

# CompTIA Security+

## Lecture 07

SY0-701

Data Protection

By  
**Dr. Khaled ELMenshawy**

## Lecture

## 07

## LEARNING OUTCOMES

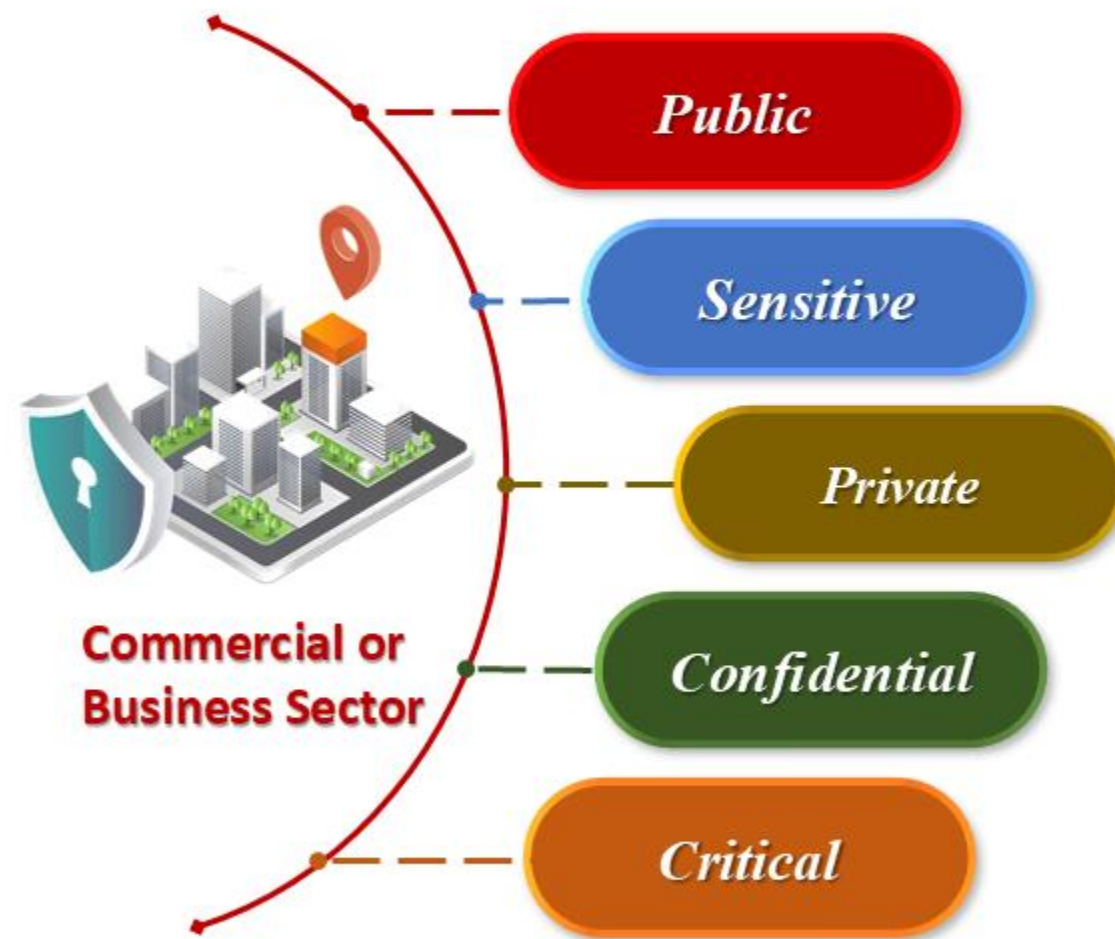
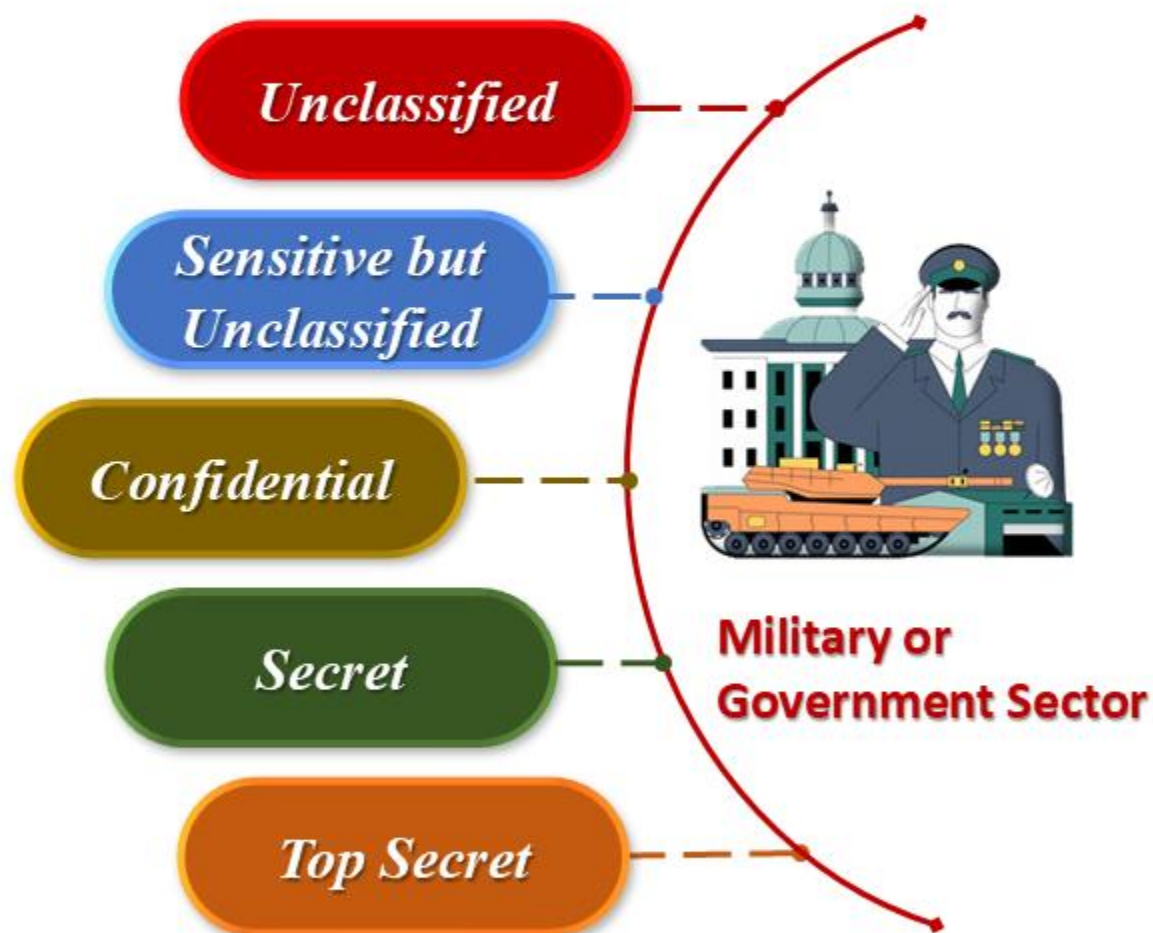
- Understanding the Data Classification.
- Understanding the Data Ownership
- Understanding data states, types, and sovereignty.
- Recognize how to secure the data.
- Understanding Data Loss Prevention.
- Understanding cyber resilience & cyber redundancy.
- Understanding high availability.
- Understanding Data Redundancy.
- Recognize different types of RAID.
- Understanding data backup.

