

# 大数据职称考试知识大纲

## 数据分析与应用-初级

2021 年版

# 目 录

|                            |   |
|----------------------------|---|
| 一、大数据法律法规、相关标准及职业道德 .....  | 3 |
| 二、计算机基础知识 .....            | 3 |
| 三、信息化基础知识 .....            | 5 |
| 四、大数据安全 .....              | 6 |
| 五、数据库系统(SQL + NoSQL) ..... | 7 |
| 六、数据仓库 .....               | 7 |
| 七、大数据技术与应用 .....           | 8 |
| 八、大数据分析模型 .....            | 8 |
| 九、数据科学 .....               | 8 |

# 第一部分 基础知识模块

## 一、大数据法律法规、相关标准及职业道德

### （一）法律法规

1、《中华人民共和国民法典》中华人民共和国主席令第四十五号

（1）了解隐私权和个人信息保护有关内容。

2、《中华人民共和国数据安全法》中华人民共和国主席令第八十四号

（1）了解数据安全制度、数据安全保护义务、政务数据安全与开放等相关内容。

3、《中华人民共和国网络安全法》中华人民共和国主席令第五十三号

（1）了解数据分类、重要数据备份和加密等措施。

4、《中华人民共和国电子签名法（2019版）》中华人民共和国主席令第二十九号

（1）了解电子签名、数据电文基本概念。

5、《中华人民共和国密码法》中华人民共和国主席令第三十五号

（1）了解核心密码、普通密码、商用密码之间的区别。

### （二）相关标准

1、GB/T 35295-2017 信息技术 大数据 术语

（1）了解本标准的范围、术语和定义，以及密切相关的通用术语。

2、GB/T 35589-2017 信息技术 大数据 技术参考模型

（1）了解本标准的范围、术语和定义，了解大数据参考架构的目的、目标，了解大数据参考架构及其各组成部分。

3、GB/T 37973-2019 信息安全技术 大数据安全管理指南

（1）了解本标准的范围、术语和定义，了解大数据安全需求、大数据分类分级、大数据活动及安全要求、大数据安全风险评估等方面的内容。

### （三）职业道德

掌握专业技术人员职业道德的基本要求

## 二、计算机基础知识

## （一）面向对象技术

- 1、了解面向对象的发展历史，为什么引入面向对象
- 2、掌握面向对象编程与非面向对象编程的优缺点
- 3、掌握抽象的 5 个层次，抽象形式
- 4、了解类和方法
- 5、了解消息、实例和初始化

## （二）数据结构与算法

- 1、掌握程序性能分析的概念和方法，包括时间复杂性与空间复杂性分析。
- 2、掌握线性表的概念，掌握堆栈、队列、跳表和散列的描述方法与应用。
- 3、了解树的描述方法与应用。
- 4、了解图的描述方法与应用。

## （三）操作系统

- 1、掌握操作系统的概念和操作系统结构。
- 2、掌握操作系统的进程管理，包括进程概念、进程调度、同步及死锁处理。
- 3、了解内存管理，包括内存管理策略和虚拟内存管理。
- 4、了解存储管理，包括文件系统、文件系统实现、大容量存储结构和 I/O 系统。
- 5、了解系统保护与安全。

## （四）计算机网络

- 1、了解计算机网络在信息时代的核心作用以及计算机网络的发展历史；理解计算机网络的分类，网络标准化，网络参考模型，网络体系结构。
- 2、掌握物理层的基本概念，理解信道极限容量的概念以及信道最大传输速率的公式，模拟传输和数字化传输的物理层标准。
- 3、掌握数据链路层的基本概念，理解停等协议和连续 ARQ 协议，滑动窗口协议，检错和纠错机制。
- 4、掌握介质访问控制子层的基本概念，理解动态多路访问控制协议，以太网，无线局域网，数据链路层的交换技术。
- 5、了解网络层的基本概念，了解路由协议，拥塞控制算法，服务质量，网

