Data Centre: A data center 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:

- **Servers**: Physical or virtual machines running applications and services.

- **Storage Systems**: Hardware or cloud-based solutions used to store data.
- **Network Equipment**: Routers, switches, and other devices used to connect systems within and outside the data centre.
- **Power Infrastructure**: Uninterrupted power supply, power distribution units, etc., to ensure constant power availability.
- **Cooling Systems**: Air conditioning, cooling towers, and related technologies to maintain an optimal temperature.
- **Security Systems**: Physical security, network security, firewalls, surveillance, etc., to safeguard the data centre.

Types of Data Centre:

On-Premises Data Centre:

- **Owned and Operated**: Owned and managed by the organization.
- **Complete Control**: Full control over hardware, software, and security.
- **High Initial Investment**: Significant upfront costs for infrastructure.
- **Suitable for**: Organizations with high security and compliance needs.

Colocation Data Centre:

- **Rented Space**: Companies rent space in a third-party data centre.
- **Lower Investment**: Lower capital costs than building an on-premises data centre.
- **Shared Infrastructure**: Resources are shared with other clients.

Cloud Data Centre:

- **IaaS, PaaS, SaaS**: Provides infrastructure, platform, and software services.
- **Virtualized Infrastructure**: Provides virtualized resources over the internet.
- **Pay as You Go**: Scalable, pay-per-use model.
- **High Scalability**: Easily scalable depending on demand.

Key Considerations of Data Centres:

- **Reliability**: Ensuring continuous uptime and availability.

- **Security**: Protecting data and physical assets from threats.
- **Scalability**: Ability to scale infrastructure as demand grows.
- **Energy Efficiency**: Reducing power consumption and carbon footprint.

- **Disaster Recovery**: Ability to recover from hardware failure or disasters.

Data Centre Infrastructure

1. **Power Infrastructure:**
   - o **Uninterruptible Power Supply (UPS)**: Ensures power continuity during outages.
   - o **Power Distribution Units (PDU)**: Distributes electricity to servers and systems.
   - o **Power Monitoring**: Tracks power usage and ensures efficiency.
   - o **Redundancy**: Backup systems (e.g., generators) to ensure uptime.
2. **Cooling Infrastructure:**
   - o **Air Conditioning Units**: Maintain temperature within optimal levels.
   - o **Cooling Towers**: Help remove excess heat from the data centre.
   - o **Hot and Cold Aisle Containment**: Separates hot and cold air to optimize cooling efficiency.
3. **Space Management:**
   - o **Modular Design**: Flexible infrastructure that can expand as needed.
   - o **Cable Management**: Organized routing and management of cables.
   - o **Raised Floors**: Provides space for cables and air circulation beneath the floor.

Types of Data Storage:

- **Primary Storage**: Direct access storage for active data (e.g., hard drives, SSDs).

Types of Storage:

- **DAS (Direct Attached Storage)**: Directly connected storage to a server or computer.
- **NAS (Network Attached Storage)**: Storage that is connected to a network and can be accessed by multiple devices.
- **SAN (Storage Area Network)**: High-speed network connecting servers to storage devices.

RAID (Redundant Array of Independent Disks):

1. **RAID 0 (Striping)**: Data is divided and stored across multiple disks for improved performance (no redundancy).

2. **RAID 1 (Mirroring)**: Data is duplicated across two or more disks for redundancy.

3. **RAID 5 (Block-level Striping with Parity)**: Data is striped across multiple disks, with parity distributed for redundancy (needs at least 3 disks).

4. **RAID 6 (Block-level Striping with Double Parity)**: Similar to RAID 5, but with two sets of parity for added fault tolerance (needs at least 4 disks).

5. **RAID 10 (Combination of RAID 1 and RAID 0)**: Combines the mirroring of RAID 1 and the striping of RAID 0 for both performance and redundancy (needs at least 4 disks).

Backup and Recovery:

Recovery Time Objective: target time within which a system or application must be restored after an outage.

Recovery Point Objective: The maximum amount of data loss that an organization can tolerate

Backup Types:

- **Full Backup**: A complete backup of all data.
- **Incremental Backup**: Backs up only the data changed since the last backup (full or incremental).
- **Differential Backup**: Backs up data changed since the last full backup.

Backup Strategies:

3-2-1 Backup Strategy:

Keep 3 copies of data.

Store on 2 different types of media (e.g., hard drive and cloud).

Keep 1 copy off-site for disaster recovery

Servers:

- **Types of Servers:**

- **Dedicated Servers**: Servers dedicated to a single organization.

- **Virtual Servers**: Multiple virtual machines run on a physical server, sharing resources.

Load Balancing:

Distributes traffic across multiple servers to ensure high availability and performance.

Load Balancing Methods:

- **Round Robin**: Distributes traffic evenly to all servers.
- **Least Connection**: Directs traffic to the server with the fewest active connections.
- **Least Response Time**: Routes traffic to the server with the fastest response time.
- **Source IP Hashing**: Routes traffic based on the source IP address.
- **Weighted Round Robin**: Routes more traffic to servers with higher capacities.

Types of Load Balancers:

- **Hardware Load Balancers**: Physical devices used to distribute traffic.

- **Software Load Balancers**: Load balancing done via software on general-purpose hardware.