INTRODUCTION TO INFORMATION STORAGE

Module 1: Introduction to Information Storage

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

- Define data and information
- Describe types of data
- Describe the evolution of storage architecture
- Describe the core elements of a data center
- List the key characteristics of data center
- Provide an overview of virtualization and cloud computing
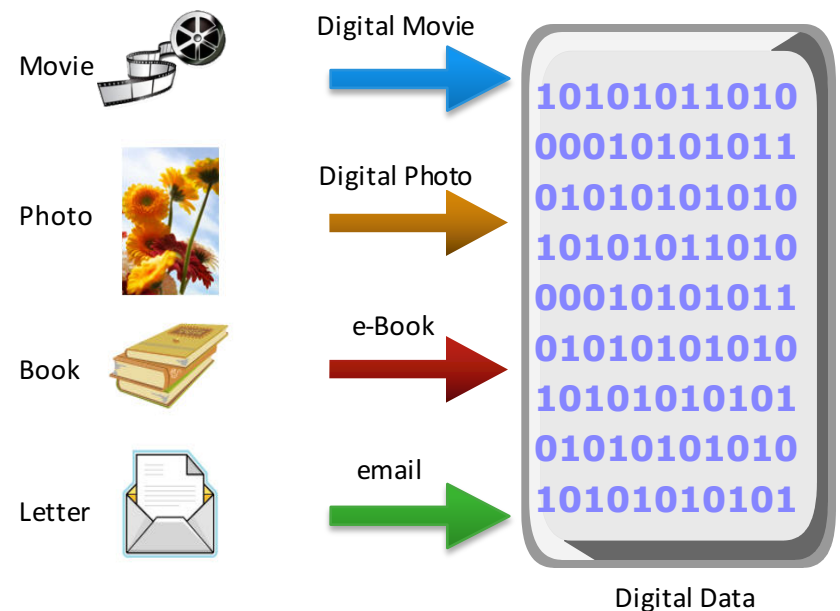
Why Information Storage and Management?

- Information is the knowledge derived from data
- Growth of digital information has resulted in information explosion
- We live in an on-command, on-demand world
  - We need information when and where required
- Increasing dependency on fast and reliable access to information
- Businesses seek to store, protect, optimize, and leverage the information
  - To gain competitive advantage
  - To derive new business opportunity
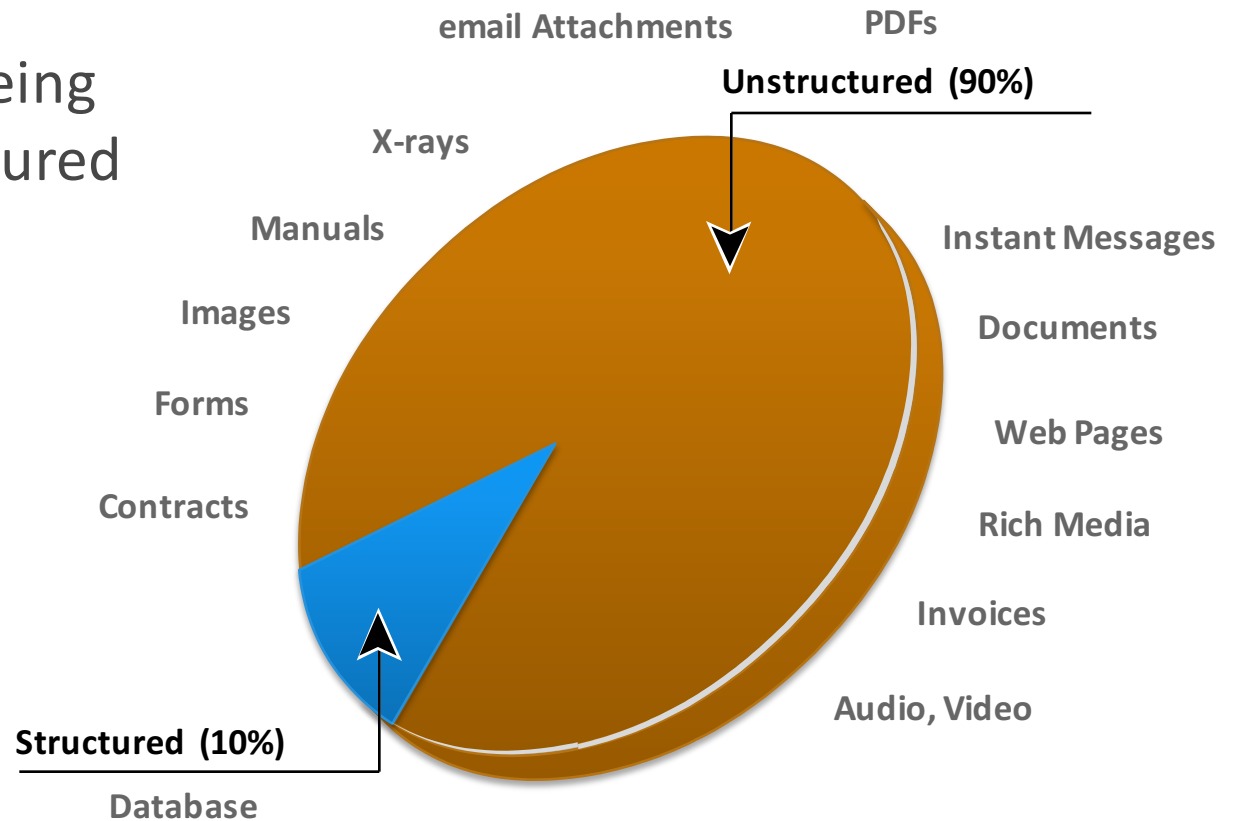
What is Data?

