

**IT8001**

# Information Storage and Management

Professional Elective

Topic

Information Storage



**Varun C.M.**

Assistant Professor, Department of Information Technology,  
St. Xavier's Catholic College of Engineering

## CO1

Understand the logical and physical components of a Storage infrastructure.

- A **data center** is a facility that contains information storage and other physical information technology (IT) resources for computing, networking, and storing information.
- Cloud computing brings in a fully automated request fulfillment process that enables users to rapidly obtain storage and other IT resources on demand.
- Through cloud computing, an organization can rapidly deploy applications where the underlying storage capability can scale-up and scale-down, based on the business requirements.

Storage is a repository that enables users to persistently store and retrieve the digital data.

## 1. Data

Data is a collection of raw facts from which conclusions might be drawn.

The following is a list of some of the factors that have contributed to the growth of digital data,

- Increase in data-processing capabilities
- Lower cost of digital storage
- Affordable and faster communication technology
- Proliferation of applications and smart devices

## Types of Data

**Structured  
Data**

Data is unstructured if its elements cannot be stored in rows and columns, which makes it difficult to query and retrieve by applications.

Structured data is organized in rows and columns in a rigidly defined format so that applications can retrieve and process it efficiently. Structured data is typically stored using a database management system (DBMS).

**Unstructured  
Data**

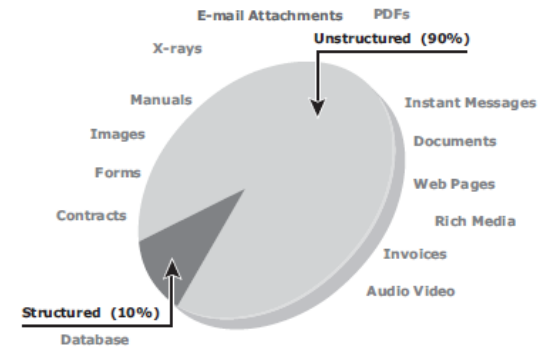


Figure 1-2: Types of data

Image courtesy EMC Corporation, "Information Storage and Management"

## 2. Big Data

Big data is a new and evolving concept, which refers to data sets whose sizes are beyond the capability of commonly used software tools to capture, store, manage, and process within acceptable time limits.

The big data ecosystem consists of the following:

1. Devices that collect data from multiple locations and also generate new data about this data (metadata).
2. Data collectors who gather data from devices and users.
3. Data aggregators that compile the collected data to extract meaningful information.
4. Data users and buyers who benefit from the information collected and aggregated by others in the data value chain.

Data science is an emerging discipline, which enables organizations to derive business value from big data.

Data science represents the synthesis of several existing disciplines, such as statistics, math, data visualization, and computer science to enable data scientists to develop advanced algorithms for the purpose of analyzing vast amounts of information to drive new value and make more data-driven decisions.

## 3. Information

Information is the intelligence and knowledge derived from data.

## 4. Storage

In a computing environment, devices designed for storing data are termed storage devices or simply storage.

The type of storage used varies based on the type of data and the rate at which it is created and used.

Devices, such as a media card in a cell phone or digital camera, DVDs, CD-ROMs, and disk drives in personal computers are examples of storage devices.



## Server-Centric Storage Architecture

In earlier implementations of open systems, the storage was typically internal to the server. These storage devices could not be shared with any other servers.

In this architecture, each server has a limited number of storage devices, and any administrative tasks, such as maintenance of the server or increasing storage capacity, might result in unavailability of information.

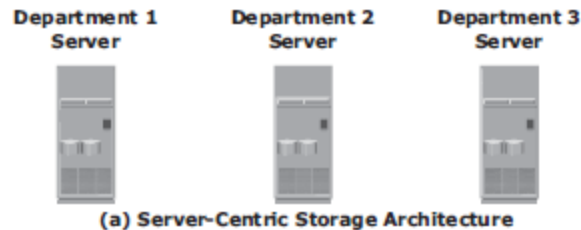


Image courtesy EMC Corporation, "Information Storage and Management"

## Information-Centric Architecture

In this architecture, storage devices are managed centrally and independent of servers.

These centrally-managed storage devices are shared with multiple servers.

When a new server is deployed in the environment, storage is assigned from the same shared storage devices to that server.

