) gets through even under congestion
 - Improves reliability for real-time apps by reducing delay and packet loss.
 - **Implementation (4)**
 - i. **Congestion Management**
 - what must be done if the data to be sent exceeds the bandwidth
 - prioritizes important packets (e.g., voice, video) and either queues or drops lower-priority ones.
 - ii. **Queue Management**
 - Packets are placed in different queues based on priority. When queues are full, low-priority packets are dropped first.
 - iii. **Link Efficiency**
 - Fragmenting large low-priority packets, letting high-priority packets sent between fragments of low qos packets.
 - iv. **Traffic Shaping**
 - Limits bandwidth for low qos and reserves bandwidth for high qos ones.
4. WAN Link Compression
 - Compresses data before sending across WAN.
 - Reduces required bandwidth, improves WAN performance.

Networking Security

Network Encryption	<p>Encryption protects data by converting it into unreadable code.</p> <p><u>Types:</u></p> <ul style="list-style-type: none"> ● Encrypting data in transit – Protects data while being transmitted across the network. <ul style="list-style-type: none"> ○ Example: HTTPS or VPN connections. ● Encrypting data at rest – Protects stored data on hard drives or SSDs. <ul style="list-style-type: none"> ○ Example: BitLocker, AES disk encryption. ● End-to-end encryption – Encrypts data between two endpoints; even intermediaries can't read it. <ul style="list-style-type: none"> ○ Example: WhatsApp or Signal messaging.
Firewalls	<p>Firewalls secure networks by separating LAN/WAN segments and blocking unauthorized traffic. Only allowed traffic passes through based on configured rules.</p> <p><u>Implementation:</u></p> <ul style="list-style-type: none"> ● Hardware appliance (e.g., Cisco ASA, Fortinet). ● Software on servers or VMs (e.g., pfSense). ● Host-based (built into OS, e.g., Windows Defender Firewall). <p><u>Traffic Control Methods:</u></p> <ul style="list-style-type: none"> ● Packet Filtering <ul style="list-style-type: none"> ○ Filters packets based on rules (IP, port, protocol). ○ Common on routers and OS-level firewalls. ● Proxy Firewall <ul style="list-style-type: none"> ○ Acts as an intermediary between clients and servers. ○ Example: Forward proxy for users, reverse proxy for web servers. ● Stateful Inspection <ul style="list-style-type: none"> ○ Tracks active sessions and packet states. ○ Allows only packets that belong to valid connections.
Intrusion Detection & Prevention (IDS / IPS)	<p>Detects or prevents unauthorized access or malicious activity.</p> <ul style="list-style-type: none"> ● Monitor traffic for suspicious patterns. ● Alerts administrators or dynamically blocks attacks. <p><u>Types:</u></p> <ul style="list-style-type: none"> ● Network-Based IDS (NIDS) <ul style="list-style-type: none"> ○ Monitors all network traffic at a strategic point. ○ Passive observation, not part of traffic flow. ● Host-Based IDS (HIDS) <ul style="list-style-type: none"> ○ Installed on individual devices or servers. ○ Monitors traffic, user actions, and system files.
De-Militarized Zone (DMZ)	<ul style="list-style-type: none"> ● A buffer network between the secure internal LAN and the public internet. ● Hosts public-facing services (e.g., web, mail, DNS servers). <p>Eg: A web server in the DMZ handles internet traffic without exposing internal systems.</p>
(RADIUS)	<ul style="list-style-type: none"> ● - Remote Authentication Dial-In User Service ● Centralized authentication protocol for managing network access. ● Verifies user credentials before granting access to network resources. <p>Eg: Used in enterprise Wi-Fi networks to authenticate users via Active Directory.</p>

C5: Storage

1 Disks – command sets

define how the computer communicates with storage devices (not the physical connector itself).

Advanced Technology Attachment (ATA)

- simple hardware and communication protocol (disks to PCs)
- (Common interface standard for connecting storage devices like HDDs and SSDs to computers.)

Small Computer System Interface (SCSI)

- set of standards for physically connecting and transferring data between computers (mostly servers) and peripheral devices, like disks and tapes
- (High-performance command set used mainly in servers and enterprise systems.)

1 Disks - type (2) + Tape (2)

Mechanical Hard Disk (HDD)	Solid State Drive (SSD)	Tape (2)...
Magnetic platters with spinning disks HDD = Balanced, cost-effective for daily use and large active storage.	Uses flash memory with no moving parts. SSD = Best for performance-critical and high-speed operations.	Magnetic tape reels Tape = Best for long-term, low-cost archiving and backups.
Performance (Speed): Moderate - limited by mechanical movement	Very fast - low latency and high throughput	slow - sequential 顺序 read/write only
Capacity: High (up to ~20 TB per drive)	Moderate (up to ~100 TB enterprise NVMe)	High (LTO-9: 18 TB per tape)
Cost per GB: Low	Higher than HDD	Lowest
Reliability: Mechanical parts may fail over time	More durable; resistant to shock	Long lifespan (~30 years), but fragile handling
SATA (Serial ATA): Standard interface for consumer PCs; cost-effective and widely used. SAS (Serial Attached SCSI): Enterprise-grade interface offering higher speed and reliability. NL-SAS (Nearline SAS): Combines large capacity of SATA with reliability of SAS; used for backup or archival storage.	SATA, SAS, NVMe (PCIe) Pros: Faster data access: SSDs access data in microseconds , compared to milliseconds for mechanical disks. No moving parts: Lower power consumption and no vibration , which improves overall system stability and lifespan of nearby components. (Modular Cons: Limited write lifespan: Flash memory wears out faster if the same areas are written repeatedly. Higher cost per GB: SSDs are more expensive than mechanical drives, especially for large storage capacities.	Cheapest option for long-term storage and archiving. Life expectancy up to 30 years (e.g., DLT, SDLT, LTO). Disadvantages: fragile, slow sequential access, prone to damage.