络互连，IP 协议，子网掩码。

6、了解传输层的基本概念，了解传输层路由协议，TCP，UDP，拥塞控制算法。

7、了解应用层的基本概念，了解 DNS，邮件系统，www，流音频与视频。

### 三、信息化基础知识

#### （一）电子政务

1、电子政务的概念、内容和技术形式

（1）了解电子政务的概念

（2）了解电子政务的内容

（3）了解电子政务的技术形式

2、中国政府信息化的策略和历程

3、电子政务应用模式

（1）了解政府对政府（G2G）模式

（2）了解政府对企业（G2B）模式

（3）了解政府对公众（G2C）模式

（4）了解政府对公务员（G2E）模式

#### （二）电子商务

1、电子商务的概念及基本特征

（1）了解电子商务的概念

（2）了解电子商务的基本特征

2、电子商务的类型

（1）了解企业与企业之间的电子商务（B2B）模式。

（2）了解商业企业与消费者之间的电子商务（B2C）模式。

（3）了解消费者与消费者之间的电子商务（C2C）模式。

（4）了解电子商务与线下实体店有机结合向消费者提供商品和服务（O2O）模式。

#### （三）新一代信息技术

1、物联网

- (1) 了解物联网的概念
- (2) 了解物联网架构
- (3) 了解物联网技术的应用

## 2、云计算

- (1) 了解云计算的概念
- (2) 了解云计算的架构及关键技术
- (3) 了解云计算服务的类型
- (4) 了解发展云计算的主要任务

## 3、移动互联网

- (1) 了解移动互联网的概念
- (2) 了解移动互联网的特点
- (3) 了解移动互联网技术的应用

## 4、人工智能

- (1) 了解人工智能的概念
- (2) 了解人工智能研究范畴
- (3) 了解人工智能实际应用

## 5、区块链

- (1) 了解区块链的概念
- (2) 了解区块链的系统架构
- (3) 了解区块链的分类
- (4) 了解区块链的应用

## 6、量子技术

- (1) 了解量子技术的概念
- (2) 了解量子技术的特点
- (3) 了解量子技术的应用

# 四、大数据安全

## (一) 网络安全

- 1、熟悉网络安全的一些基本概念、基本属性及网络安全模型等。
- 2、了解网络安全体系的安全服务、安全机制、安全服务配置和安全管理等。

3、熟悉各种网络安全技术的概念，了解物理安全、网络安全设备、系统安全、应用安全的基本知识。

4、了解网络安全运维与保障的框架、基本原理等知识。

## **(二) 数据安全**

1、了解大数据安全管理及安全需求的原则，掌握数据安全分类、定级的原则。

2、了解数据采集、存储、处理、分发、删除等活动的概念，了解相关安全技术与安全要求。

3、熟悉云存储安全体系、数据生命周期中的安全风险、保障云存储安全的原则及云计算边界安全相关的概念。

4、了解大数据安全与防护保障机制。

## **第二部分：专业知识**

### **五、数据库系统(SQL + NoSQL)**

1、掌握关系模型概念和 SQL 语言。

2、掌握关系数据库设计方法。

3、了解关系数据库事务概念和事务调度方法。

4、了解数据库并发控制技术。

5、了解非关系数据库背景、特点和分类。

### **六、数据仓库**

1、了解数据挖掘的定义、功能、常用方法。

2、了解数据仓库的产生与发展，掌握数据仓库的定义。

3、了解数据仓库与数据挖掘的联系与区别。

4、了解常用数据挖掘工具。

5、了解联机分析处理技术（OLAP）的概念、特征。

6、了解 OLAP 中的多维分析操作，包括钻取、切片和切块、旋转。

7、了解数据预处理的原因、数据预处理的方法。

8、了解数据清洗、数据集成和变换、数据归约的概念与方法，具有应用上述方法进行数据清洗的能力。

## 七、大数据技术与应用

### （一）数据可视化

1、了解常见可视化图形（散点图、折线图、饼图、环图、柱状图）的概念，具有初步的可视化图形展示数据的能力。

### （二）大数据处理技术

1、掌握云计算的概念、云计算的特征、云计算的三层 SPI 模型（IaaS、PaaS、SaaS）。

2、了解虚拟化技术的概念、常用虚拟化方法，了解常用虚拟化软件。

3、具有应用虚拟化技术搭建虚拟化平台的能力。

4、了解分布式计算、高性能计算、边缘计算的概念。

5、了解数据集市、数据仓库、数据中台、数据平台的概念。

6、了解数据采集与预处理过程，掌握数据采集、数据清洗、数据转换、数据脱敏的方法。

7、了解大数据处理环节的主要思想：大数据采集、大数据预处理、大数据存储及管理、大数据分析 & 挖掘、大数据展现和应用（包括大数据检索、大数据可视化、大数据应用、大数据安全等）。