### What is Data Classification?

**Data Classification;** is the process of *organizing data based on the value and the sensitivity to the organization*. The person that decides that level of data classification is the data owner.

**Sensitive Data;** is any data that can result in the *loss of security or loss of advantage to a company*, especially if it's accessed by unauthorized persons.

*Data classification is important for providing suitable resources to secure each class according to its sensitivity to the organization.*



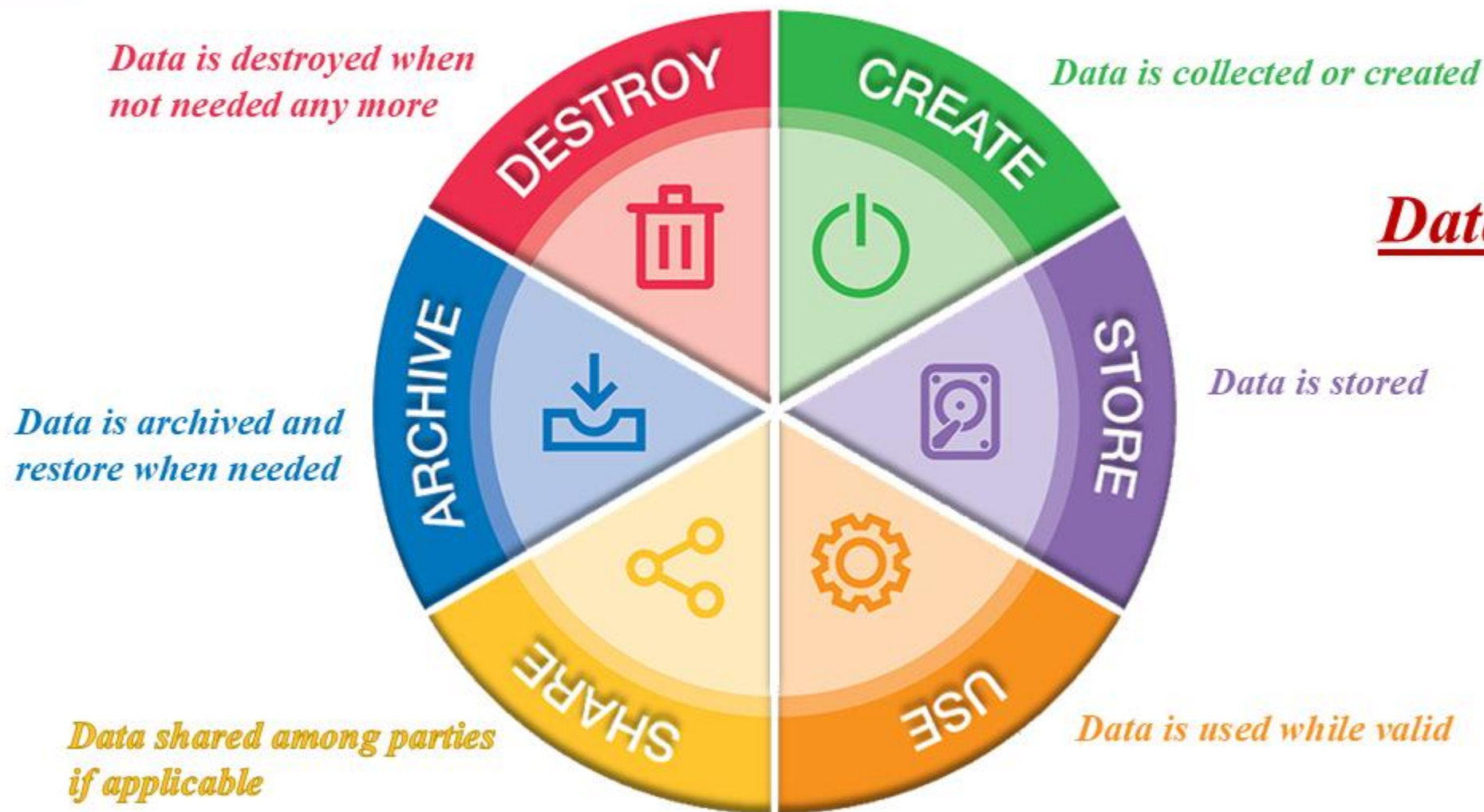
**Commercial or Business Sector:**

- **Public**; have *no impact on the organization* if it's released, and usually this is information that is posted on an open-source environment.
- **Sensitive**; Data that can result in *the loss of security or loss of advantage to an organization*, especially if it's accessed by unauthorized persons.
- **Private**; Data *not made available to the public and related to individual entities*, such as personal records, salary information, national ID numbers, etc.

- **Confidential**; Data *not made available to the public and affect the business if disclosed.*  
such as trade secrets, production lines, etc.
- **Critical**; Data *contain valuable information that can harm the organization badly if disclosed.* such as financial records.

