

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

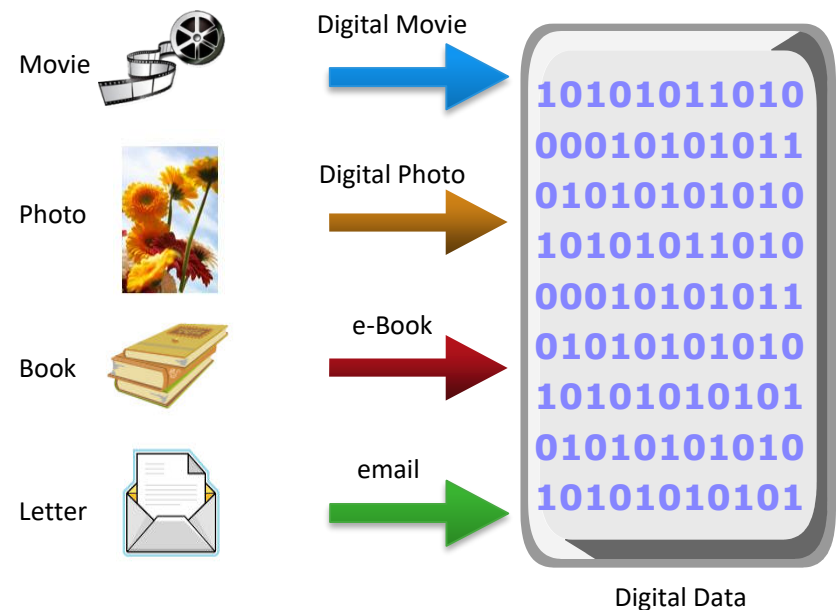
| WEB 1.0 | WEB 2.0 | WEB 3.0 |
|---------|---------|---------|
| 信息的获取 | 信息报站 | 信息过剩 |
| 信息搜索 | 归类和存储 | 筛选整合和服务 |
| 数据通信 | 数据存储 | 数据处理 |

What is Data?

Data

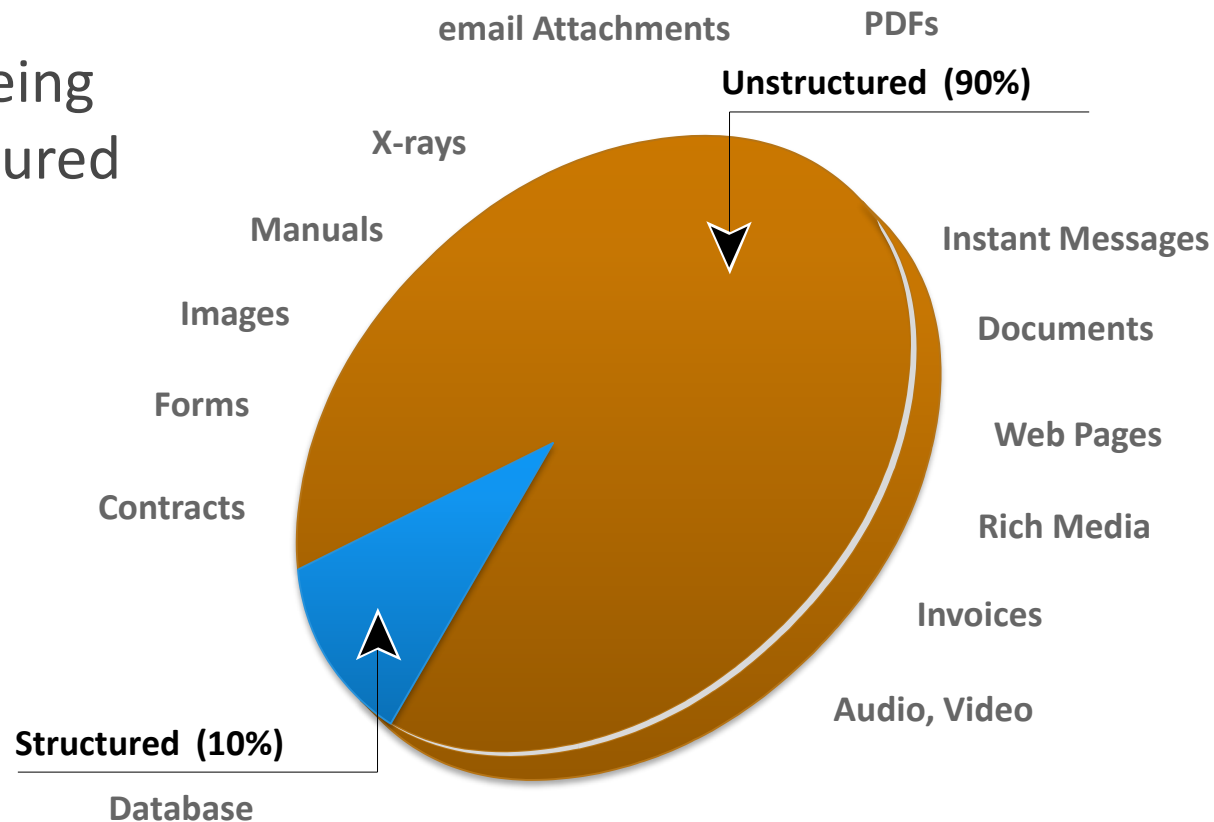
It is a collection of raw facts from which conclusions may be drawn.