8、了解大数据处理与分析的代表性流行产品。

## 八、大数据分析模型

1、了解常用的大数据分析模型，例如行为事件分析模型、点击分析模型、用户行为路径分析模型、用户分群分析模型等，具有应用大数据技术参与大数据分析的能力。

## 九、数据科学

### （一）概率统计与统计推断

1、掌握概率、条件概率的概念，并会计算简单的概率、条件概率。

2、理解随机事件的独立性和随机变量的独立性。

3、掌握乘法公式、全概率公式，并会简单应用。

4、掌握随机变量分布函数的概念，掌握连续型随机变量的密度函数和离散

型随机变量的分布列。

5、掌握常见分布（两点分布、二项分布、泊松分布、均匀分布、正态分布、指数分布）及其简单性质。

6、理解数学期望、方差（标准差）、相关系数的概念，并会简单计算。

7、理解统计量的概念，理解样本均值、样本方差（标准差）的概念，并会简单计算。

8、了解点估计的概念。

9、了解线性回归的基本概念。

## （二）统计学习与数据挖掘

1、掌握感知机的定义和基本原理。

2、掌握 Logistic 回归算法原理与特点，能够使用 Logistic 回归进行数据的分类建模与参数解释。

3、掌握朴素贝叶斯算法的定义与基本原理。

4、掌握 k 近邻算法算法定义与原理。

5、掌握支持向量机（SVM）算法的思想与原理。

6、掌握决策树的算法的定义与原理，了解决策树的剪枝理论。

7、了解常见集成方法如 boosting、bagging 等。

8、掌握聚类分析的相关概念。

9、掌握主成分分析的原理。

10、了解生成式半监督学习方法、半监督 SVM、图半监督学习方法的思想与原理。

## （三）深度学习与强化学习

1、掌握全连接网络（MLP）的定义与算法原理、并能够应用。

2、掌握卷积神经网络（CNN）基本概念，了解几种经典的卷积神经网络，如 AlexNet、VGG、GoogLeNet、ResNet。

3、掌握循环神经网络（RNN）原理与基本概念，了解几种常见的模型，如长短期记忆网络 LSTM、GRU。

4、掌握生成对抗网络（GAN、WGAN）的基本定义与原理，能够完成基本应用

5、掌握马尔可夫决策过程的定义，了解其算法应用范围，并能够实际应用

6、掌握经典强化学习算法（Q-learning、SARMA）的定义与原理。

7、掌握经典深度强化学习算法（DQN、DDPG、A3C、TRPO、PPO）的定义与原理，了解其算法特性。

# 大数据职称考试知识大纲

## 数据分析与应用-中级

2021 年版

# 目 录

|                              |    |
|------------------------------|----|
| 一、 大数据法律法规、政策文件及相关标准 .....   | 1  |
| 二、 计算机基础知识 .....             | 2  |
| 三、 信息化基础知识 .....             | 3  |
| 四、 大数据安全 .....               | 7  |
| 五、 数据库系统 (SQL + NoSQL) ..... | 8  |
| 六、 数据仓库 .....                | 8  |
| 七、 大数据技术与应用 .....            | 9  |
| 八、 大数据分析模型 .....             | 9  |
| 九、 数据科学 .....                | 10 |