...Tape: Library Types

Tape Library:	Virtual Tape Library (VTL):
<ul style="list-style-type: none"> automated storage system that manages multiple tape drives and tape cartridges using robotic arms (drives, slots, barcode/RFID, and robotic loaders) How it Works: <ul style="list-style-type: none"> Backup software sends data to the tape library. The robotic arm picks an empty tape from a slot. It inserts the tape into a drive for data writing. Once done, it returns the tape to a storage slot. 	<ul style="list-style-type: none"> disk-based backup system that emulates a traditional tape library faster, supports parallel processing; typically uses SATA/NL-SAS arrays. How it Works: <ul style="list-style-type: none"> Backup software "sees" virtual tapes, similar to physical tapes. Data is written to high-speed disk storage (e.g., SATA or NL-SAS). Later, data can be replicated or moved to real tapes if needed.

3 Controller Implementation

- Connect disks/tapes to servers via PCI or network (NAS/SAN).
- Provide virtualization, high availability, deduplication, cloning, and thin provisioning

RAID (Redundant Array of Independent Disks)

- Improves performance and data availability using multiple disks.
- Implemented in hardware (controller) or software (OS).

Common RAID Levels:

- RAID 0 (Striping): Speed only; no redundancy.
- RAID 1 (Mirroring): Two copies; high reliability, 50% efficiency.
- RAID 10: Combines striping + mirroring; high performance & availability.
- RAID 5: Striping + single parity; balanced speed and protection.
- RAID 6: Dual parity; tolerates two disk failures.

RAID Level	Description	Benefits	Drawbacks	Use Case
RAID 0 (Striping)	Data split evenly across multiple disks (no redundancy)	Increases performance	No redundancy; if one disk fails, all data is lost	Temporary or non-critical data where speed matters
RAID 1 (Mirroring)	uses 2 disks that contain the same data	High availability; most reliable	Uses 50% of disk space for redundancy; expensive	Critical data needing high reliability
RAID 10 (Striping + Mirroring)	Combination of RAID 0 and RAID 1	High performance and availability	Uses 50% disk space for redundancy; costly	Databases and critical systems needing speed + redundancy
RAID 5 (Striping + Distributed Parity)	Data and parity blocks striped across disks	Good balance of performance, availability, efficiency	Single disk failure tolerated; slower write speed due to parity calculations	General purpose with moderate reliability needs
RAID 6 (Striping + Double Parity)	Like RAID 5 but with two parity blocks	Can tolerate two disk failures	More overhead than RAID 5; slower writes	Systems requiring extra fault tolerance during rebuilds

Compression

Reduces storage needs (2–2.5× typical), depending on data type.

Data Deduplication

Removes duplicate data segments to save space.

Inline deduplication: immediate but slower.

Post-process deduplication: done later, reduces performance impact.

Cloning & Snapshots

<u>Cloning</u>	<u>Snapshot</u>
<ul style="list-style-type: none"> full duplicate of the original data Independent. To keep the copy even if the original is deleted or damaged To test or use the data without affecting the source 	<ul style="list-style-type: none"> captures the state of the data at a specific point in time, but it is not a full copy. Dependent. needs the original data to exist (unless converted to full copy). use to rollback to previous stage

Thin Provisioning

- storage optimization technique that allows users to assign (or "pretend to give") more storage capacity to users or applications than is physically available.
 - virtually allocates large storage space to each user or application.
 - However, it only uses physical disk space when data is actually written.

4 Storage Architectures (4)

Direct Attached Storage (DAS)	Storage Area Network (SAN)	Network Attached Storage (NAS)	Object Storage
<p>Storage directly connected to a single computer (no network involved). Local storage for one machine; simple and fast access for standalone systems.</p>	<p>A dedicated high-speed storage network connecting servers to a shared pool of disks. Acts like a local disk to each server, but storage is centralized.</p>	<p>A file-level storage system accessible by multiple users over a standard network. Acts as a file server allowing multiple clients to read/write shared files.</p>	<p>Stores data as objects (file + metadata + unique ID) accessible via REST API (HTTP). Stores large, mostly unchanging data like media, backups, or archives.</p>
<p>Pros:</p> <ul style="list-style-type: none"> Low cost Simple setup High performance (no network latency) <p>Cons:</p> <ul style="list-style-type: none"> Cannot be shared between servers Limited scalability and flexibility <p>Use Case: Small businesses or standalone servers use DAS for local databases or application data (e.g., a POS system in a retail store).</p>	<p>Pros:</p> <ul style="list-style-type: none"> High performance and low latency Centralized management Scalable and reliable <p>Cons:</p> <ul style="list-style-type: none"> Complex setup and maintenance Expensive hardware and expertise required <p>Use Case: SAN for virtual machine storage, databases, and transactional applications (e.g., banking systems, ERP).</p>	<p>Pros:</p> <ul style="list-style-type: none"> Easy to set up and manage Centralized file sharing Good scalability with clustered NAS <p>Cons:</p> <ul style="list-style-type: none"> Slower than SAN (file-level vs. block-level) Less suitable for database or high I/O workloads <p>Use Case: File sharing, backups, and collaborative document storage in offices.</p>	<p>Pros:</p> <ul style="list-style-type: none"> Massively scalable Geo-redundant and durable Ideal for cloud and distributed access <p>Cons:</p> <ul style="list-style-type: none"> Not suitable for frequently changing data Access latency higher than SAN/NAS <p>Use Case: Stores large amounts of unstructured data (videos, images, backups, logs) with global access.</p>

5 Software Defined Storage (SDS) (...Software Defined Networking)

- Separates **control plane(management)** from **data plane(physical storage)**.
- Pools all physical storage into a virtualized shared resource.
- allows users to **manage all storage types (SAN, NAS, Object, DAS)** as a **single virtual storage pool through software**.
- All cloud providers' storage systems are SDS-based.