Data

It is a collection of raw facts from which conclusions may be drawn.

- Data is converted into more convenient form – digital data
- Factors for digital data growth are:
  - ▸ Increase in data-processing capabilities
  - ▸ Lower cost of digital storage
  - ▸ Affordable and faster communication technology
  - ▸ Proliferation of applications and smart devices

Movie → Digital Movie

Photo → Digital Photo

Book → e-Book

Letter → email

10101011010
00010101011
01010101010
10101011010
00010101011
01010101010
10101010101
01010101010
10101010101

Digital Data

Introduction to Information Storage

Types of Data

- Data can be classified as:
  - ▸ Structured
  - ▸ Unstructured
- Majority of data being created is unstructured

email Attachments   PDFs

Unstructured (90%)

X-rays

Manuals   Instant Messages

Images   Documents

Forms   Web Pages

Contracts   Rich Media

                                             Invoices

                                             Audio, Video

Structured (10%)

Database

Big Data

Big Data

It refers to data sets whose sizes are beyond the ability of commonly used software tools to capture, store, manage, and process within acceptable time limits.

- Includes both structured and unstructured data generated by variety of sources

- Big data analysis in real time requires new techniques and tools that provide:
  - ▸ High performance
  - ▸ Massively parallel processing (MPP) data platforms
  - ▸ Advanced analytics

- Big data analytics provide an opportunity to translate large volumes of data into right decisions
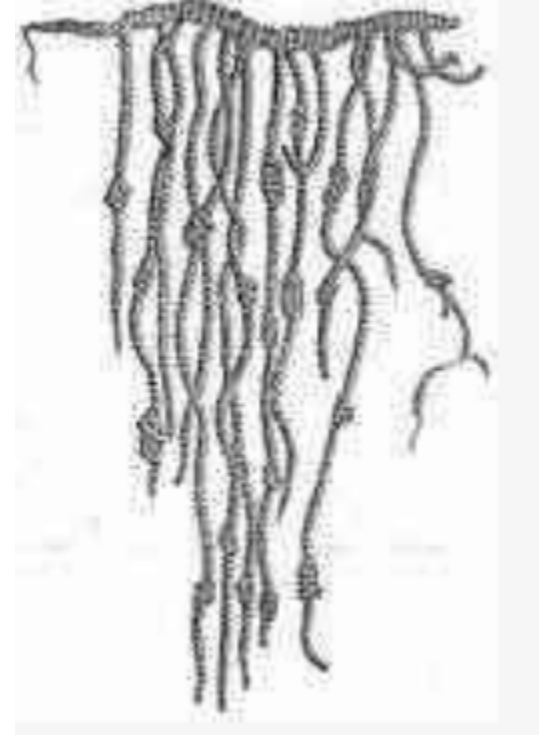
Storage

- Stores data created by individuals and organizations
  - ▶ Provides access to data for further processing
- Examples of storage devices are:
  - ▶ Media card in a cell phone or digital camera
  - ▶ DVDs, CD-ROMs
  - ▶ Disk drives
  - ▶ Disk arrays
  - ▶ Tapes