# 第一部分 基础知识模块

## 一、大数据法律法规、政策文件及相关标准

### （一）法律法规

1、《中华人民共和国民法典》中华人民共和国主席令第四十五号

（1）了解隐私权和个人信息保护有关内容。

2、《中华人民共和国数据安全法》中华人民共和国主席令第八十四号

（1）了解数据安全制度、数据安全保护义务、政务数据安全与开放等相关内容。

3、《中华人民共和国网络安全法》中华人民共和国主席令第五十三号

（1）熟悉数据分类、重要数据备份和加密等措施。

4、《中华人民共和国电子签名法（2019 版）》中华人民共和国主席令第二十九号

（1）了解电子签名、数据电文基本概念。

5、《中华人民共和国密码法》中华人民共和国主席令第三十五号

（1）了解核心密码、普通密码、商用密码之间的区别。

### （二）政策文件

1、《促进大数据发展行动纲要》国发〔2015〕50 号

（1）熟悉数据共享的范围边界和使用方式。

2、《山东省“十四五”数字强省建设规划》

（1）了解数字强省“十四五”规划发展目标及主要任务。

3、《关于促进大数据发展的意见》鲁政发〔2016〕25 号

（1）了解山东省促进大数据发展的总体目标以及大数据应用生态体系。

### （三）相关标准

1、GB/T 35295-2017 信息技术 大数据 术语

（1）了解本标准的范围、术语和定义，以及密切相关的通用术语。

2、GB/T 35589-2017 信息技术 大数据 技术参考模型

（1）了解本标准的范围、术语和定义，熟悉大数据参考架构的目的、目标，熟悉大数据参考架构及其各组成部分。

### 3、GB/T 37973-2019 信息安全技术 大数据安全管理指南

(1) 了解本标准的范围、术语和定义，了解大数据安全需求、大数据分类分级、大数据活动及安全要求、大数据安全风险评估等方面的内容。

### 4、GB/T 38667-2020 信息技术 大数据 数据分类指南

(1) 了解本标准的范围、术语和定义，熟悉数据分类的过程，以及分类视角、分类维度和分类方法。

### 5、GB/T 38673-2020 信息技术 大数据 大数据系统基本要求

(1) 了解本标准的范围、术语和定义，熟悉大数据系统框架、功能要求、非功能要求。

## (四) 职业道德

1、掌握专业技术人员职业道德的基本要求

## 二、计算机基础知识

### (一) 面向对象技术

- 1、了解面向对象的发展历史，为什么引入面向对象。
- 2、掌握面向对象编程与非面向对象编程的优缺点。
- 3、掌握抽象的 5 个层次，抽象形式。
- 4、掌握类和方法。
- 5、掌握消息、实例和初始化。
- 6、了解继承和替换。
- 7、了解静态行为和动态行为。

### (二) 数据结构与算法

- 1、掌握程序性能分析的概念和方法，包括时间复杂性与空间复杂性分析。
- 2、掌握线性表的概念，掌握堆栈、队列、跳表和散列的描述方法与应用。
- 3、了解树的描述方法与应用。
- 4、了解图的描述方法与应用。

### (三) 操作系统

- 1、掌握操作系统的概念和操作系统结构。
- 2、掌握操作系统的进程管理，包括进程概念、进程调度、同步及死锁处理。

- 3、掌握内存管理，包括内存管理策略和虚拟内存管理。
- 4、了解存储管理，包括文件系统、文件系统实现、大容量存储结构和 I/O 系统。
- 5、了解系统保护与安全。

#### **（四）计算机网络**

- 1、了解计算机网络在信息时代的核心作用以及计算机网络的发展历史；理解计算机网络的分类，网络标准化，网络参考模型，网络体系结构。
- 2、掌握物理层的基本概念，理解信道极限容量的概念以及信道最大传输速率的公式，模拟传输和数字化传输的物理层标准。
- 3、掌握数据链路层的基本概念，理解停等协议和连续 ARQ 协议，滑动窗口协议，检错和纠错机制。
- 4、掌握介质访问控制子层的基本概念，理解动态多路访问控制协议，以太网，无线局域网，数据链路层的交换技术。
- 5、了解网络层的基本概念，理解路由协议，拥塞控制算法，服务质量，网络互连，IP 协议，子网掩码。
- 6、了解传输层的基本概念，理解传输层路由协议，TCP，UDP，拥塞控制算法。
- 7、了解应用层的基本概念，理解 DNS，邮件系统，www，流音频与视频。

### **三、信息化基础知识**

#### **（一）电子政务**

- 1、电子政务的概念、内容和技术形式
  - （1）掌握电子政务的概念
  - （2）熟悉电子政务的内容
  - （3）熟悉电子政务的技术形式
- 2、中国政府信息化的策略和历程
- 3、电子政务应用模式
  - （1）熟悉政府对政府（G2G）模式
  - （2）熟悉政府对企业（G2B）模式

- (3) 熟悉政府对公众 (G2C) 模式
- (4) 熟悉政府对公务员 (G2E) 模式

## （二）电子商务

### 1、电子商务的概念及基本特征

- （1）熟悉电子商务的概念
- （2）熟悉电子商务的基本特征

### 2、电子商务的类型

- （1）熟悉企业与企业之间的电子商务（B2B）模式
- （2）熟悉商业企业与消费者之间的电子商务（B2C）模式
- （3）熟悉消费者与消费者之间的电子商务（C2C）模式
- （4）熟悉电子商务与线下实体店有机结合向消费者提供商品和服务（O2O）

