Module 1: Introduction to Information Storage

Upon completion of this module, you should be able to:

- Describe digital data, types of digital data, and information
- Describe data center and its key characteristics
- Describe key data center management processes
- Describe the evolution of computing platforms

The Growth of the Digital Universe

- The digital universe is created and defined by software
  - Digital data is continuously generated, collected, stored, and analyzed through software
- The digital universe generates approximately 4.4 trillion GB of data annually(44 in 2020)
  - Proliferation of IT, Internet usage, social media, and smart devices adds to data growth
- The Internet of Things (IoT) is also adding to data growth
  - IoT is made up of Internet-connected equipment and sensors

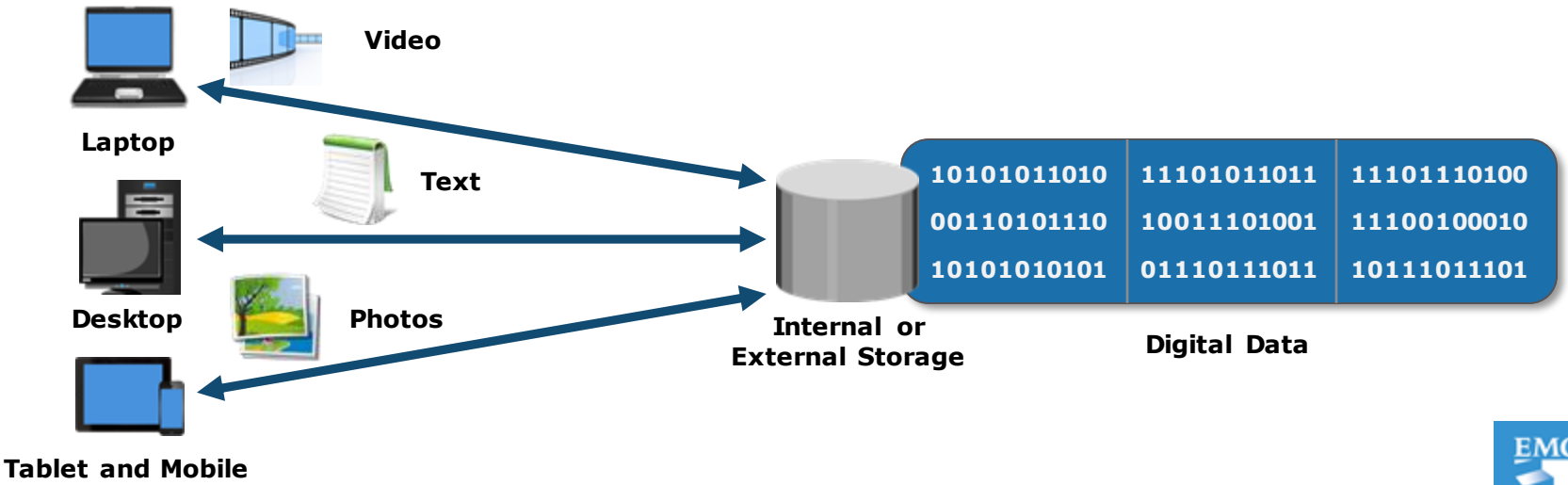
Why Information Storage and Management?

- Organizations are dependent on continuous and reliable access to information
- Organizations seek to effectively store, protect, process, manage, and leverage information
- Organizations are increasingly implementing intelligent storage solutions
  - To efficiently store and manage information
  - To gain competitive advantage
  - To derive new business opportunities