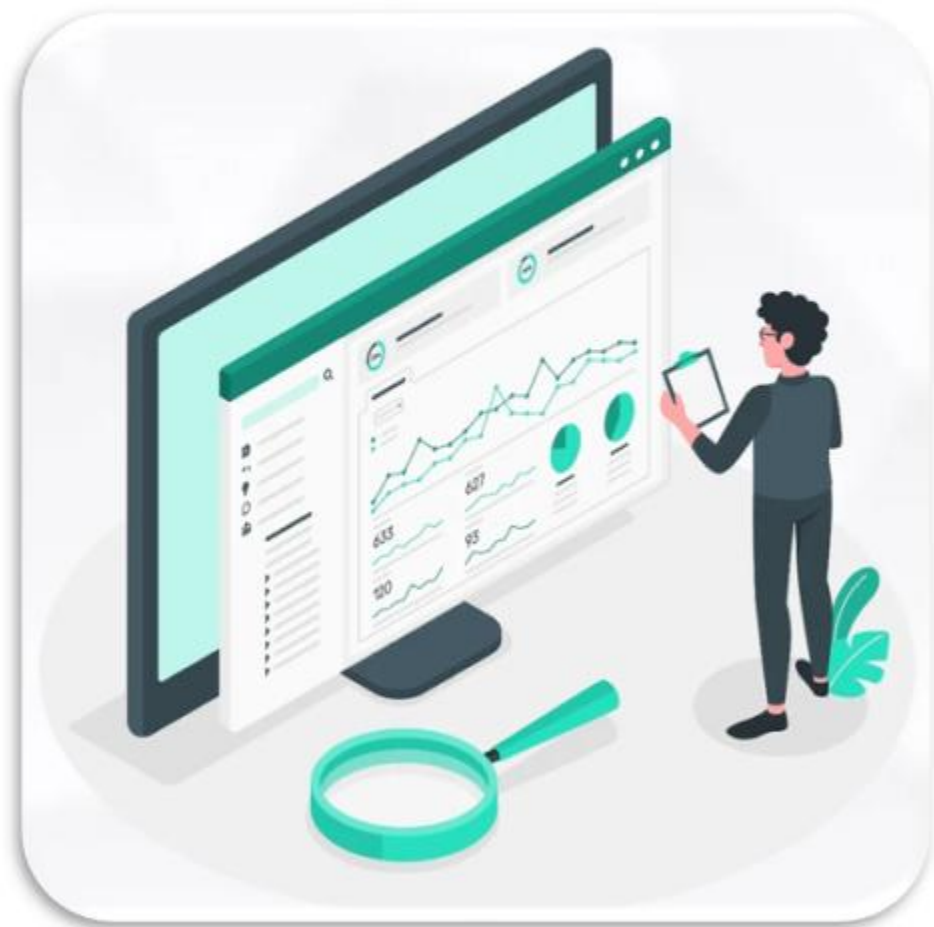
**Military or Government Sector:**

- **Unclassified**; Data that *can be released to the public*.
- **Sensitive but Unclassified**; Data that *would not harm national security* if disclosed but could *have an impact on whose data belong* such as medical records.
- **Confidential**; Data that *can seriously affect the government* if unauthorized disclosure were to happen.
- **Secret**; Data that could seriously *harm national security* if disclosed.
- **Top Secret**; Data that could seriously *damage national security* if disclosed.

**Data lifecycle**

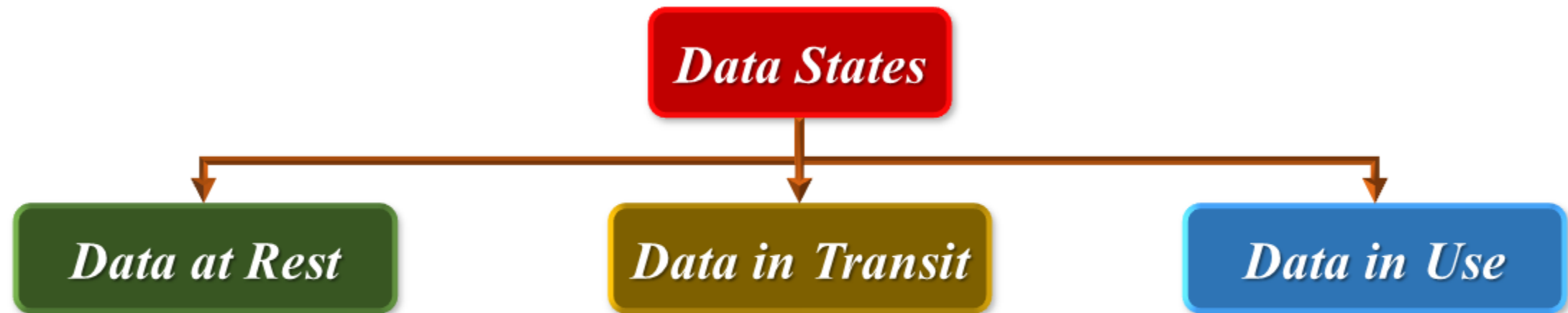
### What is Data Ownership?

**Data Ownership;** is the process of *identifying the person responsible for the confidentiality, integrity, availability, and privacy of the data assets.*



- **Data Owner;** is a *senior executive role* who have the *ultimate responsibility for maintaining the confidentiality, integrity, and availability of the data asset.*
- **Data Stewards;** *focused on the quality of the data and the associated metadata* They're going to be involved with making sure that the data is appropriately labeled and classified.
- **Data Controllers;** The entity that holds the responsibility for *deciding the purposes and methods of data storage, collection, and usage, and the legality of these processes.* The data controller holds the ultimate accountability for any breaches of privacy and *can not delegate this responsibility to another party.*

- **Privacy Officers;** The role that *holds the responsibility responsible for any kind of data privacy*. They have to make sure that the data privacy complying with legal and regulatory frameworks of the privacy statement of the organization.
- **Data Administrators (custodian);** the role that is responsible for handling the *management of the system which data assets are stored* and responsible for enforcing the access control, the encryption, and the backup recovery measures that protect this data.



**Data at Rest:**

- It refers to any *data stored in databases, file systems, or other storage systems*. This data is not actively moving through the network or processes.
- Data at rest is a prime target for threat actors, thus deploying encryption for stored data is necessary, There are several types of encryption methods used for data at rest:
  - ***Disc Encryption***; which *encrypts the entire hard drive*, when the user is logged in, the data is then decrypted.
  - ***Partition Encryption***; which *encrypts specific partitions of a hard drive* leaving other partitions unencrypted.

- ***File Encryption***; which *encrypts individual files*. It's often used when specific files need to be secured.
- ***Volume Encryption***; which *encrypts a set of selected files or directories* leaving other files or directories unencrypted.
- ***Database Encryption***; which *encrypts data stored in the database*. It can be done at the column, row or table level.
- ***Record Encryption***; which *encrypts specific fields within a database record*. This could be beneficial when multiple users are accessing the same database.