模式

## （三）新一代信息技术

### 1、物联网

- （1）掌握物联网的概念
- （2）熟悉物联网架构
- （3）了解物联网技术的应用

### 2、云计算

- （1）掌握云计算的概念
- （2）熟悉云计算的架构及关键技术
- （3）了解云计算服务的类型
- （4）了解发展云计算的主要任务

### 3、移动互联网

- （1）掌握移动互联网的概念
- （2）熟悉移动互联网的特点
- （3）了解移动互联网技术的应用

### 4、人工智能

- （1）掌握人工智能的概念
- （2）熟悉人工智能研究范畴
- （3）了解人工智能实际应用

### 5、区块链

- (1) 掌握区块链的概念
- (2) 熟悉区块链的系统架构
- (3) 了解区块链的分类
- (4) 了解区块链的应用

## 6、量子技术

- (1) 掌握量子技术的概念
- (2) 熟悉量子技术的特点
- (3) 了解量子技术的应用

## (四) 项目管理

### 1、项目管理的理论与体系

- (1) 了解项目管理基础知识
- (2) 掌握项目管理知识体系的构成
- (3) 了解 IPMP/PMP、PRINCE2 等项目管理认证体系
- (4) 掌握项目管理成熟度模型
- (5) 了解项目管理的量化方法

### 2、组织结构对项目的影响

- (1) 了解组织结构对项目的影响。

### 3、信息系统项目典型生命周期模型

- (1) 掌握瀑布模型及其瀑布模型基础上改进的模型
- (2) 了解原型化模型
- (3) 了解敏捷开发模型
- (4) 了解软件生命周期其它模型

### 4、单个项目的管理过程

- (1) 掌握项目整体管理的含义、作用
- (2) 了解项目整体管理过程
- (3) 掌握项目进度管理
- (4) 了解项目成本管理
- (5) 了解项目质量管理
- (6) 了解项目人力资源管理

- (7) 了解项目沟通管理和干系人管理
- (8) 了解项目风险管理
- (9) 了解项目单个项目管理的其它方法

#### 5、信息系统安全管理

- (1) 掌握信息系统安全策略
- (2) 了解信息安全系统工程
- (3) 了解 PKI 公开密钥基础设施
- (4) 了解 PMI 权限（授权）管理基础
- (5) 了解信息安全审计
- (6) 了解信息系统安全管理其它方法

## 四、大数据安全

### （一）网络安全

1、熟悉网络安全的一些基本概念、基本属性及网络安全模型等；了解网络安全评估；掌握各种类型的网络安全服务。

2、了解网络安全体系的安全服务、安全机制、安全服务配置和安全管理等；掌握 TCP/IP 协议安全体系。

3、熟悉各种网络安全技术的概念，了解物理安全、网络安全设备、系统安全、应用安全的技术知识。

4、了解网络安全运维与保障的框架、基本原理等知识，掌握安全管理的技术、设施、工具和方法。

### （二）数据安全

1、了解大数据安全管理及安全需求的原则，掌握数据安全分类、定级的原则和方法。

2、了解数据采集、存储、处理、分发、删除等活动的概念，了解相关安全技术与安全要求。

3、熟悉云存储安全体系、数据生命周期中的安全风险、保障云存储安全的原则及云计算边界安全相关的概念及云计算安全架构。

4、了解大数据安全与防护保障机制、大数据安全审计、大数据安全评估与

安全管理系统。

## 第二部分：专业知识

### 五、数据库系统(SQL + NoSQL)

- 1、掌握关系模型概念和 SQL 语言。
- 2、掌握关系数据库设计方法。
- 3、掌握关系数据库索引的概念和使用方法。
- 4、掌握关系数据库查询处理与查询优化方法。
- 5、掌握关系数据库事务概念和事务调度方法。
- 6、掌握数据库并发控制技术。
- 7、了解非关系数据库背景、特点和分类。

### 六、数据仓库

- 1、了解数据挖掘的定义、功能、常用方法。
- 2、了解数据仓库的产生与发展，掌握数据仓库的定义。
- 3、了解数据仓库与数据挖掘的联系与区别。
- 4、了解常用数据挖掘工具。
- 5、了解数据仓库的体系结构，掌握元数据、粒度、分割的概念，了解数据仓库中的数据组织形式。
- 6、了解数据仓库的数据模型，包括概念数据模型、逻辑数据模型、物理数据模型。
- 7、掌握联机分析处理（OLAP）技术的概念、特征。
- 8、掌握 OLAP 中的多维分析操作，包括钻取、切片和切块、旋转。
- 9、了解多维联机分析处理、关系联机分析处理、MOLAP 和 ROLAP 的差异、混合型联机分析处理；OLAP 的衡量标准。
- 10、了解数据预处理的原因、数据预处理的方法。
- 11、掌握数据清洗、数据集成和变换、数据归约的概念与方法，具有应用上述方法进行数据清洗的能力。
- 12、了解关联规则的概念和分类。