Functions:

- Provides data services: deduplication, compression, caching, snapshotting, cloning, replication, and tiering.
- Enables policy-based provisioning (e.g., automatic setup for databases with snapshots and tiering).
- Managed via API, CLI, or GUI, ensuring desired performance, availability, and security.

Scenarios

- Cloud Provider Infrastructure
 - AWS, Azure, and Google Cloud all run on SDS — users can create, resize, or delete storage via API or web console.
- Enterprise Data Center
 - A bank uses Dell EMC PowerFlex to virtualize SAN + SSD + HDD storage under one SDS system.

5.2 Storage: Availability, Performance & Security

1 Improve Storage Availability (4)

Refers to how often storage systems remain accessible for data operations (store, retrieve, manage).

1	<p><u>RAID (Redundant Array of Independent Disks)</u></p> <ul style="list-style-type: none"> • Combines multiple disks to provide redundancy and fault tolerance.
2	<p><u>Redundancy and Data Replication (2)</u></p> <p>Synchronous replication:</p> <ul style="list-style-type: none"> • Data is written to both primary and secondary storage before confirming write completion. • Ensures data consistency; risk of latency or downtime if connection fails. <p>Asynchronous replication:</p> <ul style="list-style-type: none"> • Data is first written to primary storage and later copied to secondary storage. • Reduces latency but may risk small data loss if failure occurs.
3	<p><u>Backup and Recovery</u></p> <ul style="list-style-type: none"> • Protects data from deletion, corruption, or disasters. <p>Follows the 3-2-1 rule:</p> <p>3 Copies: Keep three copies of your data.</p> <p>2 Different Media Types: Store the copies on two different media types.</p> <p>1 Offsite Copy: Keep one copy at a separate location.</p> <p>Importance: This rule ensures data is protected against various risks, including hardware failure, human error, and natural disasters. Keeping multiple copies on different media types and at different locations reduces the risk of data loss.</p> <p>Backup types:</p> <ol style="list-style-type: none"> i) Full Backup – Complete data copy. ii) Incremental Backup – Changes since last backup. iii) Differential Backup – Changes since last full backup. iv) Incremental Forever Backup – One full backup + continuous incrementals. v) Continuous Data Protection (CDP) – Captures every data change in real time (zero RPO).
4	<p><u>Archiving</u></p> <ul style="list-style-type: none"> • Long-term storage for compliance and regulations.

- Data is read-only, encrypted, and stored on durable media (e.g., WORM tapes, optical disks).
- Use open formats (e.g., XML) for future readability; migrate to new media every 10 years.

2 Storage Performance (3)

Refers to how often storage systems remain accessible for data operations (store, retrieve, manage).

To improve storage availability:

1	<u>Factors affecting disk performance:</u> <ul style="list-style-type: none"> i) Disk Rotation Speed (RPM): Faster rotation → lower delay (e.g., 5400 RPM = 5.6ms; 15000 RPM = 2ms). ii) Seek Time: Time for disk head to locate track (3-9ms typical). iii) Interface Protocol: Determines data transfer rate (e.g., SATA 6Gb/s, SAS 12Gb/s, NVMe 7GB/s).
2	<u>Metrics to measure performance:</u> IOPS (Input/Output Operations Per Second): Number of reads/writes per second. <ul style="list-style-type: none"> • HDD: <500 IOPS • SSD: ~400,000 (read), 150,000 (write) • NVMe: up to 1,000,000 (read)
3	<u>To improve storage performance:</u> <ul style="list-style-type: none"> i) Cache: <ul style="list-style-type: none"> • Read cache stores frequently accessed data. • Write-back cache speeds up write operations. ii) Storage Tiering: <ul style="list-style-type: none"> • Assigns data to storage based on importance and usage (e.g., SSD for Tier 1, SAS for Tier 2, tape for Tier 4). • Automated tiering moves data between tiers based on access patterns. iii) Load Optimization: <ul style="list-style-type: none"> • Balances workloads to prevent bottlenecks. • Example: Oracle uses a mix of RAID 1 & 5 for database optimization.

3 Improve Storage Security(2)

1	<u>Data at Rest (on disk or tape)</u> <p>Data Encryption: Prevents unauthorized access without a decryption key.</p> <p>Self-Encrypting Drives (SEDs): Built-in encryption requiring password at startup.</p> <p>Cryptographic Disk Erasure (CDE): Deletes encryption keys to render data unreadable.</p>
2	<u>Data in Transit</u> <p>SAN Zoning: Divides Fibre Channel SANs into logical groups; only devices in the same zone can communicate.</p> <p>SAN LUN Masking: Restricts Logical Unit Numbers (LUNs) to specific hosts; implemented at HBA level.</p> <p>Combining zoning and masking enhances access control.</p>