The capacity of shared storage can be increased dynamically by adding more storage devices without impacting information availability

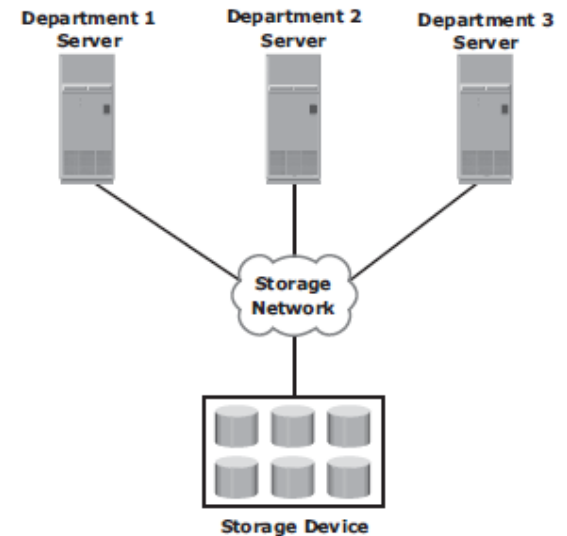


Image courtesy EMC Corporation, "Information Storage and Management"

Data centers house and manage large amounts of data.

The data center infrastructure includes **hardware components**, such as computers, storage systems, network devices, and power backups; and **software components**, such as applications, operating systems, and management software. It also includes **environmental controls**, such as air conditioning, fire suppression, and ventilation.

## Core Elements of a Data Center

### Application

A computer program that provides the logic for computing operations

### DBMS

Provides a structured way to store data in logically organized tables that are interrelated

### Host

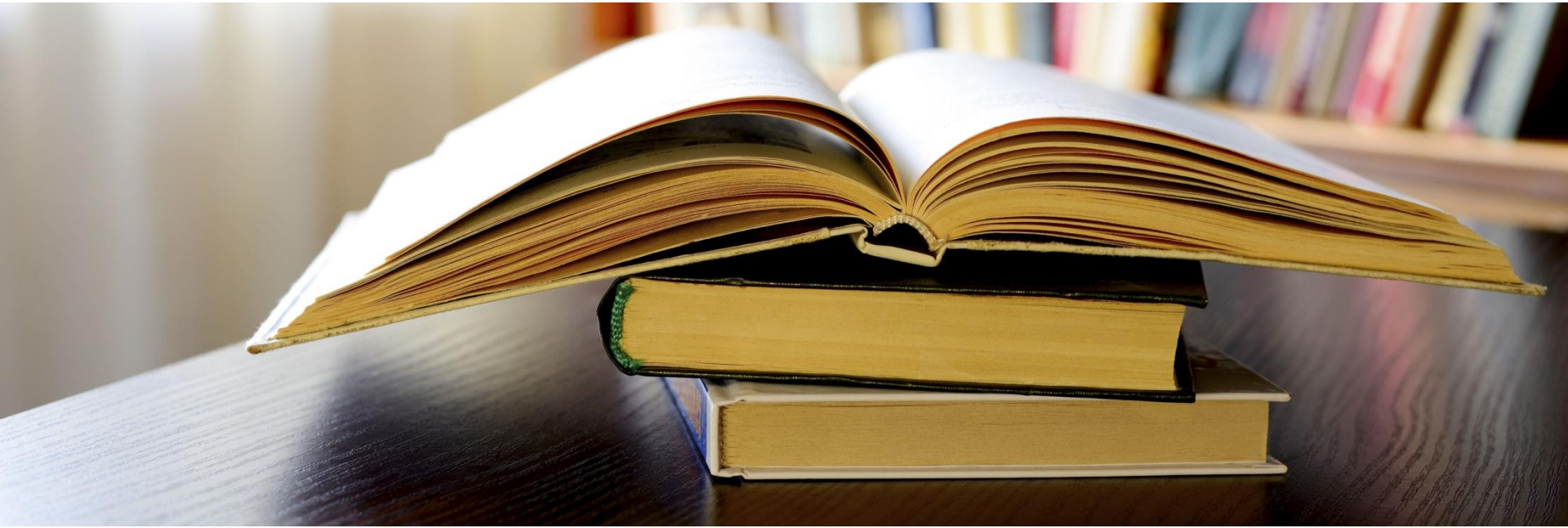
A computing platform (hardware, firmware, and software) that runs applications and databases

### Network

A data path that facilitates communication among various networked devices

### Storage

A device that stores data persistently for subsequent use



## References

EMC Corporation, "Information Storage and Management" , Wiley, India.  
Robert Spalding, "Storage Networks: The Complete Reference", Tata McGraw Hill, Osborne, 2003.



Varun C.M.

Assistant Professor, Department of Information Technology, St. Xavier's Catholic College of Engineering