### *Data in Transit:*

- It refers to *data being moved from one location to another*, such as across the internet or through a private network.
- This data is vulnerable to interception during its journey. There are several types transport or communication encryption methods are used:
  - *TLS/SSL or VPN*; which *provide secure communication over a computer network.*, when the user is logged in, the data is then decrypted.
  - *IPSec*; which *is used to secure the internet protocol communications by authenticating and encrypting each IP packet in a data stream.*

**Data in Use:**

- It refers to *the data in the process of being created, retrieved, updated, or deleted*. In other words, it's data that's being processed.
- Techniques used to protect data in use such as *encryption at the application level*, *access controls* and *secure enclaves*, where data can be processed in a protected isolated environment.
- Also, some mechanisms like the *Intel software guard* are able to *encrypt data as it exists in memory* so that an untrusted process cannot decode the information.

- Data is like the lifeblood of organizations. It drives decision making and provides insights into business operations. However, not all data is created equal. *Different types of data require varying levels of protection and are subject to different regulations.*
- Data type may classify using different methods such as structured and unstructured data, qualitative and quantitative data, or even by nature such as graphs, numbers, etc.
- But when it comes to cybersecurity, data types classify according to *trade secrets, intellectual property, legal information, financial information, and human-readable versus non-human readable data.*

### Data types in cybersecurity:

➤ Regulated Data; is *data or information that is controlled by laws, regulations, or industry standards*. This includes data such as *personal identification information (PII)*, *Protected health Information (PHI)*, and credit card information.

Organizations that handle regulated data *must comply with specific security and privacy requirements* such as the *General Data Protection Regulation (GDPR)* in the European Union, or the *Health Insurance Portability and Accountability Act (HIPAA)* in the United States.

➤ **Trade Secrets**; are *a type of confidential business information that provides a company with a competitive edge*. This can include manufacturing processes, marketing strategies, proprietary software, and customer lists.

Trade secrets are *protected by law*, and unauthorized disclosure can result in severe legal penalties.

➤ **Intellectual Property**; is *the creations of the mind*, such as inventions, designs, etc.

This data is *protected by Intellectual property protection laws* like patents, copyrights, and trademarks to encourage innovation and creativity. unauthorized use of intellectual property can lead to legal action.

- **Financial Information**; includes *data related to organization's financial transactions*, such as sales records, invoices, tax documents, and bank statements.

Financial information is usually subject to regulations such as the *Payment Card Industry (PCI)*, and *Data Security Standard (DSS)*.

- **Human and non-human readable data**; Human-readable data is the information that can be understood directly by humans without the need for a machine or software, otherwise if information requires a machine or a software to interpret it called non-human readable data.

### What is Data Sovereignty?

**Data Sovereignty;** is the concept that digital information is *subject to the laws of the country in which it is located and within which it is collected or processed.*

### Geographical Considerations:

The geographical location of data storage and processing can have significant implications for businesses. *Different countries have different laws and regulations concerning data protection, privacy, and cross-border data transfers.*



*The strategies and techniques used to safeguard valuable data:*

- *Geographic Restrictions (geofencing)*; involve *setting up virtual boundaries to restrict data access based on the geographic location*. This method can help organizations comply with data sovereignty laws and prevent unauthorized access from high-risk locations, *for example*; preventing login requests from specific locations.
- *Encryption*; is a fundamental data security method that transforms readable data, plaintext, into unreadable data or ciphertext. There is different methods of encryption stated in Lecture (2); Cryptographic solutions.

- **Hashing**; is a technique that *converts data into a fixed size of numerical or alphanumeric characters*, known as the hash value. Unlike encryption, hashing is a one-way function. Once data is hashed, it cannot be reversed or decrypted.
- **Masking**; is replacing some or all of the data in a field with a placeholder, such as \* or X, to conceal the original content.
- **Tokenization**; is *replacing sensitive data with non-sensitive substitutes, known as tokens*. The original data is stored securely in a separate database with a token serving as a reference to the original data.

- **Obfuscation**; is *making data unclear or unintelligible*, making it difficult for unauthorized users to understand. This can involve various techniques, such as steganography, tokenization, data masking.
- **Permission Restrictions**; is defining *who has access to specific data and what they can do with it*. This is often managed through access control list or role-based access control.

### What is Data Loss Prevention (DLP)?

**Data Loss Prevention;** is set up to *monitor the data of a system while it's in use, in transit, or at rest* in order to detect any attempt to steal the data.



### What is Data Loss Prevention System?

***Data Loss Prevention System;*** is monitor and protect sensitive information across on-premises systems, cloud-based locations, and endpoint devices. These systems come as either software or hardware solutions.



***Network DLP***



***Cloud DLP***



***Storage DLP***



***Endpoint DLP***

- **Endpoint DLP System;** is usually a *piece of software that is installed on a workstation or a laptop*. It's going to *monitor the data that's in use on the computer*. If someone tries to do a file transfer, it'll either stop the file transfer or alert the admin of the occurrence based on certain rules and policies.
- **Network DLP system;** is a *piece of software hardware that is a solution placed at the perimeter of your network*. A sole function in life is to check all of the data going into and out of your network with a special focus on things going out of the network.

- **Storage DLP System;** is a *software that is installed on a server in the data centers* and inspects the data while it's addressed on the server.
- **Cloud-based DLP System;** is a *software as a service (SaaS) solution*, and that's a part of the cloud service. They're going to protect the data when it's stored inside of those cloud services.