Data Creation: Individuals

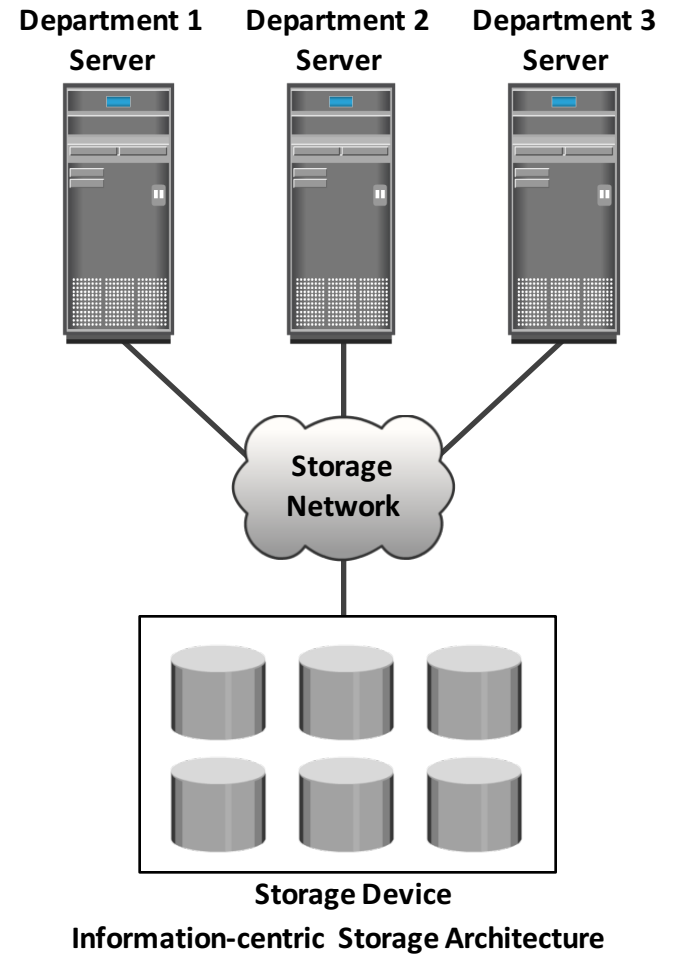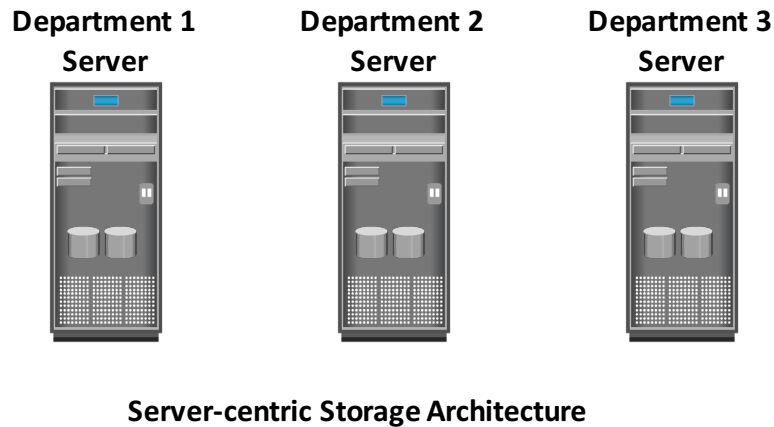- What data is created by a individuals?

殷墟文字

上古结绳而治，

后世圣人易之以书契

Data Creation: Business

- What data is created by a business?
- Examples of data created by a business include:
  - Product data: inventory, description, pricing, availability, sales numbers and projections
  - Customer data: orders, shipping details
  - Account data: banking, financial services industry
  - Medical data: health care providers, insurance industry, hospitals