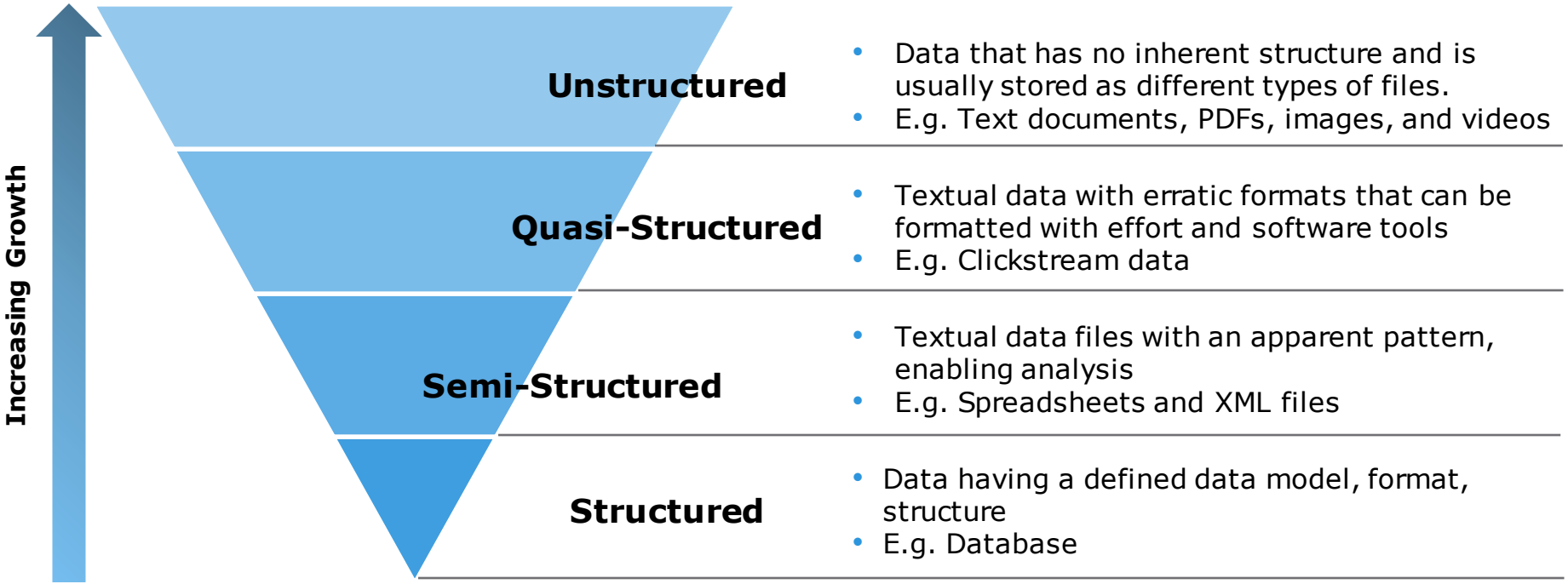
What is Digital Data?

Digital Data

A collection of facts that is transmitted and stored in electronic form, and processed through software.



Types of Digital Data



- Data that has no inherent structure and is usually stored as different types of files.
- E.g. Text documents, PDFs, images, and videos

- Textual data with erratic formats that can be formatted with effort and software tools
- E.g. Clickstream data

- Textual data files with an apparent pattern, enabling analysis
- E.g. Spreadsheets and XML files

- Data having a defined data model, format, structure
- E.g. Database

What is Information?

Information

Processed data that is presented in a specific context to enable useful interpretation and decision-making.

- Example: Annual sales data processed into a sales report
  - Enables calculation of the average sales for a product and the comparison of actual sales to projected sales
- New architectures and technologies have emerged for extracting information from non-structured data

Information Storage

- Information is stored on storage devices on non-volatile media
- Types of storage devices:
  - **Magnetic storage devices**: Hard disk drive and magnetic tape
  - **Optical storage devices**: Blu-ray disc, DVD, and CD
  - **Flash-based storage devices**: Solid state drive, memory card, and USB thumb drive
- Storage devices are assembled within a storage system or "array"
  - Provides high capacity, scalability, performance, reliability, and security
- Storage systems along with other IT infrastructure are housed in a data center

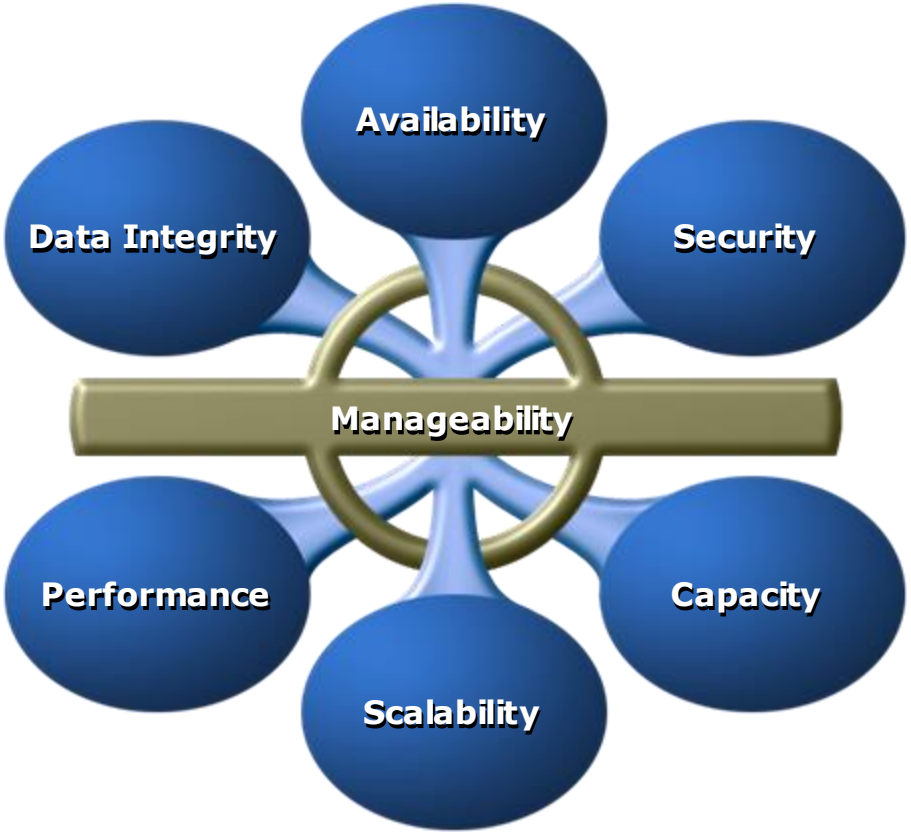
What is a Data Center?