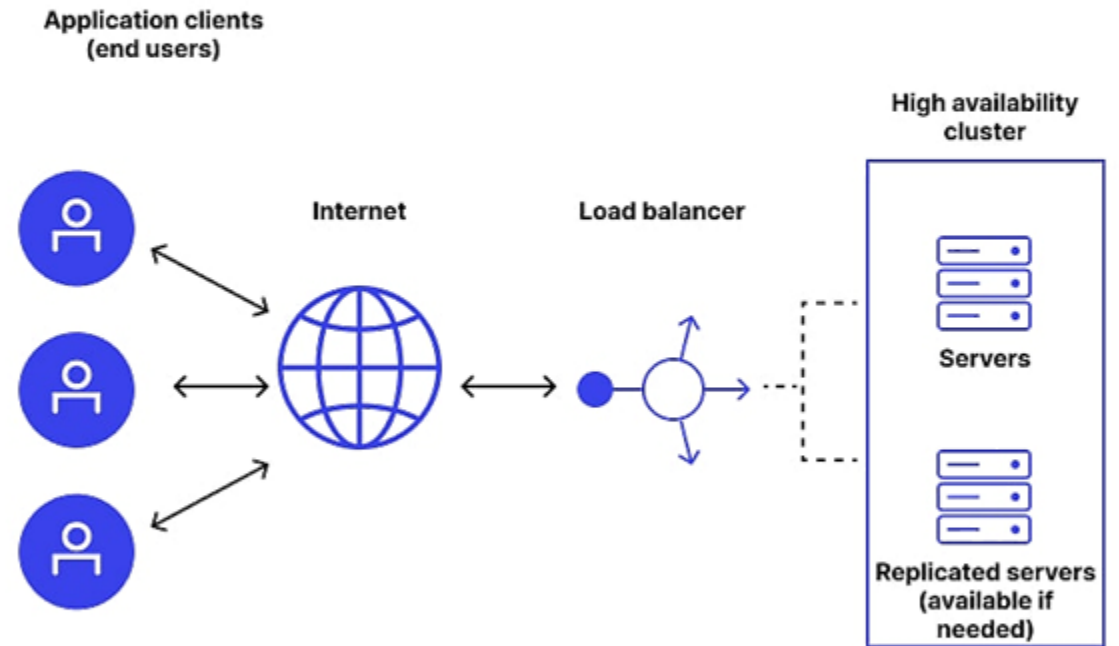
## Cyber Resilience & Cyber Redundancy

---

- **Cyber Resilience;** is the entity's *ability to continuously deliver the intended outcome, despite adverse cyber events.*
- **Cyber Redundancy;** is *having additional systems, equipment, or processes, to ensure the continued functionality,* if the primary ones are going to fail.

### What is high availability?

*High availability;* is the *ability of a service to be continuously available* by minimizing the downtime to the lowest amount possible.



➤ **Uptime (%)**: is used to *measure the levels of availability* and it is the number of minutes or hours that the system remains online over a given period.

➤ **Five nines of availability**: is the gold standard in availability and means *having 99.999 % uptime* which means that the services can only have a maximum of about *five minutes of downtime per year*.



Some cloud-based networks, organizations have gone even further by setting their availability target at *six nines of availability, or 99.9999%*, which equates to only *31 seconds of downtime per year*.

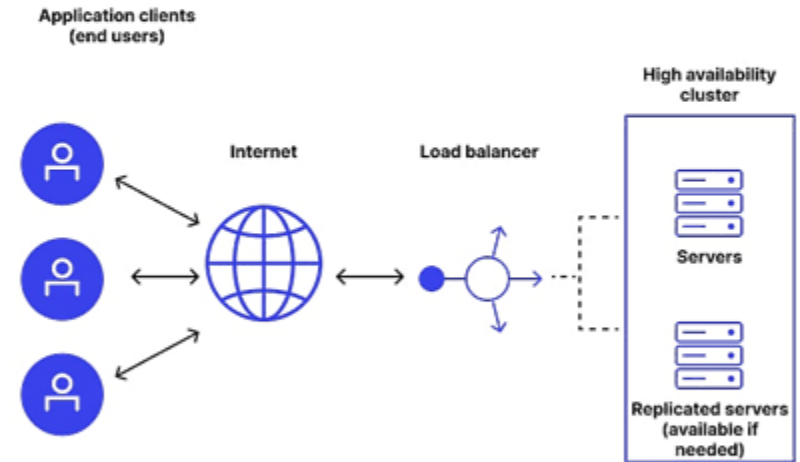
*To achieve higher levels of availability:*

➤ *Apply load balancing;* is the process of *distributing workloads across multiple computing resources* in order to:

- Optimize the resource use.
- Maximize their throughput.
- Minimize their response time.
- Prevent the overloading of any single resource.

➤ **Clustering**; is the *use of multiple computers, multiple storage devices, or redundant network connections that all work together as a single system* in order to:

- Provide higher levels of availability.
- Provide reliability.
- Provide scalability.



Clustering is keeping an application available even in the event of a hardware failure.

Load balancing and clustering can also be combined together in the same architecture.

➤ **Redundancy;** is the *duplication of critical components or functions of a system* with the intention of *increasing the reliability* of a given system. To create a highly available infrastructure, redundancy has to be built into your designs by installing:

- Multiple power supplies (Generators, UPS, Power Grid).
- Multiple network connections (wired and wireless, multiple cable connections).
- Multiple servers and storage devices (clusters, load balance).
- Multiple software services.
- Multiple service providers (multiple ISP).

### What is Data Redundancy?

*Data Redundancy;* is keeping data in two or more places within a database or data storage system.

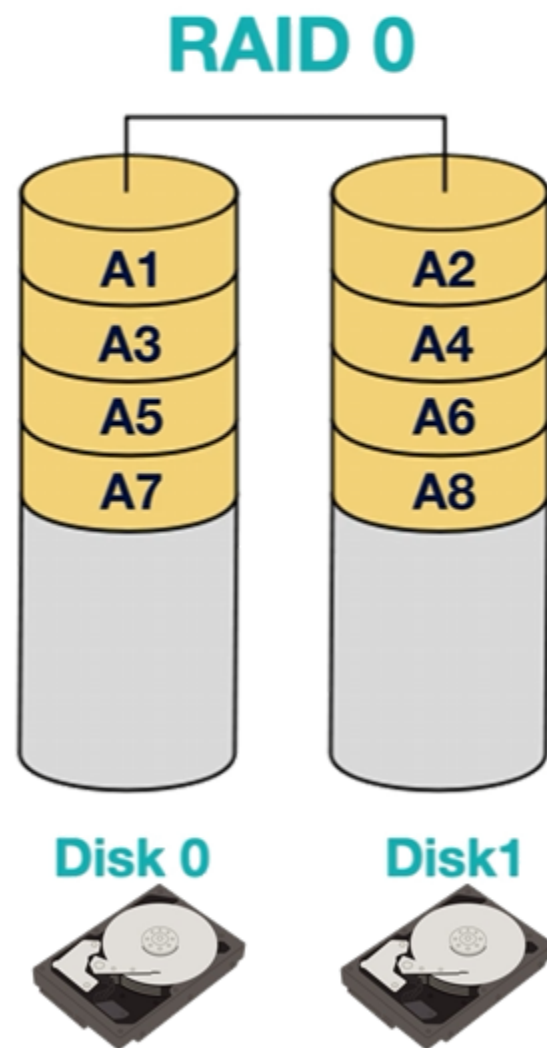
To create data redundancy is to use a *RAID*. A *RAID* is *a redundant array of independent disks*, which allows an organization to *combine multiple physical storage devices into a single logical storage device*.

**R**edundant  
**A**rray of  
**I**ndependent  
**D**isks

**RAID 0 (Striping):**

**RAID 0;** is used for performance only and does not care about fault tolerance, thus there is no data redundancy in **RAID 0**.