- 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



Types of Data

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



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

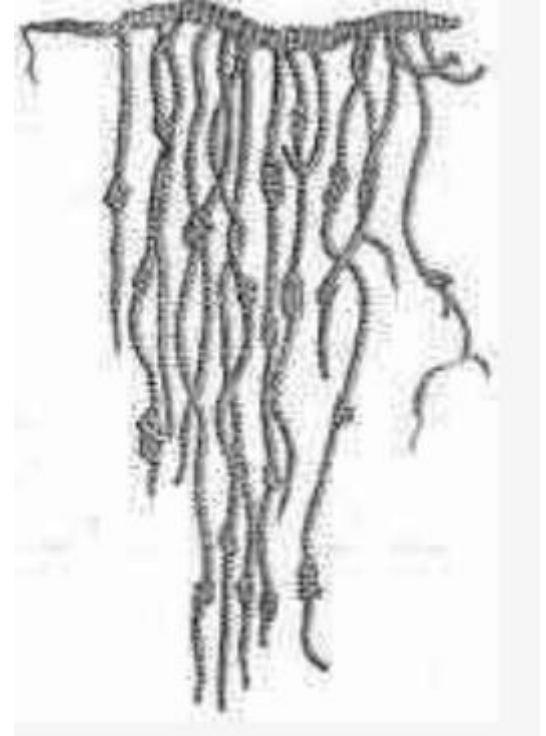
- Stores data created by individuals and organizations
 - ▶ Provides access to data for further processing
- Examples of storage devices are:
 - ▶ Media card in a cell phone or digital camera
 - ▶ DVDs, CD-ROMs
 - ▶ Disk drives
 - ▶ **Disk arrays**
 - ▶ Tapes

Data Creation: Individuals

- What data is created by a individuals?