C6: Compute

6.1 Compute building blocks

1	<p><u>Compute Housing</u></p> <p>Tower (Pedestal): Standalone; placed on floor.</p> <p>Rack Servers: Standardized frames for multiple systems.</p> <p>Blade Servers: Shared enclosure components (power, fans, network, SAN); reduced wiring, cost, and failure points. Blade enclosure: houses 8-16 blades; shared redundant power, Ethernet, SAN, and management modules.</p> <ul style="list-style-type: none"> • Reduced wiring then fewer failure points: it only need to run one cable to the enclosure/chassis机箱, making it easier to manage and avoid safety concerns from tangled wires电线缠绕. • Lower initial deployment costs: it uses the enclosure's shared components like power supplies and fans, also shared redundant components (Power supplies, Backplane for interconnection, Network switches (redundant Ethernet connections), SAN switches (redundant Fibre Channel connections)).
2	<p><u>Processor</u></p> <p>① Central Processing Unit (CPU)</p> <ul style="list-style-type: none"> • <u>Executes program instructions</u> (arithmetic, logic, I/O). • Works using an instruction set (machine code). • Operates via clock cycles (GHz = billions of ticks/sec). • Word size: data handled per operation (modern = 64-bit). <p>② Graphics Processing Unit (GPU)</p> <ul style="list-style-type: none"> • <u>Thousands of cores for parallel processing.</u> • Accelerates AI/ML workloads and graphics rendering. <p>Example: NVIDIA Tesla GP100 (3840 cores, 150B transistors).</p> <p>#GPU usage increases? growing demand for the artificial intelligence, machine learning, big data analytics application GPUs are highly parallel processors that much faster than CPUs for workloads of these application</p>
3	<p><u>Memory</u></p> <p>Evolution</p> <p>Early: vacuum tubes → relays → magnetic core memory → RAM chips. Core memory was replaced by transistor-based RAM in the 1970s.</p> <p>① RAM (Random Access Memory)</p> <p>Volatile; data lost without power.</p> <p>SRAM(Static RAM): 6 transistors/bit, implement for cache</p> <ul style="list-style-type: none"> • fast, costly. <p>DRAM(Dynamic RAM): 1 transistor + 1 capacitor/bit, implement for main memory DRAM loses its data after a short time due to the leakage of the capacitors, refresh regularly to keep data available, cheaper.</p> <p>② BIOS (Basic Input/Output System)</p> <p>Firmware initializing hardware before OS loads. Stored in flash memory; regularly updated (BIOS flashing).</p>
4	<p><u>Interfaces</u></p> <p><u>External Interfaces:</u></p> <p>① USB: Replaced older interfaces (since 1996). USB 3.1 = 10 Gbps; USB-C = 40 Gbps (USB 4) & up to 20V power.</p>

2 Thunderbolt (Light Peak):
Up to 80 Gbps; 100W power; uses USB-C connector.

Internal Interfaces:

3 PCI: (cheaper and wide adoption) slower, uses shared parallel bus, limited bandwidth.

4 PCIe: (routed by a hub that allows multiple devices to communicate **simultaneously**) faster, uses point-to-point serial links, and scalable bandwidth

6.2 Compute Virtualization

LPAR (Logical Partition)	<ul style="list-style-type: none"> Hardware-based virtualization full and can different OS per LPAR divides single physical sys into multiple isolated env. Common on IBM mainframes and Power Systems. 	Enterprise systems, mainframes, midrange
VM (Virtual Machine)	<ul style="list-style-type: none"> Software-based virtualization full and can different OS per VM (<i>Hypervisor</i> create vm) creates and manages multiple virtual machines (VMs) on one physical host. 	Data centers, servers
Container	<ul style="list-style-type: none"> OS-level Shared host OS (Container Engine - Docker,) Pay for container runtime or host instance 	Cloud-native apps
Serverless Computing	<ul style="list-style-type: none"> Function-level by third-party cloud providers (AWS,) Pay per function execution 	Event-driven cloud tasks

1	<u>Compute Virtualization vs. Public Cloud</u>	
	Aspect	Compute Virtualization (Virtual Machine)
Definition	Uses hypervisor 虚拟机管理程序 to create multiple virtual machines (VMs) on a single physical server.	Provides virtualized compute resources over the Internet, managed by cloud providers.
Ownership	Managed by organization on its own servers. Full control over infrastructure.	Managed by third-party providers (AWS, Azure, GCP).
Scalability	Limited by on-premise hardware capacity.	Highly scalable on demand.
Cost Model	High initial cost, lower long-term cost.	Pay-as-you-go model, no upfront hardware.
Use Case (Takoyaki Example)	A local takoyaki shop uses virtualization to run POS and inventory VMs on one physical server.	A takoyaki franchise uses cloud VMs to host its online ordering system that scales automatically during lunch rush hours.
2	<u>Compute Virtualization Technology</u>	
	1 Emulation	
	<ul style="list-style-type: none"> Simulates different hardware architecture entirely through software. Allows running software for one system on another (e.g., PlayStation emulator on PC). 	

<p>Example (Takoyaki): A takoyaki shop runs an old POS app made for PowerPC on a modern x86 server using emulation.</p> <p>2 Logical Partitions (LPARs)</p> <ul style="list-style-type: none"> • <u>Hardware-based virtualization</u> that divides a single physical system into multiple isolated env. • Common on IBM mainframes and Power Systems. • Each LPAR can run its own OS. <p>Example (Takoyaki): A large takoyaki chain uses an IBM Power server — one LPAR for sales, another for payroll, another for supply management — all isolated but sharing hardware.</p> <p>3 Hypervisor</p> <ul style="list-style-type: none"> • <u>Software-based virtualization</u> layer that creates and manages multiple virtual machines (VMs) on one physical host. • Each VM has its own virtual CPU, memory, storage, and OS. <p>Types:</p> <ul style="list-style-type: none"> • Type 1 (Bare Metal): Runs directly on hardware (e.g., VMware ESXi, Microsoft Hyper-V). • Type 2 (Hosted): Runs on top of existing OS (e.g., Oracle VirtualBox). <p>Example (Takoyaki): The main takoyaki shop uses a hypervisor to host separate VMs — one for accounting, one for recipe management, and one for online orders.</p>
--