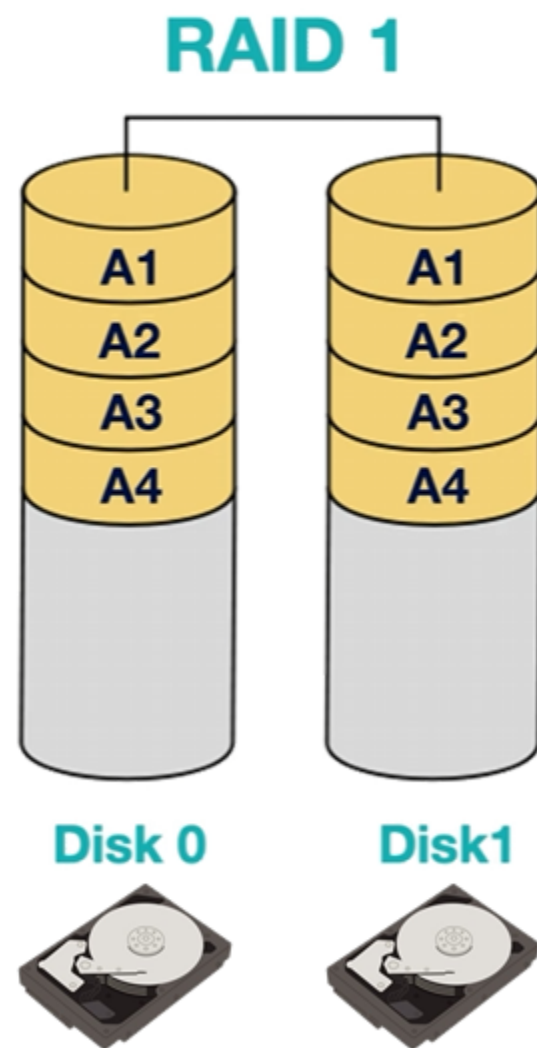
The data is striped into pieces equal to the number of existing storage devices.



**RAID 1 (Mirroring):**

**RAID 1;** is used for fault tolerance and data redundancy by mirroring the data identically onto two identical storage devices.

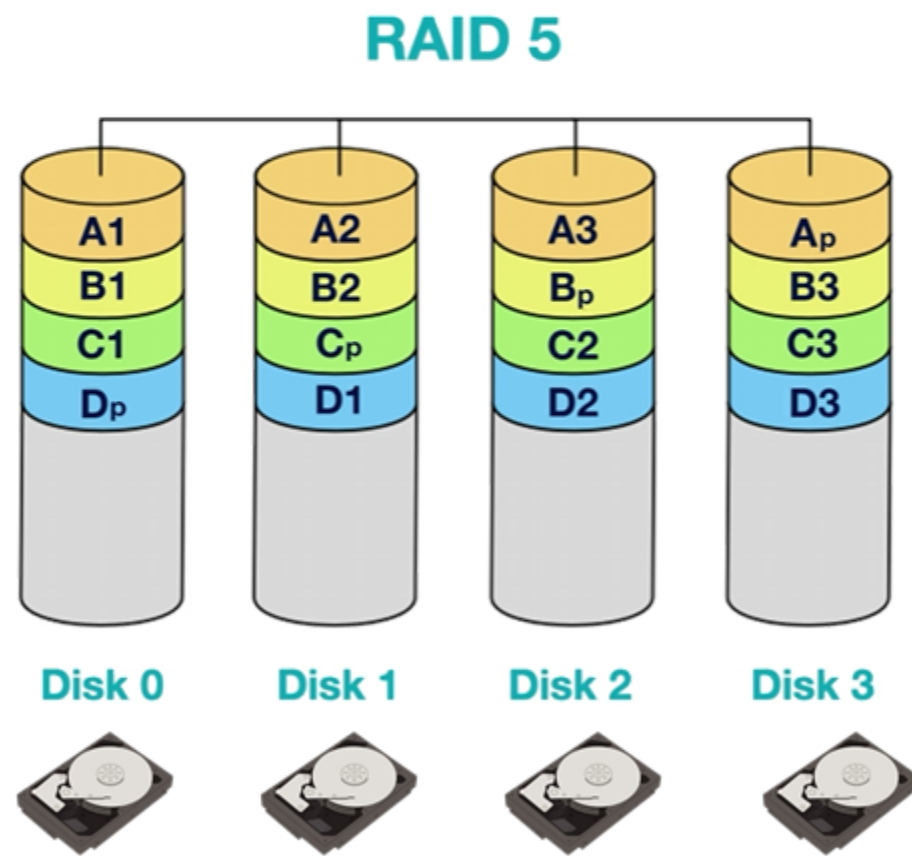
The data is copied into the second storage device so, if one of those storage devices fails, the other one can continue to operate.



### RAID 5 (Striping with Parity):

**RAID 5;** is used to achieve fault tolerance and data redundancy it requires the use of at least three physical storage devices in order for it to work.

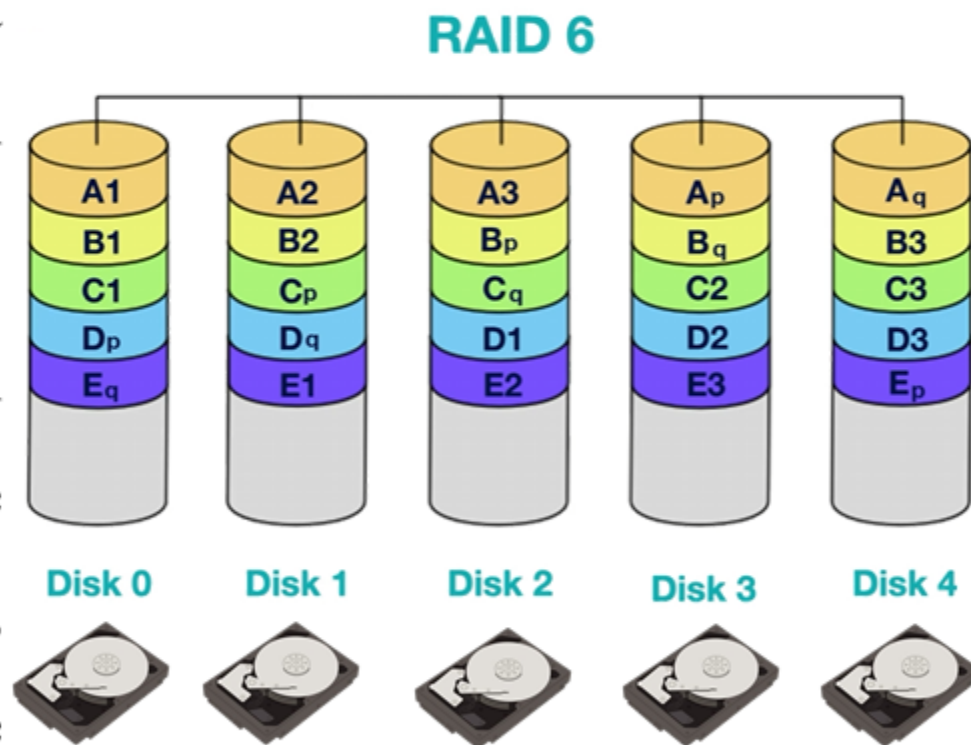
**RAID 5** is striping the data across multiple physical storage devices, as well as writing the parity data across the disk array. If one of storage devices fail, replace it with new one and repair the RAID to restore the missing data.



