Day – 4

Data centres

A data centre is a centralized facility equipped with computing resources such as servers, storage systems, networking equipment, and cooling infrastructure that is used for the delivery of cloud services over the Internet.

Key components of datacentre

- **Servers:** For data processing and storage.
- **Storage Systems:** To store large amounts of data.
- **Networking Equipment:** To ensure communication within the data centre and externally.
- **Power Supply and Backup:** To ensure consistent and reliable power.
- **Cooling Systems:** To manage temperature and prevent overheating.
- **Physical Security:** To protect against unauthorized access.
- **Rack and Enclosures:** For organizing and housing equipment.
- **Fire Suppression System:** To safeguard against fires.
- **Management and Monitoring Tools:** To track and manage performance.
- **Cloud Infrastructure (in cloud data centres):** For scalable, virtualized resources. In a cloud environment, **virtual machines** (VMs) are created to handle different workloads.

TYPES OF DATACENTERS

1. Enterprise Data Center:

- **Description:** These are private data centers owned and operated by individual organizations (e.g., a company or a government agency). They typically serve the needs of the organization's internal IT infrastructure and data processing requirements.

2. Colocation Data Center:

- **Description:** In a **colocation** data center, multiple organizations or companies share space in the same facility but with their own separate servers and hardware. The facility provides physical space, power, cooling, and network connectivity, while each organization manages its own equipment.

3. Cloud Data Center:

- **Description:** Cloud data centers are operated by cloud service providers (like AWS, Microsoft Azure, or Google Cloud) to support cloud computing services. They offer on-demand, scalable resources over the internet, allowing customers to rent virtualized computing resources.

1. Capacity and Scalability:

- **Consideration:** The data center should be designed with future growth in mind. This means having enough space for hardware, cooling, and power needs as the business or infrastructure grows.
- **Why it matters:** It allows the data center to scale seamlessly without requiring major redesigns in the future.
- **How to address it:** Ensure enough physical space, power capacity, and network bandwidth are available to handle future demands. Use modular designs to add components as needed.

2. Power Supply and Redundancy:

- **Consideration:** Power is a critical aspect of any data center, and a reliable power source with backup systems (like **Uninterruptible Power Supply (UPS)** and **backup generators**) is essential.
- **Why it matters:** Data centers must maintain operations even during power failures or fluctuations. Redundant power ensures high availability and minimal downtime.
- **How to address it:** Implement dual power feeds, UPS systems, and backup generators to provide continuous power in case of outages.

3. Cooling and Environmental Control:

- **Consideration:** Proper cooling is essential to prevent overheating of servers and equipment. Efficient cooling systems reduce energy consumption and maintain an optimal operating temperature.
- **Why it matters:** Excessive heat can damage hardware, reduce performance, and lead to system failures.
- **How to address it:** Design with efficient cooling technologies such as hot aisle/cold aisle containment, liquid cooling, or in-row cooling units. Use energy-efficient air conditioning or free cooling where possible.

4. Security:

- **Consideration:** Physical and network security are both critical in data center design. This includes preventing unauthorized access to the building, network, and data.
- **Why it matters:** Data centers house sensitive data, and any security breach can lead to theft, data loss, or reputational damage.
- **How to address it:** Implement physical security features like biometric access, surveillance cameras, security guards, and restricted access areas. For network security, use firewalls, intrusion detection systems, and encryption.

5. Network Design and Connectivity:

- **Consideration:** A data center must have fast, reliable network connectivity to ensure high-performance data transmission. The network infrastructure should be designed to handle large amounts of data without bottlenecks.

Types Of Storage

1. Primary Storage

- **Description:** This is the storage used by the computer or server to store data that is actively being used. It is typically fast and easily accessible by the system.
- **Types:**
    - **RAM (Random Access Memory):** Temporary, volatile storage used to store data that is currently being processed. Data is lost when the system is powered off.
    - **Cache Memory:** A small, high-speed memory located near the processor that stores frequently accessed data to speed up processing.
- **Use Case:** RAM is used for immediate data access during tasks, and cache memory helps improve system performance by reducing access time.