6.3 Types of Compute

<p>1 <u>Compute: Cloud – Containers Technology</u></p>
<p>i) Technology: Container</p> <ul style="list-style-type: none"> • Containers package applications with their dependencies to ensure consistency across environments. <p>ii) Implementation</p> <ul style="list-style-type: none"> • Chroot or Jail: Isolate processes and file systems for lightweight containment. • Namespaces: Provide process isolation by separating system resources (PID, network, users). • Cgroups: Control and limit resource usage (CPU, memory, I/O) for containers. <p>iii) Container Orchestration</p> <ul style="list-style-type: none"> • Tools like Kubernetes automate deployment, scaling, and management of containerized applications.
<p>2 <u>Compute: Cloud – Serverless Computing</u></p>
<ul style="list-style-type: none"> • Allows running code without managing servers. • Automatically scales by providers based on demand <ul style="list-style-type: none"> ◦ charges only for actual execution time. <p>#Serverless computing means running code without managing servers yourself. The cloud provider (like AWS Lambda, Azure Functions, or Google Cloud Functions) handles server provisioning, scaling, and maintenance automatically. Just upload your function/code → provider runs it when triggered → you pay only for runtime.</p>
<p>3 <u>Mainframe</u></p>
<p>#x86 = general-purpose; Mainframe = high reliability, enterprise-scale workloads.</p> <p>i) Components</p> <ul style="list-style-type: none"> • PU (Processing Unit): Executes instructions and handles computations. • Memory: Stores active data and instructions. • I/O Channels: Manage data transfer between mainframe and peripherals. • Control Units: Direct and coordinate input/output operations. <p>ii) Mainframe Virtualization: Logical Partitions (LPARs)</p> <ul style="list-style-type: none"> • Divides mainframe hardware into multiple isolated logical systems, each running its own OS.
<p>4 <u>Midrange Systems</u></p>

	i) Midrange Virtualization: Logical Partitions (LPARs) Similar to mainframes but designed for medium-scale workloads in enterprises.
5	X86 Servers Industry-standard servers based on x86 architecture, commonly used for virtualization, cloud, and enterprise workloads.
6	Supercomputers #Supercomputer = classical parallel processing; Quantum = quantum bit-based, experimental. Extremely powerful systems designed for intensive scientific and engineering computations requiring parallel processing.
7	Quantum Computers Use quantum bits (qubits) and quantum mechanics principles to perform complex calculations exponentially faster than classical computers.

6.4 Compute: Availability, Performance & Security

1	<p><u>Improve Compute Availability</u></p> <ul style="list-style-type: none"> • Hot swappable components <ul style="list-style-type: none"> ○ Components (e.g., memory, CPUs, interface cards, power supplies) can be installed, replaced, or upgraded while the server is running. ○ Why it improves availability: It prevents downtime because the system stays operational during • Parity memory 奇偶校验内存 <ul style="list-style-type: none"> ○ Uses parity bits as a simple error detection code to detect memory data corruption. ○ Why it improves availability: Early detection of memory faults prevents crashes and data corruption that could cause service interruptions. ○ it can only detect errors, not correct them. • Error-Correcting Code (ECC) memory <ul style="list-style-type: none"> ○ Detects and corrects single-bit memory errors automatically using Hamming Code or TMR. ○ Why it improves availability: Continuous error correction keeps servers stable and prevents system failure due to memory faults, which are more frequent in 24/7 environments. • Virtualization availability <ul style="list-style-type: none"> ○ Provides failover clustering where VMs automatically restart on another host if a hardware or OS failure occurs. ○ • Why it improves availability: Ensures minimal service downtime and automatic recovery during failures. ○ • Clustering requires spare or underloaded hosts to take over workloads, maintaining service continuity.
2	<p><u>Performance</u></p> <p>Factors affect compute performance</p> <ul style="list-style-type: none"> • Architecture of the server <ul style="list-style-type: none"> ○ Refers to CPU design, memory hierarchy, and bus interconnections. ○ • How it affects performance: Efficient architecture improves data throughput and reduces bottlenecks between components. • Speed of the memory and CPU <ul style="list-style-type: none"> ○ Determines how fast data is processed and accessed. ○ • How it affects performance: Faster CPUs execute instructions more quickly; faster memory reduces data wait time. • Bus speed <ul style="list-style-type: none"> ○ Refers to the data transfer rate between CPU, memory, and I/O devices.

- • How it affects performance: A faster bus minimizes latency when moving data between components.

To improve compute performance:

Moore's Law

- States that transistor count doubles every ~2 years.
- Why it affects performance: More transistors enable more processing units and higher efficiency, increasing CPU power — although physical limits are being reached.
 - Increasing clock speed (more instructions per second)
 - CPU executes instructions at each clock pulse.
 - • Why it improves performance: Higher clock rates mean more instructions per second, improving compute speed — limited by heat and power constraints.
 - Cache memory
 - Small, high-speed memory located on the CPU (L1/L2/L3).
 - • Why it improves performance: Stores frequently accessed data, reducing delays from slower main memory access.
 - Pipeline 多个指令阶段可以重叠执行
 - Allows overlapping execution of multiple instruction stages.
 - • Why it improves performance: Increases instruction throughput by utilizing CPU resources continuously.
 - Prefetching 加载即将到来的指令 and branch prediction 预测猜测下一个执行路径
 - Prefetching loads upcoming instructions; branch prediction guesses the next execution path.
 - • Why it improves performance: Reduces idle CPU cycles and cache misses, enabling faster instruction delivery.
 - Superscalar CPUs
 - Executes multiple instructions per clock cycle via parallel execution units.
 - • Why it improves performance: Maximizes CPU utilization and throughput.
 - Multi-cores CPUs
 - Integrates multiple cores into a single chip.
 - • Why it improves performance: Enables true parallel processing; improves multitasking and workload distribution while reducing heat and energy per operation.
 - Hyper-threading
 - Allows one core to execute multiple threads simultaneously.
 - • Why it improves performance: Keeps execution pipelines active, improving CPU efficiency for multi-threaded workloads.
 - Virtualization performance
 - Multiple VMs share one physical machine.
 - • Why it affects performance: Efficient consolidation reduces idle time but can cause I/O bottlenecks if overloaded.
 - • High CPU/memory capacity and fast storage improve performance.
 - • Overhead from hypervisor processing is typically <10%.
 - • Databases are I/O-intensive and may require dedicated physical servers or Raw Device Mapping to maintain speed.