### RAID 6 (Striping with double Parity):

**RAID 6;** is used to achieve fault tolerance and data redundancy it requires the use of at least four physical storage devices in order for it to work.

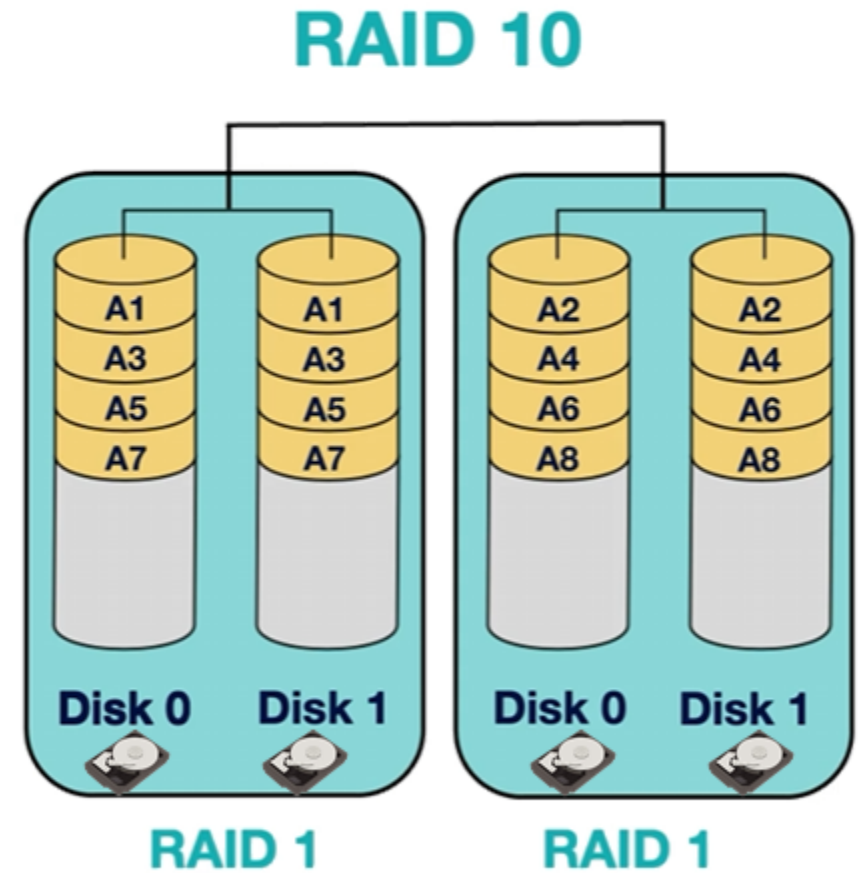
**RAID 6** is striping the data across multiple physical storage devices, as well as writing the parity data twice across the disk array. If one of storage devices fail, replace it with new one and repair the RAID to restore the missing data.



### RAID 10 (Striping with Mirroring):

**RAID 10;** is going to be a striped array that contains two sets of mirrored arrays within it. This is why we call it a **RAID 10**.

**RAID 10** first mirrors the two sets of storage devices into a RAID 1, and then it stripes the data across both sets of mirrored storage devices using a RAID 0 configuration. RAID 1 plus RAID 0 gives us RAID 10.



***Data Redundancy******Failure-Resistance***

*Designed to withstand hardware malfunctions without losing the data.*

***RAID 1, RAID 10***

***Fault-Tolerant***

*Designed to be able to quickly rebuild any lost data from the remaining healthy storage devices.*

***RAID 1, RAID 5, RAID 6,  
RAID 10***

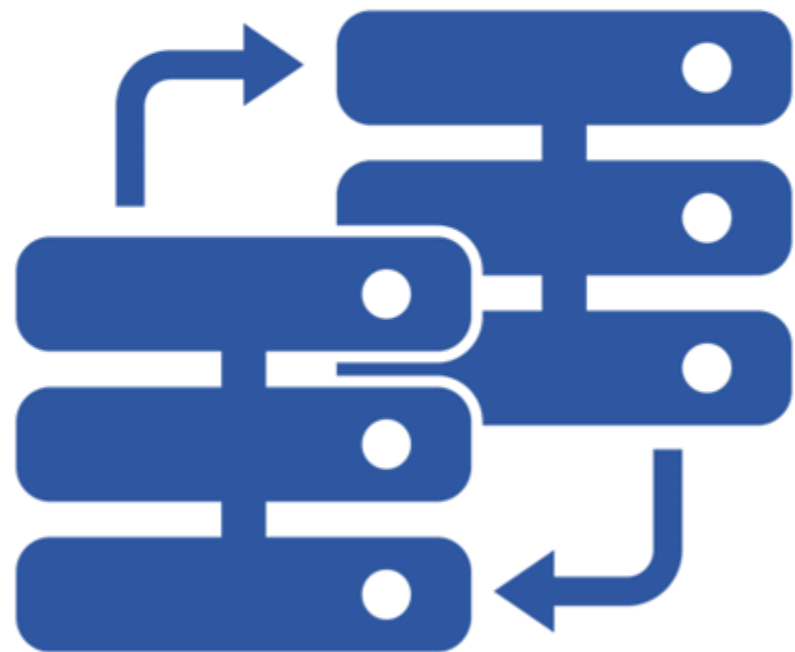
***Disaster-Tolerant***

*Designed to protect data against catastrophic events by utilizing two independent zones that always have full access to all of the data*

***RAID 1, RAID 10***

### What is Data Backup?

***Data Backup;*** is creating a copy of the data on another system that may use for recovery in case the original data is lost or corrupted.