殷墟文字



上古结绳而治，

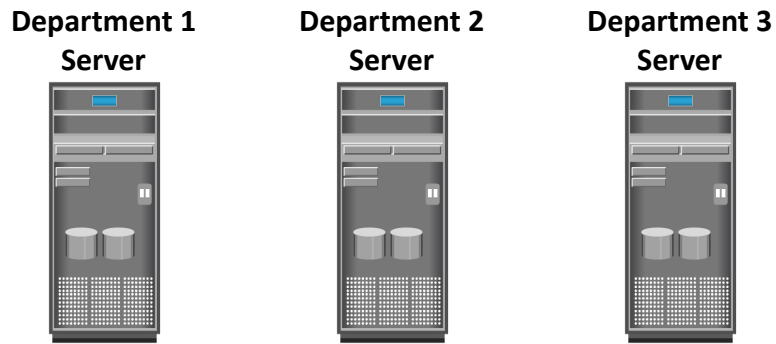
后世圣人易之以书契

Data Creation: Business

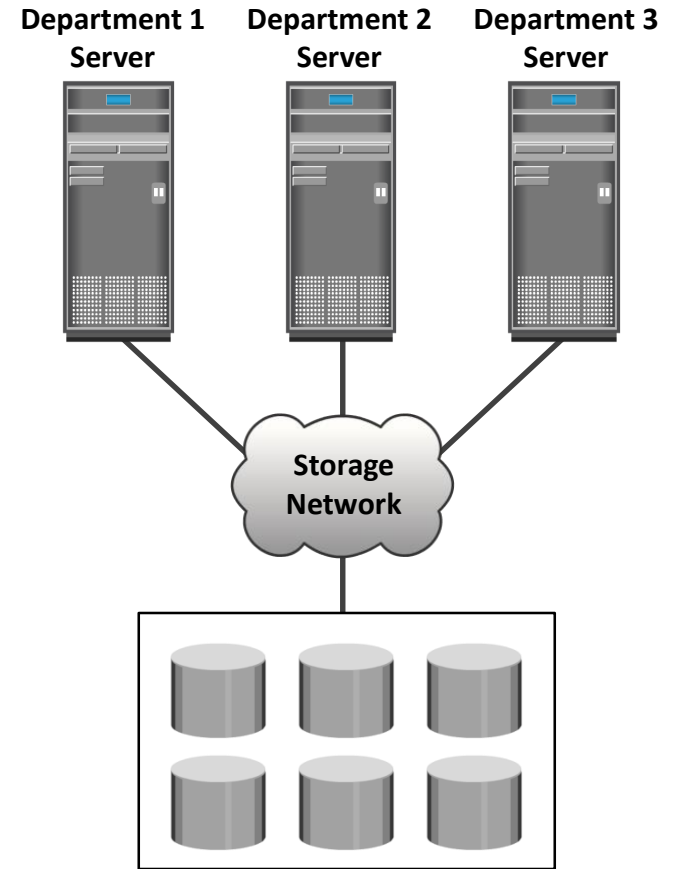
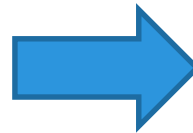
- What data is created by a business?
- Examples of data created by a business include:
 - ▶ Product data: inventory, description, pricing, availability, sales numbers and projections
 - ▶ Customer data: orders, shipping details
 - ▶ Account data: banking, financial services industry
 - ▶ Medical data: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