3 Security

Physical security

i) Physical server

- USB port
- BIOS setting
- Server housing

Disable external USB ports and secure BIOS with passwords.

- Why it improves security: Prevents unauthorized access, data theft, or malware introduction via external devices.
- Detecting case openings can trigger alerts for tampering attempts.

ii) Data in use

	<ul style="list-style-type: none">• Trusted Execution Environment (TEE)• Virtualization security <p>i) To minimize attacks</p> <ul style="list-style-type: none">• Firewalls• IDS• Patching• Minimize complexity of hypervisor <p>ii) To improve virtualization security</p> <ul style="list-style-type: none">• De-militarized Zone (DMZ)<ul style="list-style-type: none">○ The DMZ isolates external-facing services (e.g., web servers) from the internal private network.○ How it affects virtualization: It prevents direct access from the internet to virtual machines in the internal network, reducing the attack surface and protecting sensitive data or management interfaces.• Systems management console<ul style="list-style-type: none">○ Provides centralized control and monitoring of all virtual machines and hosts.○ How it affects virtualization: Enables administrators to manage user permissions, track configurations, detect unauthorized changes, and apply patches or updates consistently, ensuring the virtualized environment stays secure and compliant.
--	---

C7: Operating systems

7.1 Popular Operating Systems

IBM z/OS
 IBM i (OS/400)
 UNIX
 Linux
 Berkeley Software Distribution (BSD)
 Windows
 MacOS
 OS for mobile
 Special purpose OSs

7.2 Operating systems building blocks

1	<p><u>OSs building blocks</u></p> <ul style="list-style-type: none"> An operating system allows multiple users, processes, and applications to share hardware and hides hardware complexity. <p>Kernel</p> <ul style="list-style-type: none"> The heart of the OS - starts/stops programs, manages files, and controls hardware access to prevent conflicts. <p>Drivers</p> <ul style="list-style-type: none"> Connect hardware devices (printers, NICs, keyboards, video) to the kernel. <p>Utilities</p> <ul style="list-style-type: none"> Built-in tools like user interfaces, editors, loggers, and update managers. <p>Applications</p> <ul style="list-style-type: none"> Software that communicates with the OS through system calls and APIs.
2	<p><u>OSs functions</u></p> <p>Process Scheduling</p> <ul style="list-style-type: none"> Simulates parallelism by rapidly switching between processes (preemptive multitasking). Ensures fair CPU allocation through complex scheduling algorithms. <p>File Systems</p> <ul style="list-style-type: none"> Organize data in directories and files, managing disk communication and permissions. Support multiple file system types: FAT, NTFS, UFS, VxFS, Ext (Linux). Journaling file systems track changes for fast recovery. File systems must be mounted before use and may use drive letters (Windows) or mount points (UNIX/Linux). File sharing via NFS (UNIX) or SMB/CIFS (Windows). <p>APIs and System Calls</p> <ul style="list-style-type: none"> System calls provide a hardware-independent interface for applications. APIs define how software can use these system calls (e.g., POSIX for UNIX/Linux, Windows API). <p>Device Drivers</p> <ul style="list-style-type: none"> Software that allows the OS to communicate with hardware components through defined APIs. <p>Memory Management</p> <ul style="list-style-type: none"> Allocates/deallocates memory for applications. Includes cache, paging, swapping, and DMA (Direct Memory Access) for efficient transfers. Prevents memory shortage by moving pages to/from disk as needed. <p>Shells, CLIs, and GUIs</p> <ul style="list-style-type: none"> User interfaces to interact with the OS. <ul style="list-style-type: none"> - CLI: text-based (bash, sh, cmd.exe). - GUI: graphical (Windows, X Windows). <p>OS Configuration</p> <ul style="list-style-type: none"> Stores settings in databases or text files (Windows Registry, Linux /etc, AIX ODM). Editable via configuration tools that simplify text-based settings.

7.3 Operating systems: Availability, Performance & Security

Availability

Failover Clustering - Improves system uptime by automatically transferring workloads from a failed node to a standby node.

- Failover cluster is a group of independent servers running identical operating systems, known as nodes, connected through a network and managed by cluster software.
- Every active application has a standby counterpart on a passive node, which remains idle until a failover occurs.
- When a failure happens, the standby application automatically becomes active and continues serving clients [ensures minimal downtime and uninterrupted service].
- The cluster manages each running application within a node as a package of application components, called a resource pool or application package [organizes applications for automatic monitoring and failover control].

i) Shared storage

- All nodes can access the same data; ensures data continuity during failover.

ii) Shared node

- Provides redundancy; another node takes over automatically during failure.

iii) Voting and quorum disks

- Prevents “split-brain” errors by deciding which node stays active during network disconnections.

iv) Cluster-aware applications

- Run on multiple nodes simultaneously; improve both scalability and failover recovery time.

Performance

- Factors affecting OS performance:

i) Hardware performance: Faster CPU, RAM, and storage directly enhance OS responsiveness.

ii) Application load: Heavy workloads slow down the OS; optimization ensures smoother multitasking.

iii) OS configuration: Inefficient settings or unnecessary services waste system resources.

- To improve performance:

i) Increase memory: Reduces paging/swapping to disk, enabling faster data access and smoother multitasking.

ii) Decrease kernel size: Frees up RAM, shortens boot time, reduces crash risk, and lowers attack surface.

- Why it improves performance:

Optimized memory and kernel management maximize hardware utilization, making the system faster and more stable.

Security

To improve OS security:

- **Patching:** Fixes vulnerabilities, bugs, and design flaws 缺陷 to close potential attack vectors.
- **Hardening:** Disables unnecessary services, users, and protocols to reduce attack surface.
 - step by step process of configuring an operating system to protect it against security threats
 - used to instantiate new operating systems. Ensure security is optimal and is consistent in all deployment
- Virus Scanning: Detects and removes malicious software that can harm data or performance.
- Host-Based Firewalls: Filters incoming/outgoing traffic to block unauthorized access.
 - Most operating systems, including Windows, Linux, and UNIX, provide a built-in host-based firewall
- Limiting User Accounts: Minimizes risk from privileged misuse; enforces principle of least privilege.
- Hashed Passwords: Protects credentials by preventing recovery of original passwords.

