

课堂练习答案：1-5DCDCA 6-11DDAABC

课后作业答案：1-5DBBCD 6-10ABADC

11-15CADAD 16-19AAAB

## 1.4 数据管理与安全 1.5 数据与大数据

### 【课时目标】

1. 能选择合适的方法管理数据，理解文件的基本属性。
2. 了解数据保护的一般方法，提高数据安全意识。
3. 了解大数据的概念，理解大数据的特征。
4. 了解大数据对人类思维方式的影响，能正确面对大数据。

### 【知识梳理-课前预习】

#### 一、数据管理与安全

1. 数据管理是利用计算机硬件和软件技术对数据进行有效收集、存储、处理和应用的过程，其目的在于充分、有效地发挥数据的作用。
2. 在大数据出现之前，计算机数据的管理已经经历了 人工管理、文件管理和 数据库管理 三个阶段。
3. 数据一般以 文件 的形式存储。根据编码规则的不同，文件的格式也不相同，可以用文件 扩展名 来表示某些特定的 文件类型，如网页文件的文件扩展名为.htm 或.html，而 JPEG 图像文件扩展名为.jpg 等，以区分不同类型的存储数据，如文本、图像、音频等。

文件类型	扩展名
文本文件	.txt .doc .docx .wps .pdf
图像文件	.bmp .jpg .gif .psd .png .tif
声音文件	.wav .mid .mp3 .wma
视频文件	.avi .mpg .flv .mov .dat .rm .rmvb .wmv
网页文件	.htm .html

4. 计算机一般采用 树形目录结构 来管理文件。在 Windows 系统中则采用了更为形象的文件 夹 来管理文件。
5. 传统数据库技术基于 结构化数据 进行开发，凭借数据独立性、数据可共享等特点，成为现代社会数据管理的主要方式，在安全性、易使用、可靠性等方面都取得了令人瞩目的成绩。
6. 数据管理技术不断发生着变化，目前借助云计算、大数据来管理大量半结构化、非结构化数据。
7. 威胁数据安全的因素有很多，如硬盘驱动器损坏、操作失误、黑客入侵、感染计算机病毒、遭受自然灾害等。
8. 目前主要是采用 主动防护 的手段保护政府部门或企业的数据安全，如通过磁盘阵列、数据备份、异地容灾等手段，来保证数据的安全。
9. 对于数据安全，不仅要做好防护上的安全，还应提高数据本身的安全，如通过 数据加密 数据校验 等方法，可以提高数据的保密性和完整性。
10. 所谓数据加密，是指通过加密算法和加密密钥将明文转变为密文，而解密则是通过解密算法和解密密钥将密文恢复为明文。
11. 所谓数据校验是为保证数据的完整性进行的一种验证操作，通常用一种指定的算法对原始数据计算一个校验值，接收方用同样的算法计算一个校验值，如果两次计算得到的检验值相同，则说明数据是完整的。常见的 数据校验 方法有 MD5、CRC、SHA-1 等。

#### 二、数据与大数据

1. 符合大数据概念的数据一般具有 数据规模大、处理速度快、数据类型多、价值密度低 四个特征，可以用 4 个 V 来概括，即数量(Volume)、速度(Velocity)、多样(Variety)和价值(Value)。
  - ① 数据体量巨大。大数据的数据量庞大，其存储单位需要用到 EB、ZB 级别，且时刻在增长中。
  - ② 速度快。速度快不仅是指数据产生的速度快，还指数据处理速度快。
  - ③ 数据类型多。大数据来源多样，其数据往往是多种形式共存。