Evolution of Storage Architecture



Server-centric Storage Architecture



Storage Device
Information-centric Storage Architecture

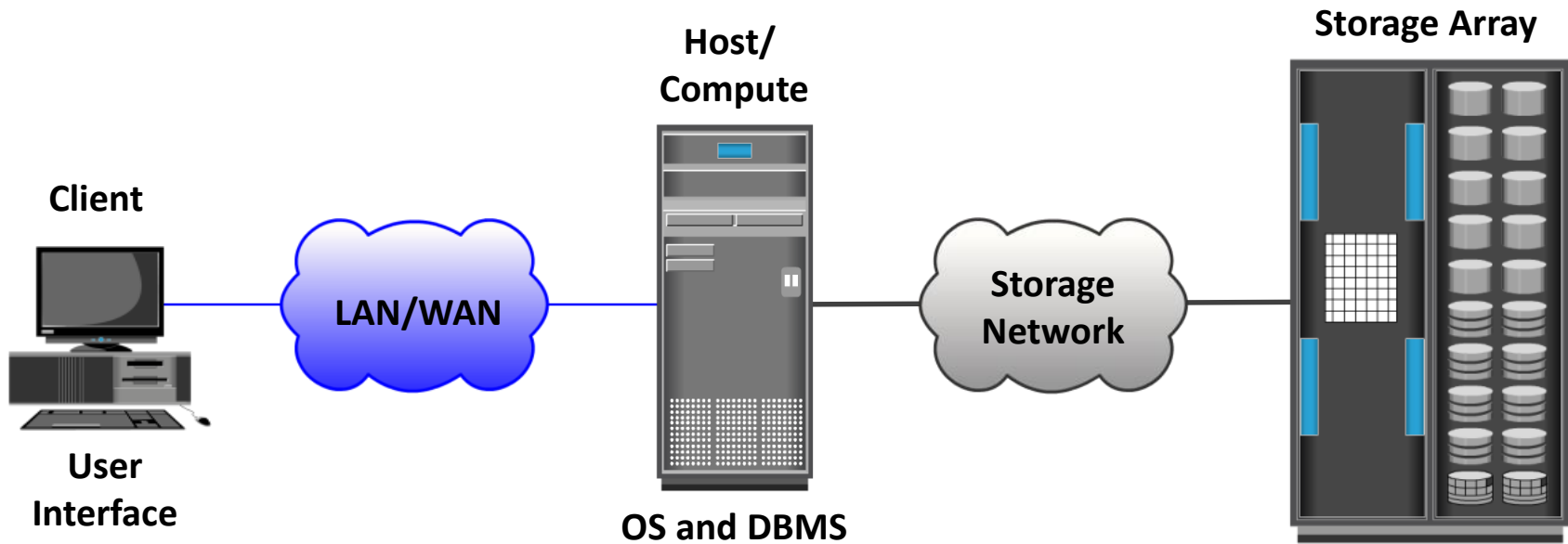
Data Center

Data Center

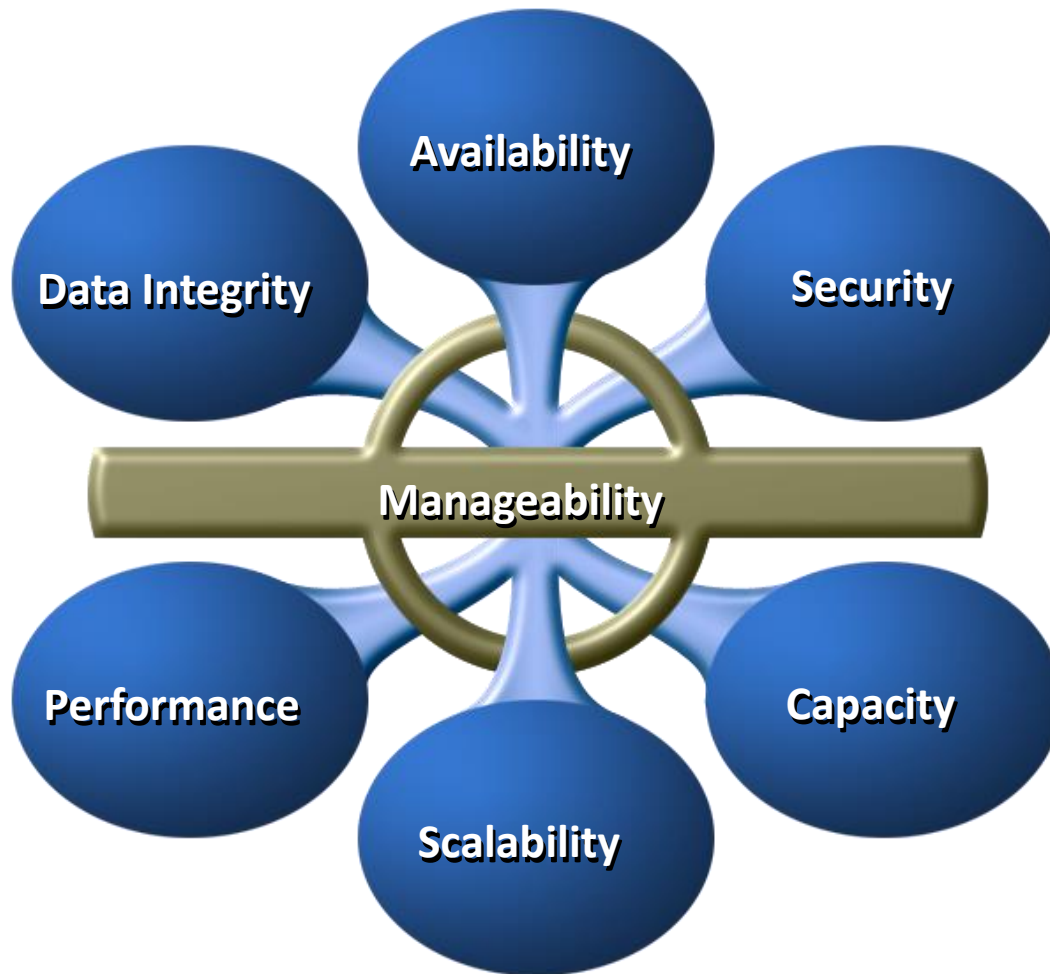
It is a facility that contains storage, compute, network, and other IT resources to provide centralized data-processing capabilities.