Evolution of Storage Architecture

**Department 1
Server**

**Department 2
Server**

**Department 3
Server**

Server-centric Storage Architecture

**Department 1
Server**

**Department 2
Server**

**Department 3
Server**

**Storage
Network**

Storage Device

Information-centric Storage Architecture
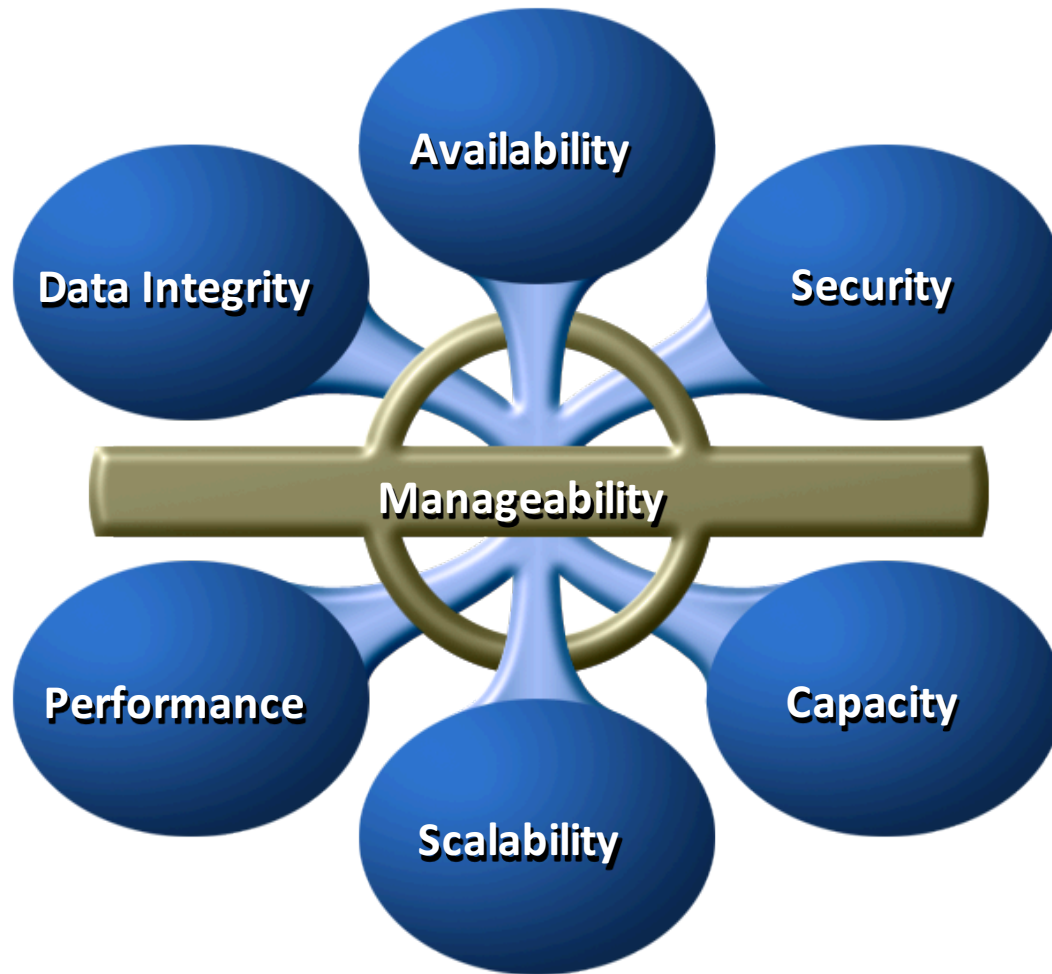
Data Center

Data Center

It is a facility that contains storage, compute, network, and other IT resources to provide centralized data-processing capabilities.

- Core elements of a data center
  - ▸ Application
  - ▸ Database management system (DBMS)
  - ▸ Host or Compute
  - ▸ Network
  - ▸ Storage
- These core elements work together to address data-processing requirements