④价值密度低。尽管大数据有很高的价值，但与其海量的数据量相比，则显得密度较低。

5.大数据是一场变革，改变的不仅是数据，还有人们的思维。

①大数据采用**全体数据**分析，而不是抽样数据。

②大数据不再追求个别数据的精确性，允许**数据的混杂性**

③不一定强调事物之间的因果关系，更加注重它们的**相关性**。

6.大数据对社会的影响

①大数据已经渗透到社会各个层面，不断提高工作效率与生活便利性。

②大数据让决策更精准。

③大数据带来了新的就业机会。

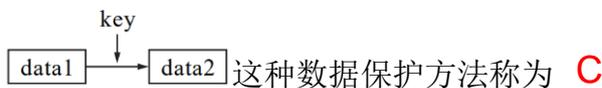
④大数据带来信息泄露、数据安全、个人隐私甚至伦理道德等方面的社会问题。

### 【课堂练习】

1. 下列关于扩展名和文件类型的组合，不正确的是 **D** **png是图像文件**

A. 网页文件 \*.htm B. 文本文件 \*.txt C. 音频文件 \*.wav D. 视频文件 \*.png

2. 为了保护数据的安全性，对数据做如下处理：



A. 异地容灾 B. 数据校验 C. 数据加密 D. 数据备份

3. 某市所有医疗机构联网使用的医疗系统，在管理全市所有病患信息、医疗记录等数据时，应选择 **D**

A. 人工管理 B. 文件管理 C. 表格管理 D. 数据库管理

4. 数据在传输过程中容易受到干扰导致接收的数据不完整，针对这一问题的有效手段是 **C**

A. 数据备份 B. 异地容灾 C. CRC 校验 D. 保护存储介质

5. 下列有关数据管理与安全说法不正确的是 **A** **数据校验保证数据的完整性**

A. 安装磁盘阵列系统是常见的数据加密手段 **磁盘阵列是主动防护**

B. CRC 是数据通信领域的一种查错校验算法

C. 借助云计算与大数据等技术，能提高数据管理水平

D. 为防止系统停止工作时造成数据丢失，可通过容灾系统来解决

6. 下列关于数据安全的说法，不正确的是 **D**

A. 自然灾害也有可能对数据造成损坏 B. 数据加密有助于提高数据的安全 **还要提高数据本身的安全**

C. 数据校验主要是为了保证数据的完整性 D. 数据安全只要做好防护上的安全即可

7. 为了保护数据的安全，下列措施不合理的是 **D**

A. 定期升级系统、安装系统补丁 B. 采用备份、镜像等技术

C. 对数据进行加密存储

D. 为降低系统负荷，关闭防火墙和杀毒软件

8. 下列选项中，不属于大数据的是 **A**

A. 某校学生高考报名数据

B. 气象卫星采集的数据

C. 电商平台用户浏览、交易时产生的数据

D. 交通摄像头记录的数据

9. 小明了解到某省气象大数据云平台可以存储地面、高空、海洋等气象数据，这些数据中有二维表格数据、文本数据、图片数据、地理位置数据等。这最能体现大数据特征中的 **A**

A. 数据类型多

B. 价值密度低

C. 处理速度快

D. 数据规模大

10. 手机导航能**实时**提示前方路况，主要体现了大数据的特征是 **B**

A. 数据体量巨大

B. 处理速度快

C. 数据类型多

D. 价值密度低

11. 大数据环境下的隐私担忧，主要表现为 **C**

A. 恶意广告的推送

B. 病毒入侵

C. 个人信息的暴露

D. 用户画像的生成

## 【课后作业】

- 为减少存储容量，小强想把音频文件“二泉映月.wav”转换为MP3格式文件，小强的方法是将“二泉映月.wav”文件重命名为“二泉映月.mp3”。关于此操作后的文件的文件格式和存储容量说法正确的是 **D**
  - 文件格式改变，存储容量没变
  - 文件格式没变，存储容量变小
  - 文件格式改变，存储容量变小
  - 文件格式和存储容量都没变
- 2014年，铁路售票网站12306发生个人数据泄露事件，犯罪嫌疑人通过收集某游戏网站以及其他多个网站泄露的用户名与密码信息，尝试登录12306网站进行“撞库”，非法获取用户信息，谋取非法利益。结合该事例，为保护个人信息安全，下列行为最合理的是 **B**
  - 用户要提高密码强度
  - 不同网站采用不同的密码
  - 每天改变一次密码
  - 在计算机中安装杀毒软件
- 对于上述描述，下列说法不正确的是 **B**
  - 你的每次上网行为及相关数据可能会被采集
  - 你在上网时的每次操作，蕴含着巨大的价值
  - 你在聊天、购物、看新闻和短视频等过程中，有可能泄露个人隐私
  - 网站提供给你的“个性化推荐”，依赖于你的上网行为
- 有下列数据文件：①wuzhen.bmp ②wuzhen.gif ③wuzhen.wav ④wuzhen.mpg ⑤wuzhen.jpg ⑥wuzhen.doc。一般属于图像文件的有 **C**
  - ①②④
  - ①④⑤
  - ①②⑤
  - ①②④⑤
- 为了预防自然灾害引起的数据损坏，一般可采用的措施有 **D**
  - 安装磁盘阵列系统
  - 对数据进行加密
  - 安装防火墙
  - 建设异地容灾系统
- 下列有关大数据的说法，正确的是 **A**
  - 大数据采集的数据有结构化、半结构化和非结构化的
  - 大数据分析的是抽样数据
  - 用大数据进行数据处理时，必须保证每个数据都准确无误
  - 在大数据时代，数据的主要来源还是人力
- 对于数据安全防范措施，下列做法正确的是 **B**
  - 在公用计算机上使用个人账号时，密码选项中选择保存密码
  - 对于陌生用户发送的电子邮件，不轻易打开邮件中的附件
  - 安装防木马和杀毒软件后就可高枕无忧，不再升级更新
  - 在网上聊天时向对方透露你的账号密码
- 下列关于数据安全的说法，不正确的是 **A**
  - 自然灾害不会导致数据安全问题
  - 威胁数据安全的因素有操作失误、黑客入侵、感染计算机病毒等
  - 通过数据加密和数据校验等方法，可以提高数据的保密性和完整性
  - 保护数据的安全也需要保护存储数据的介质
- 下列选项中，哪个不是大数据的典型应用 **D**
  - 基于交易大数据分析用户的购买习惯
  - 基于搜索引擎的搜索关键词分析社会热点
  - 基于道路摄像头、地感线圈等数据分析城市交通情况
  - 基于科技文献数据库检索某一领域研究进展
- 大数据已经进入了人们的生活，关于大数据的说法，不正确的是 **C**
  - 大数据是未来信息技术产业的一种原料，可以从中提取很多有效信息
  - 在很多情况下，大数据的来源可以不依靠人工
  - 大数据信息量很大，蕴藏的信息价值密度高
  - 在大数据中，个别极端数据不会影响大数据分析结果