2. Secondary Storage

- **Description:** This refers to non-volatile storage that stores data persistently and is typically used for long-term data storage.

- **Types:**
  - **Hard Disk Drives (HDD):** Mechanical storage devices that use spinning disks to read/write data. HDDs offer high capacity but are slower compared to SSDs.
  - **Solid-State Drives (SSD):** Flash-based storage that is much faster and more durable than HDDs. SSDs have no moving parts and provide quicker access to data.
  - **Optical Storage:** Uses laser technology to read and write data, typically found in CDs, DVDs, and Blu-ray discs.
  - **Magnetic Tape Storage:** Uses magnetic tape to store data and is primarily used for archival storage due to its low cost, but it is slower than other options.
- **Use Case:** SSDs are used in devices that require fast access (e.g., laptops, smartphones), while HDDs are often used for large, cost-effective storage. Optical and tape storage are used for backups and archiving.

3. Tertiary Storage

- **Description:** This is used for infrequently accessed data, often as part of a long-term archiving strategy. It is typically slower but cheaper.
- **Types:**
  - **Cloud Storage:** Data is stored on remote servers and accessed over the internet. Examples include services like Google Drive, Dropbox, and Amazon S3.
  - **Network-Attached Storage (NAS):** A dedicated file storage system that allows multiple users and devices to retrieve data from a central location over a network.
  - **Object Storage:** A system that stores data in objects, often used for large-scale unstructured data such as photos, videos, and backups. Cloud providers like AWS S3 offer object storage solutions.
- **Use Case:** Cloud storage is widely used for backup, file sharing, and collaboration. NAS is commonly used in small-to-medium businesses for centralized file storage.

4. Distributed Storage

- **Description:** Distributed storage systems split data into smaller parts and store them across multiple physical devices or locations. This provides redundancy and fault tolerance.
- **Types:**

    - **RAID (Redundant Array of Independent Disks):** A method of combining multiple hard drives to improve performance and/or provide redundancy (e.g., RAID 1 for mirroring, RAID 5 for striping with parity).
    - **Cloud Storage (Distributed Version):** Many cloud providers use distributed storage models to ensure data is replicated across multiple data centers for reliability.
- **Use Case:** RAID is commonly used in enterprise servers and data centers for redundancy and performance. Distributed cloud storage is used to ensure data durability and high availability.

5. Hybrid Storage

- **Description:** Hybrid storage combines different types of storage (e.g., SSD and HDD) in one system to balance speed and cost-efficiency.
- **Types:**
    - **Hybrid Storage Arrays:** These arrays use a mix of SSDs and HDDs, with frequently accessed data stored on the SSDs for speed, and less frequently accessed data stored on the HDDs for cost-effectiveness.
- **Use Case:** Businesses may use hybrid storage to provide the benefits of both speed and capacity, optimizing cost and performance.

6. Cloud Storage

- **Description:** A type of tertiary storage where data is stored on remote servers, accessible via the internet. It is highly scalable and offers flexible storage options.
- **Types:**
    - **Public Cloud Storage:** Offered by providers like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud. It is a shared environment where users pay for the storage they use.
    - **Private Cloud Storage:** Dedicated cloud storage for an individual organization, offering more control and security than public cloud options.
    - **Hybrid Cloud Storage:** Combines private and public cloud storage for more flexible, scalable, and secure data management.
- **Use Case:** Cloud storage is widely used for backups, file sharing, and accessing data remotely across devices. Enterprises may use private or hybrid cloud storage for enhanced security and control.

7. Network-Attached Storage (NAS)

- **Description:** A dedicated file storage system that connects to a network, allowing multiple devices to access the stored data. It's like a centralized data repository for small-to-medium-sized businesses.
- **Use Case:** NAS is commonly used in small and medium-sized businesses (SMBs) for file sharing, backups, and collaboration among multiple users.