Data Center: Online Order Transaction System Example

Storage Array

Host/ Compute

Client

LAN/WAN

Storage Network

User Interface

OS and DBMS
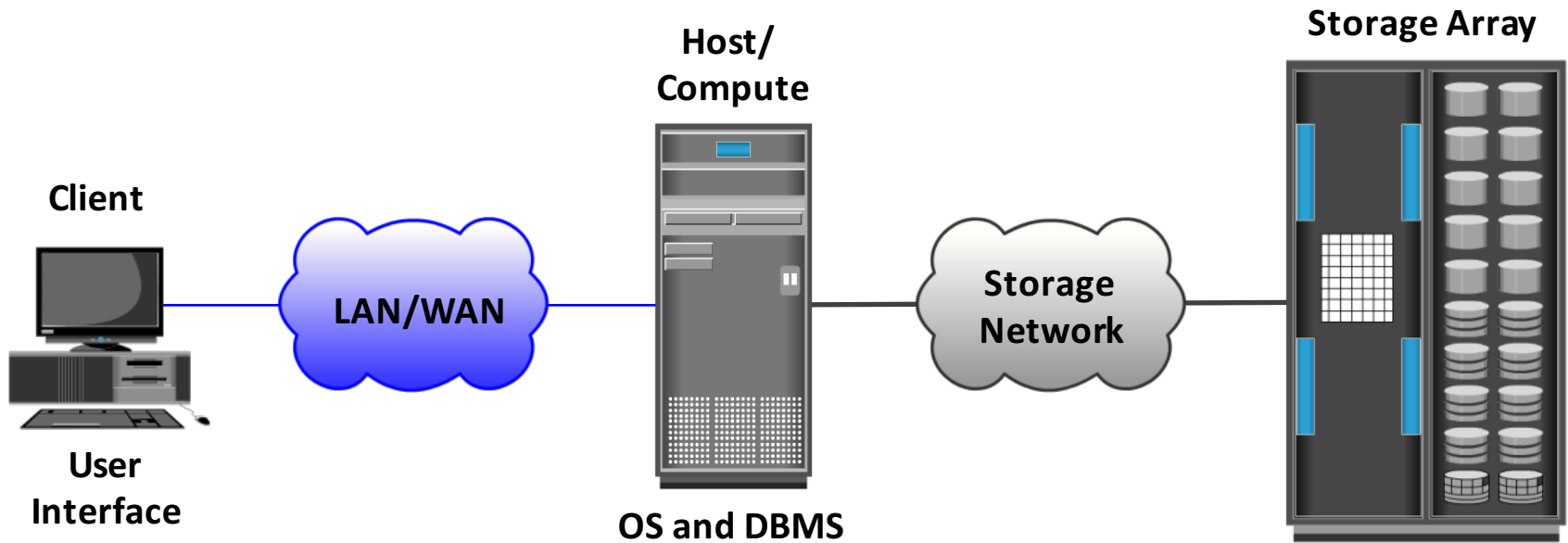
Key Characteristics of a Data Center

Managing Data Center

- Key management activities include
  - ▸ Monitoring
    - ⏩ Continuous process of gathering information on various elements and services running in a data center
  - ▸ Reporting
    - ⏩ Details on resource performance, capacity, and utilization
  - ▸ Provisioning
    - ⏩ Configuration and allocation of resources to meet the capacity, availability, performance, and security requirements
- Virtualization and cloud computing have changed the way data center infrastructure resources are provisioned and managed

Virtualization: An Overview

- Virtualization is a technique of abstracting physical resources and making them appear as logical resources

  ▸ For example partitioning of raw disks

- Pools physical resources and provides an aggregated view of physical resource capabilities

- Virtual resources can be created from pooled physical resources

  ▸ Improves utilization of physical IT resources

Cloud Computing: An Overview

- Enables individuals and organizations to use IT resources as a service over network

- Enables self-service requesting and automates request-fulfillment process
  - ‣ Enables users to scale up or scale down the usage of computing resources quickly

- Enables consumption-based metering
  - ‣ Consumers pay only for the resources they use
    - ‣‣ Example: CPU hours used, amount of data transferred, and Gigabytes of data stored

学院 云平台

Search

◎ 虚拟机管理

◎ 实验虚拟机

▦ 实验信息

虚拟机表

➕新增虚拟机   ✏修改   ✖删除

☐	虚拟机名称	是否开机	状态

Search

◎ 虚拟机管理

◎ 实验虚拟机

▦ 实验信息

▦ 申请记录

实验信息

未完成实验

序号	实验名	所属课程	教师	学期	开始时间	结束时间
1	在线协同团队教学系统	生产实习	噜噜	2016-2017 第二学期	2017-07-21	2017-09-15
2	对方答复熟悉	中标麒麟	路新喜	2017年秋季	2017-09-01	2017-09-22

Showing 1 to 2 of 2 rows

所有实验

学期 当前学期(2017年秋 ▼)   课程 全部课程 ▼

序号	实验名	所属课程	教师	学期	开始时间	结束时间

课程微信群

面对面建群：
密码：2233

Module 1: Summary

Key points covered in this module:

- Data and information
- Types of data
- Big data
- Evolution of storage architecture
- Core elements of data center
- Key characteristics of data center
- Virtualization and cloud computing