11. 下列关于大数据和数据安全的描述, 正确的是 **C**
- A. 数据安全不包括存储数据介质的安全
  - B. 数据加密是为了数据的完整性而进行的操作
  - C. 大数据蕴含巨大价值, 但价值密度相对较低
  - D. 大数据因数据体量大, 只能抽样处理
12. 上海浦东机场采集了过去两年中每个时刻的旅客到达量和陆路交通情况的数据, 开发出一个智能出租车调度系统, 从而使旅客在机场等候出租车的时间大幅缩短。下列关于该事例的说法, 正确的是 **A**
- A. 人们在机场候车的行为产生了数据
  - B. 机场主要是通过人工方式采集数据的 **流数据是实时数据, 过去的数据是静态数据**
  - C. 机场在过去两年采集到的数据属于流数据 (“流数据”属于第四章 4.2.1 中内容)
  - D. 大数据处理只需要分析最近一个月的抽样数据
13. 北斗导航一般指北斗卫星导航系统, 是中国自行研制的全球卫星导航系统。随着“北斗”全球系统的发展, 接入用户的业务类型将不断增加。用户数据可涵盖文字、语音、视频等多种形式, 这体现了大数据的特征是 **D**
- A. 数据体量巨大
  - B. 数据产生和处理的速度快
  - C. 价值密度低
  - D. 数据类型多
14. 在信息社会中, 数据安全关乎国计民生, 为了数据安全需要采用各种防护手段, 下列技术手段不属于存储介质的主动防护的是 **A**
- A. 数据校对
  - B. 磁盘阵列
  - C. 数据备份
  - D. 异地容灾
15. 下列关于数据管理与安全的说法正确的是 **D**
- A. 计算机数据的管理先后经历了文件管理、人工管理、数据库管理
  - B. 安装正版的杀毒软件并及时更新就一定能保证计算机的数据安全
  - C. 数据校验提高了数据的**保密性** **完整性**
  - D. 常用的数据校验方法有 MD5、CRC、SHA-1 等
16. 某校学生在某学习平台进行学习, 所以该平台保存了大量的学生日常学习数据, 从数据安全方面考虑, 下列做法不恰当的是 **A**
- A. 将学生观看教学微课偏好等数据共享给其他软件公司 **侵犯隐私**
  - B. 发布时对数据中的个人信息作匿名处理
  - C. 对数据定期备份
  - D. 采集数据时对用户履行告知义务
17. 下列关于数据管理与安全的说法, 不正确的是 **A**
- A. 常见的数据**加密**方法有 MD5、CRC、SHA-1 等
  - B. 对于数据安全, 不仅要做好防护上的安全, 还应提高数据本身的安全
  - C. 需要多种字符组合且符合一定长度的密码称为强密码, 强密码安全性高, 但记忆较困难
  - D. 结构化数据是由二维表结构来进行逻辑表达和实现的数据
18. 下列不能体现大数据对社会的影响的是 **A**
- A. 扫描餐厅二维码进行自助点餐
  - B. 数据分析师等职业需求急剧增长
  - C. 通过分析车辆导航数据, 预测不同时间不同道路的车流量, 指导用户灵活调整出行路线
  - D. 通过案件数据分析和趋势预测系统, 及时调整警力布控, 变被动为主动巡逻防控
19. 在大数据时代, 各行各业都可以利用大数据创造出更多的价值。下列选项中, 没有用到大数据技术的是 **B**
- A. 节假日时期, 用户利用“城市热力图”提供的信息, 避开人流密度高的地区
  - B. 在超市、银行等场所, 用手持测温仪测量入场人员的体温数据
  - C. 某平台以网民行为数据为基础, 研究关键词搜索趋势, 洞察网民消费特征
  - D. 某网站利用可视化“迁徙”地图, 分析各城市人口流动轨迹与特征