- Onsite and Offsite backups; *onsite backup* is a backup of the data where it's physically going to be *stored in same place as the original data*. While *offsite backup* is a backup of the data where *it is geographically separate locations from the original data source* in order to provide protections against physical disasters.
- Frequency of backups; is *how many should perform backup operation*, in fact that depends on *how frequently your data is going to be changed inside of the business* and *the organization's Recovery Point Objective (RPO)* which is a time-based measurement of the maximum amount of data loss that is tolerable to an organization. The amount of resources is also considered such as time, storage space, bandwidth, and cost.

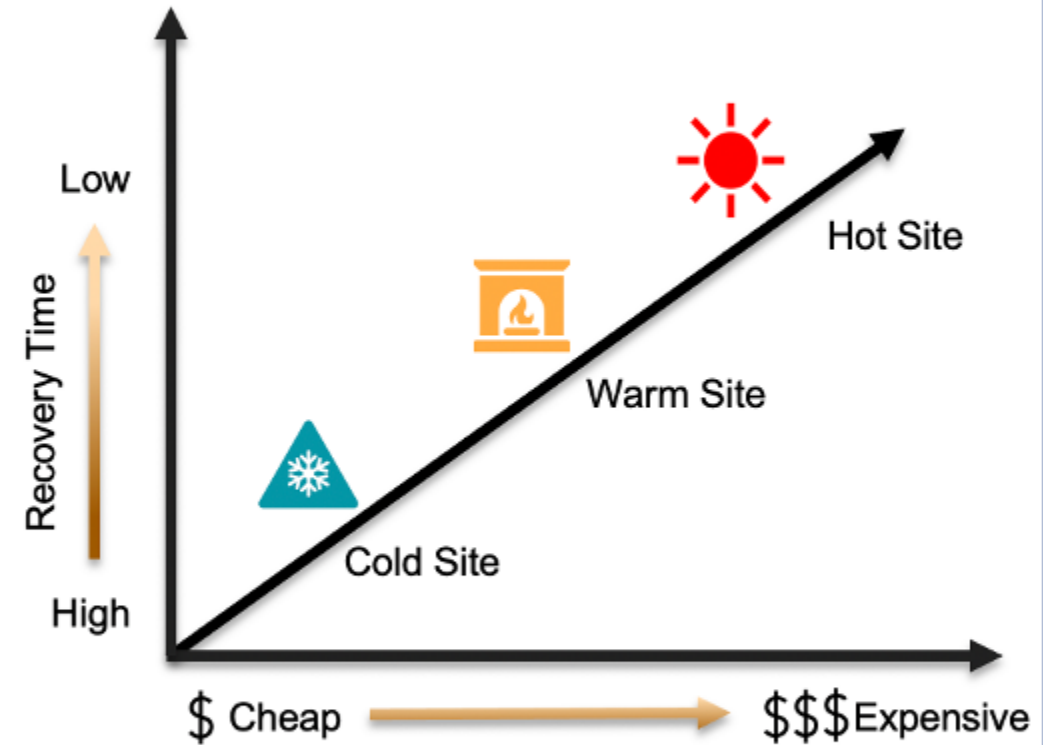
- **Encryption with backups**; is encrypting the backup files to ensure that even if the backup media or files do fall into the wrong hands, the data on them will remain unintelligible without the proper decryption key. The encryption with backups should also *use both data-at-rest encryption and data-in transit encryption*.
- **Snapshot**; is taking a picture of the server data at one specific moment in time and they are only going to *record the changes made since the previous snapshot*. This efficiency not only *reduces your storage requirements*, but it also allows for a *quicker backup process* and more frequent capture of any data changes.

➤ **Data Recovery**: is the process of restoring lost, corrupted, or accidentally deleted data to its original location (server, computer, mobile device or storage device). There are several key steps in the data recovery process:

- Selection of the right backup.
- Initiating the recovery process.
- Data validation and testing the recovery process.
- Documentation and reporting.
- Notification.

- **Replication**; is making *real-time or near real-time copies of the data*. It will keep the data stored in two places simultaneously. If one server crashes, the other will be able to continue without interruption.
- **Journaling**; is making *maintaining a record of every change made to an organization's data over time*. This comprehensive journal or log will serve as a *historical account of the data modifications* and will offer valuable insights to be able to enable the precise data recovery to a specific point in time if a disaster strikes.

In cybersecurity and *disaster recovery planning*, *hot, warm, and cold sites* are backup facilities that organizations can use to restore operations if their primary data center goes down due to a cyberattack, natural disaster, or other disruption. They differ in cost, and recovery speed:



➤ **Hot Site;**

- A fully equipped and operational backup site that mirrors the primary site *in real time*.
- Hardware, software, and network connectivity are already in place.
- Data is synchronized continuously (or nearly real time).
- Staff can start working almost immediately.
- **Recovery Time Objective (RTO):** Minutes to a few hours.
- **Cost:** Very high (most expensive).
- **Use Case:** Critical organizations (banks, stock exchanges, healthcare, defense) that cannot tolerate downtime.

➤ **Warm Site;**

- A partially equipped site that has the necessary hardware and network connections *but may not have the latest data or fully installed software.*
- Systems may need configuration or restoration of recent backups.
- Data may be a few hours to a few days old.
- Some manual effort is required before operations resume.
- **Recovery Time Objective (RTO):** Several hours to a few days.
- **Cost:** Moderate.
- **Use Case:** Businesses that can tolerate short downtime but still need relatively quick recovery.

➤ **Cold Site;**

- A backup facility with basic infrastructure (power, cooling, and space) but *no active equipment or live data*.
- No servers, storage, or networking ready.
- Organization must bring in hardware and restore data from backups.
- Long delay before full operation.
- **Recovery Time Objective (RTO):** Days to weeks.
- **Cost:** Low (least expensive).
- **Use Case:** Smaller organizations or those with higher tolerance for downtime but still needing a place to recover if disaster strikes.

## 3 TYPES OF DR SITES



### Cold Site

- A cold site is an off-site location with basic infrastructure but lacks the necessary equipment, data, and applications.
- It is the most cost-effective option but has the longest recovery time.
- Organizations need to set up and configure the site before it becomes functional.



### Hot Site

- A hot site is a fully operational duplicate of the primary data center with up-to-date data and applications.
- It can take over operations immediately in case of a disaster.
- Offers the fastest recovery time but is also the most expensive option.



### Warm Site

- A warm site is a partially equipped backup facility with essential infrastructure and some data.
- It requires more time to become fully operational compared to a hot site.
- It provides a balance between cost and recovery time.