8. Storage Area Network (SAN)

- **Description:** A high-speed network of storage devices that connects multiple servers, enabling them to access shared storage as if it were a local hard drive.
- **Use Case:** SAN is used in enterprise environments where large amounts of data need to be shared between multiple servers, providing high-speed access and storage.

Backup and Recovery:

Backup refers to creating copies of data to protect it from loss due to hardware failure, data corruption, human error, or disasters. **Recovery** is the process of restoring that data from backups when needed.

Types of Backup:

1. **Full Backup:**
    - **Description:** A full backup copies all selected data, regardless of whether it has changed since the last backup.
    - **Advantages:**
        - Simple to restore because all data is in one place.
        - Provides the most complete backup.
    - **Disadvantages:**
        - Time-consuming and requires more storage space.
        - Can be slow if the dataset is large.
    - **Use Case:** Best for initial backups or when creating periodic snapshots of important data.
2. **Incremental Backup:**
    - **Description:** An incremental backup only copies data that has changed or been added since the last backup (whether full or incremental).
    - **Advantages:**

        - Faster and requires less storage than full backups.
        - Efficient for regular backups.
    - **Disadvantages:**
        - Restoration requires the last full backup and all subsequent incremental backups, making recovery slower.
    - **Use Case:** Ideal for frequent backups where only a small amount of data changes each time, such as daily backups.
3. **Differential Backup:**
    - **Description:** A differential backup copies data that has changed since the last full backup. Unlike incremental backups, it does not reset after each backup.
    - **Advantages:**
        - Faster recovery than incremental backups because only the last full backup and the most recent differential backup are needed.
        - More efficient than full backups for regular data updates.
    - **Disadvantages:**
        - Takes more time and storage than incremental backups as the size of the differential backup grows over time.
    - **Use Case:** Useful for balancing between recovery speed and storage requirements, typically used for backup systems that require daily or weekly protection.

3-2-1 Backup Strategy:

- **Description:** The 3-2-1 strategy recommends having:
    - **3 copies of data** (the original and two backups).
    - **2 different media types** (e.g., hard drives, cloud storage).
    - **1 offsite backup** (to protect against local disasters like fire or flooding).
- **Advantages:** Highly reliable for data protection and disaster recovery, ensuring redundancy.
- **Disadvantages:** Requires careful planning for multiple backup locations and types.
- **Use Case:** Highly recommended for businesses and critical data protection.

Key Network Security Concepts:

1. **Confidentiality:** Protects data from unauthorized access using encryption and access controls.

2. **Integrity:** Ensures data remains unchanged using hashing and digital signatures.
3. **Availability:** Ensures data is accessible when needed using redundancy and DDoS protection.
4. **Authentication:** Verifies identity using passwords, biometrics, and 2FA.
5. **Authorization:** Manages access to resources using RBAC, DAC, and MAC.
6. **Firewalls:** Monitors and controls incoming/outgoing network traffic.

   Types:

   Packet Filtering Firewall: Inspects packets of data and blocks or allows them based on security rules.

   Stateful Inspection Firewall: Keeps track of the state of active connections and makes decisions based on the state of traffic.

   Next-Generation Firewall (NGFW): Includes advanced features such as application awareness and deep packet inspection (DPI).

7. **IDS/IPS:** Detects and prevents intrusion by monitoring network traffic.
8. **VPNs:** Creates secure connections for remote users and networks.
9. **Encryption:** Secures data during transmission or storage.
10. **Security Policies:** Defines rules and guidelines to secure networks and data.
11. **Training:** Educates users to recognize and avoid security threats.
12. **ACLs:** Defines access rights to resources based on IP addresses and permissions.
13. **Zero Trust Security:** Assumes no trust and continuously authenticates users.
14. **Malware Protection:** Safeguards systems from malicious software.