Why it improves security:

Each measure reduces the likelihood of unauthorized access, malware infection, and privilege abuse, ensuring system integrity and reliability.

C8: End User Devices

8.1 End user devices building blocks

End user devices are tools humans use to interact with applications.

Examples: desktops, laptops, virtual desktops, mobile devices, printers.

[They connect users to IT infrastructure and deliver input/output functions.]

Desktop & laptop

i) Desktop

- High performance, can store large data and run complex software.
- Issues: complex management, high maintenance cost, and local security risks.
- [Best for stationary, high-power tasks.]

ii) Laptop

- Portable and as powerful as desktops.
- Common risks: theft, damage, and malware from external networks.
- Usually connected to docking stations for power and peripherals.
- [Ideal for mobility but less secure.]

Mobile devices

- Examples: smartphones, tablets, smartwatches, cameras, cars.
- Connect via wireless networks (UMTS/LTE/Wi-Fi).
- Features: small form factor, low bandwidth, variable reliability.
- [Enable mobile access but require adaptive app design and secure connectivity.]

Bring Your Own Device (BYOD)

- Users bring personal devices to access organizational data, instead of using only company-provided devices.
- user paid for the device (they brought their own device), it will not be acceptable to:
 - Have systems managers erase the device (including all family photos or purchased music) in case of an incident
 - Have personal data visible to the systems managers
- Conflict: user freedom vs. IT security control.
- **Solution:**
 - **virtualization(personal + work vm)** + Mobile Device Management (MDM) → separates personal and work environments.
 - **Mobile Device Management (MDM)**
 - to monitor organization data on personal devices.
 - It can enforce password strength, device encryption, and remote wipe in case of device loss or theft.
 - **Network Access Control and segmentation**
 - require devices to **meet security compliance standards (updated OS, antivirus, encryption)** before connecting to internal systems.
 - **Provide DMZ or VLAN for personal devices to connect**, instead of the core banking network.

Printers

1. Laser Printers: use toner + drum + laser → high quality, fast.
2. Inkjet Printers: spray ink droplets → cheaper, energy-efficient, high-quality color prints.
3. Multi-Functional Printers (MFPs):
combine printer, scanner, copier, fax; include OS, storage, and network features → require patching and data protection.
4. Specialized Printers:
 - a. Dot Matrix: uses pins; reliable, low cost, noisy.
 - b. Line Printer: prints full lines; durable for industrial use.
 - c. Thermal Printer: uses heat-sensitive paper; quiet, compact, fast, but prints fade.

8.2 Desktop virtualization

Technology	Runs On	User Rss	Cost Model	Key Benefit
Application Virtualization	Local or Server	Shared OS	License-based	Compatibility & easy deployment
SBC	Server	Shared VM	Per-user/server license	Centralized management
VDI	Server(VM per user)	Dedicated VM	Pay-per-use (cloud)	Full OS isolation
Thin/Zero Client	Minimal hardware	Server resources	Low device cost	Easy to manage, low maintenance
PXE Boot	Network OS image	Network-based	Shared infra	Diskless, fast deployment

Application virtualization

- Runs apps in isolated virtual environments.
- OS resources are virtualized, not the app itself.
- Requests and file operations are redirected to virtualized locations.
- **[Allows running old/incompatible apps side-by-side and easier migration.]**
- Examples: Microsoft App-V, VMware ThinApp.

Server based computing (SBC)

- Applications/desktops **run on remote servers**.
- **Only display updates, keyboard, and mouse data are exchanged.**
- [Low bandwidth, centralized management.]
- Products: Windows RDS, Citrix XenApp.
- Advantages: Easier maintenance and consistent configuration.
- Disadvantages: **Depends on network latency, limited local performance.**

Virtual Desktop Infrastructure (VDI)

- Each user runs a full desktop OS in their own VM.
- Managed by a hypervisor that allocates resources.
- **[Provides isolation but uses more CPU/storage than SBC.]**
- Available as cloud services (Azure Virtual Desktop, Amazon WorkSpaces, Google Cloud VDI).
- Issue: **"Logon storm" when many users start VMs simultaneously.**

Thin clients (2)

- Lightweight devices connecting to SBC/VDI servers.
 - **No local storage or configuration;** easy to replace.
 - [Reduce cost, maintenance, and security risk.]
- i) **Zero client**
- No local OS; relies entirely on the server for boot and operation.
- ii) **Preboot eXecution Environment (PXE) boot**
- Boots OS over network instead of local disk; used for diskless workstations.
 - Requires constant network connection.
 - [Useful for managed office environments but unsuitable for mobile devices.]

8.3 End user devices: Availability, Performance & Security

Availability

To improve end user device's availability:

- Use **high-quality components** to reduce failure rates.
- Perform **regular maintenance** (software updates, cleaning, battery checks).
- Add **redundancy** (spare devices, backup data copies).

Performance

To improve end user devices' performance:

- **Add more RAM** – improves multitasking and system speed.

- **Use SSD instead of HDD** – faster data access and startup.
- **Ensure sufficient network bandwidth** for all users.
- **Optimize mobile apps** for low-bandwidth or offline operation.
- Use **Server-Based Computing (SBC)** to reduce the effect of slow connections.

Security

Challenges:

- Devices are **spread across offices, homes, and client sites**.
- Not locked in a datacenter, thus prone to **theft, malware, and misuse**.