Data Center

A facility that houses IT equipment including compute, storage, and network components, and other supporting infrastructure for providing centralized data-processing capabilities.

- A data center comprises:
  - **Facility**: The building and floor space where the data center is constructed
  - **IT equipment**: Compute, storage, and network equipment
  - **Support infrastructure**: Power supply, fire detection, HVAC, and security systems

Key Characteristics of a Data Center



Key Data Center Management Processes

Management Process	Description
Monitoring	Continuously gathering information on data center resources
Reporting	Presenting the details on resource performance, capacity, and utilization
Provisioning	Configuring and allocating resources to meet the capacity, availability, performance, and security requirements
Planning	Estimating the amount of resources required to support business operations
Maintenance	Ensuring the proper functioning of resources and resolving incidents

Evolution of Computing Platforms

BILLIONS OF USERS

PLATFORM 3

Cloud Big Data Mobile Social

Mobile Devices

MILLIONS OF APPS

HUNDREDS OF MILLIONS OF USERS

PLATFORM 2

LAN/Internet Client/Server

PC

TENS OF THOUSANDS OF APPS

MILLIONS OF USERS

PLATFORM 1

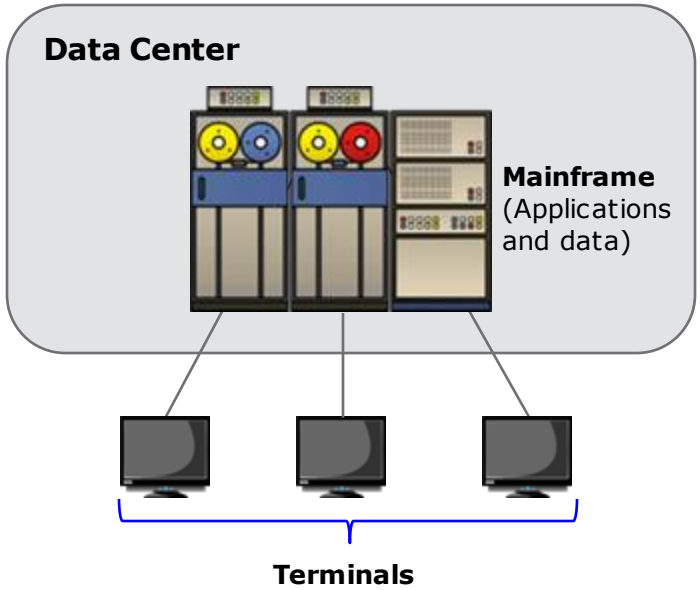
Mainframe, Mini Computer

Terminals

THOUSANDS OF APPS

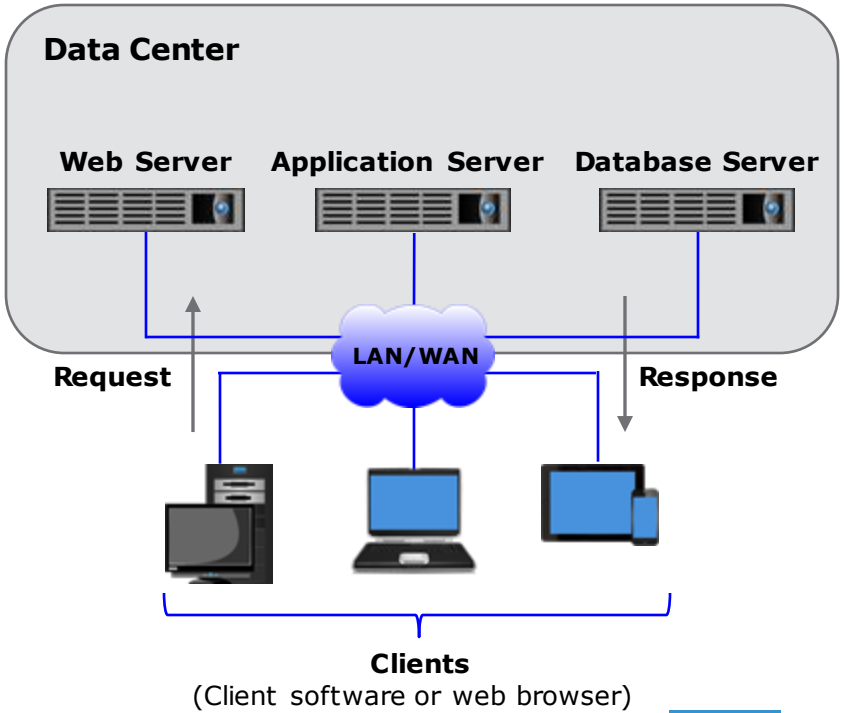
First Platform

- Based on mainframes
  - Applications and databases hosted centrally
  - Users connect to mainframes through terminals
- Challenges with mainframes
  - Substantial CAPEX and OPEX
    - High acquisition costs
    - Considerable floor space and energy requirements

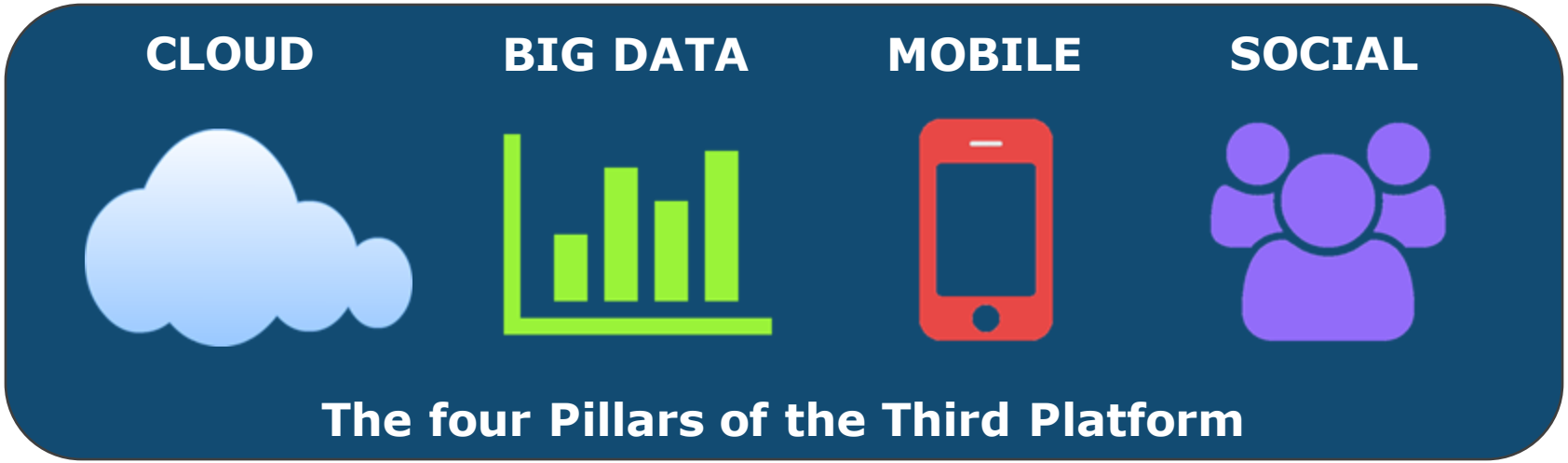


Second Platform

- Based on client-server model
  - Distributed application architecture
  - Servers receive and process requests for resources from clients
  - Users connect through a client program or a web interface
- Challenges with client-server model
  - Creation of IT silos
  - Hardware and software maintenance overhead
  - Scalability to meet the growth of users and workloads



Third Platform



- The four pillars are transforming the way organizations are using technology for business operations

Module 1: Summary

Key points covered in this module:

- Digital data, types of digital data, and information
- Data center and its key characteristics
- Key data center management processes
- Evolution of computing platforms