- 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



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

课程|实验平台: scs.buaa.edu.cn



用户登录



[忘记密码](#)

[激活账户](#)

登 录

[统一认证登录](#)

新增课程

新增实验

新增作业

序号

课程名称

1

网络存储B

2

网络存储A

3

大数据技术综合实验

4

程序设计实践

5

大数据和云计算[B3J213240]

6

走进软件

7

大数据技术综合实验

8

网络存储

9

程序设计实践

10

云计算[B3J213240]

查看课程《网络存储A》中的学生

全部删除

刷新

显示设置

搜索

学生学号

学生姓名

所在学院

学生邮箱

操作

20373913

宗恩立

软件学院

798417513@qq.com

删除

20373922

黄志杰

软件学院

1525388533@qq.com

删除

20373952

邹寒羽

软件学院

852081297@qq.com

删除

20373966

刘禹宏

软件学院

787288968@qq.com

删除

20373981

彭志朋

软件学院

hironin@foxmail.com

删除

21210110

杜佳男

软件学院

2603605747@qq.com

删除

21210111

石庚

软件学院

shigeng@buaa.edu.cn

删除

21210115

梅煜坤

软件学院

534590526@qq.com

删除

22210103

待激活用户

软件学院

waitinput

删除

22210108

待激活用户

软件学院

waitinput

删除

共 210 条

<

1

...

19

20

21

>

10 条/页

跳至

21

页

取消

确定

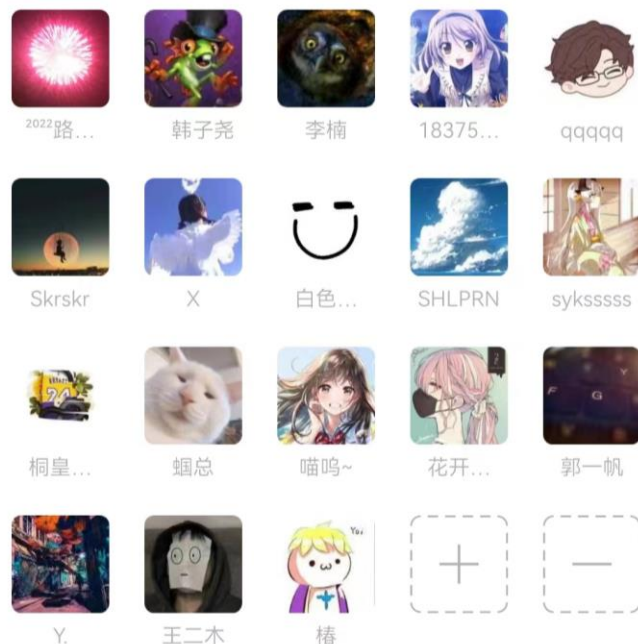
课程微信群

学号姓名实名 (学号-姓名命名) ***

选课人数较多 (210)

难题求助

问题讨论



2022网络存储课程群



该二维码7天内(9月8日前)有效, 重新进入将更新

部分已经完成, 提前入群

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

思考作业

- 学生个人数据2TB数据，你将如何存储？
- 企业数据20TB，有什么好的存储方案？
- 分析并陈述理由？