To improve security:

i) General Protection

- Use **laptop cable locks** to prevent theft.
- Install **malware protection** (antivirus, firewall).
- Enable **full-disk encryption** to protect data.
- Erase disks completely before disposal or repair.

ii) Mobile Device Management (MDM)

- **Monitor, maintain, and secure** mobile devices remotely.
- **Remote wipe** feature to delete data on lost or stolen devices.
- **Locate** stolen devices via tracking software.

iii) End User Authorization & Awareness

- Restrict user privileges (no admin rights).
- Set **BIOS passwords** and disable USB/DVD booting.
- Conduct **security awareness training** (social engineering, strong passwords, handling sensitive data).

iv) Network Access Control (NAC)

- Control network access based on:
 - **Device identity** (known/trusted device)
 - **User/group roles**
 - **Compliance status** (latest antivirus, OS patches)
- Non-compliant devices are placed in an **isolated network** until updated.

C9: Infrastructure Management

9.1 Infrastructure Deployment Options

Infrastructure/ Cloud deployment models (4)

<p>On-Premises</p> <p>You own and manage all hardware and software in your own data center.</p> <p>Requires sufficient space, UPS, cooling, fire safety, redundant networks, and strong floor load capacity.</p> <ul style="list-style-type: none"> • Large enterprises with strict data control needs • Industries with regulatory requirements (e.g., finance, defense) 	<p>✓ Full control and high security ✗</p> <ol style="list-style-type: none"> 1. High Costs – Expensive setup and maintenance (hardware, licenses, IT staff, energy, cooling). Vendor lock-in risk. 2. Poor Scalability – Scaling requires physical installation and manual setup. 3. Limited Redundancy & Disaster Recovery – Single points of failure; complex and costly DR solutions. 4. Security Burden – Needs dedicated IT security, tools, and staff.
<p>Public Cloud</p> <p>You rent computing resources (e.g., servers, storage) from providers like AWS, Azure, or GCP. Management level depends on the model (IaaS, PaaS, SaaS).</p> <ul style="list-style-type: none"> • Startups or new businesses (greenfield projects) • Workloads that need rapid scaling or global reach 	<ul style="list-style-type: none"> • Cost-effective and scalable • Fast innovation and deployment • Pay-as-you-go model • Less control and data security • Vendor dependency • Requires reliable internet connection
<p>Private Cloud</p> <p>Infrastructure dedicated to one organization, either hosted internally or by a provider. Functions as a software-defined data center (SDDC) with automation and orchestration.</p> <ul style="list-style-type: none"> • Enterprises requiring strict security and compliance • Organizations needing isolated environments with virtualization 	<ul style="list-style-type: none"> • High control, privacy, and customization • Automated management (costing, reporting, scaling) • Comparable to IaaS model • High initial investment • Limited scalability • No true pay-per-use model
<p>Hybrid Cloud</p> <p>Combines on-premises or private cloud with public cloud for flexibility and phased migration. Sensitive data stays private, while non-critical workloads use public resources.</p> <ul style="list-style-type: none"> • Enterprises transitioning to cloud • Workloads needing both secure storage and dynamic scaling 	<ul style="list-style-type: none"> • Flexibility and cost optimization • Enables gradual migration • Best of both environments • Complex integration and management • Requires dual-environment expertise • Mixed cost model (cloud + on-premises)

9.2 Infrastructure Automation

<p>Configuration Management Tools</p> <ul style="list-style-type: none"> • Automate the configuration of servers, network devices, and infrastructure components, ensuring consistency. • Continuously check systems against blueprints; if deviations occur (e.g., missing software, changed settings), tools automatically correct them.
<p>Infrastructure as Code (IaC)</p> <ul style="list-style-type: none"> • Treats infrastructure like software code — configurations are defined in code files to automate

deployment and management.

- Terraform (by HashiCorp) is an open-source IaC tool using HashiCorp Configuration Language (HCL) or JSON to manage infrastructure across cloud and on-premises environments.

Version Control

- Tracks code changes over time using repositories that create new versions automatically when updates are pushed.
- Tools: Git (distributed CLI tool), GitHub (web-hosted Git), GitLab (self-hosted Git platform).

Orchestration Tools

Coordinate and automate complex workflows across multiple infrastructure components.

Act like a conductor ensuring all parts work in harmony (e.g., for deployments or scaling).

Key Functions:

- Workflow Automation – Automates provisioning, configuration, deployment, and scaling.
- Dependency Management – Handles retries, rollbacks, and alternative paths when tasks fail.
- Resource Coordination – Manages resources efficiently across cloud and on-premises platforms.

Cloud Management Platforms (CMPs)

- Provide centralized automation for provisioning, monitoring, and managing cloud resources.

9.3 Infrastructure Documentation

Infrastructure documentation tools / techniques

- Documentation ensures reliability, security, and effective management.
- Preserves knowledge when staff leave, aids troubleshooting, and supports disaster recovery by showing how infrastructure is configured and functions.

Configuration Management Database (CMDB)	<p>Inventory of hardware, software, and networking components, detailing make, model, location, and function.</p> <p>Supports ITIL processes; should be updated regularly (preferably automated).</p> <p>Enables correlation of components to identify changes or root causes of failures.</p>
Diagrams	<p>Topology Diagrams – Show relationships and connections between components. Tools: Microsoft Visio, Diagrams.net.</p> <p>ArchiMate – Open standard for modeling enterprise architecture; describes IT systems, business processes, and information flows.</p>
IaCs tools	<p>IaC can serve as living documentation, showing how infrastructure is built and why.</p> <p>Documentation updates can be made alongside code changes.</p> <p>Advantage: Always up-to-date with modifications.</p> <p>Disadvantage: Requires reading code; doesn't provide instant visual overview like diagrams.</p>
Documenting procedures	<ol style="list-style-type: none"> 1. Procedures for routine tasks 2. Infrastructure naming convention 3. IP addressing plan 4. DNS naming convention 5. Fallback procedure 6. Disaster recovery plan