## 七、大数据技术与应用

### （一）数据可视化

1、了解常见可视化图形（散点图、折线图、饼图、环图、双层环图、柱状图、堆积柱状图、分组柱状图、正负条形图、雷达图）的概念，具有初步的可视化图形展示数据的能力。

### （二）大数据处理技术

1、掌握云计算的概念、云计算的特征、云计算的三层 SPI 模型（IaaS、PaaS、SaaS）。

2、掌握虚拟化技术的概念、常用虚拟化方法，了解常用虚拟化软件。

3、具有应用虚拟化技术搭建虚拟化平台的能力。

4、了解分布式计算、高性能计算、边缘计算的概念。

5、了解有监督学习、无监督学习、强化学习的概念和区别。

6、了解数据集市、数据仓库、数据中台、数据平台的概念。

7、掌握数据采集与预处理过程，掌握数据采集、数据清洗、数据转换、数据脱敏的方法。

8、了解数据存储与管理技术、大数据处理架构 Hadoop、分布式文件系统 HDFS、NoSQL 数据库、分布式数据库 HBase 的相关概念。

9、了解大数据处理环节的主要思想：大数据采集、大数据预处理、大数据存储及管理、大数据分析及挖掘、大数据展现和应用（包括大数据检索、大数据可视化、大数据应用、大数据安全等）。

10、了解大数据处理与分析的代表性流行产品。

## 八、大数据分析模型

1、了解常用的大数据分析模型，例如行为事件分析模型、点击分析模型、用户行为路径分析模型、用户分群分析模型等，具有应用大数据技术进行大数据分析能力。

## 九、数据科学

### (一) 概率统计与统计推断

- 1、掌握概率、条件概率的概念，并会计算简单的概率、条件概率。
- 2、理解随机事件的独立性和随机变量的独立性。
- 3、掌握乘法公式、全概率公式、贝叶斯公式，并会简单应用。
- 4、掌握随机变量分布函数的概念，掌握连续型随机变量的密度函数和离散型随机变量的分布列。
- 5、掌握常见分布（两点分布、二项分布、泊松分布、均匀分布、正态分布、指数分布）及其简单性质。
- 6、理解数学期望、方差（标准差）、相关系数的概念，并会简单计算。
- 7、了解多元随机变量的概念，掌握多元正态分布。
- 8、理解统计量的概念，理解样本均值、样本方差（标准差）的概念，并会简单计算。
- 9、了解点估计的概念，了解点估计的无偏性、相合性。
- 10、了解参数的矩估计、最大似然估计。
- 11、了解线性回归的基本概念。
- 12、了解方差分析（ANOVA）的基本概念。

### (二) 统计学习与数据挖掘

- 1、掌握感知机的定义和基本原理。
- 2、掌握 Logistic 回归算法原理与特点，能够使用 Logistic 回归进行数据的分类建模与参数解释。
- 3、掌握朴素贝叶斯算法的定义与基本原理。
- 4、掌握 k 近邻算法算法定义与原理。
- 5、掌握支持向量机（SVM）算法的思想与原理。
- 6、掌握决策树的算法的定义与原理，了解决策树的剪枝理论。
- 7、了解常见集成方法如 boosting、bagging 等。
- 8、掌握聚类分析的相关概念。
- 9、掌握主成分分析的原理。

10、了解生成式半监督学习方法、半监督 SVM、图半监督学习方法的思想与原理。

### **(三) 深度学习与强化学习**

1、掌握全连接网络 (MLP) 的定义与算法原理、并能够应用。

2、掌握卷积神经网络 (CNN) 基本概念，了解几种经典的卷积神经网络，如 AlexNet、VGG、GoogLeNet、ResNet。

3、掌握循环神经网络 (RNN) 原理与基本概念，了解几种常见的模型，如长短期记忆网络 LSTM、GRU。

4、掌握生成对抗网络 (GAN、WGAN) 的基本定义与原理，能够完成基本应用

5、掌握马尔可夫决策过程的定义，了解其算法应用范围，并能够实际应用

6、掌握经典强化学习算法 (Q-learning、SARMA) 的定义与原理。

7、掌握经典深度强化学习算法 (DQN、DDPG、A3C、TRPO、PPO) 的定义与原理，了解其算法特性。

# 大数据职称考试知识大纲

## 数据分析与应用-高级

2021 年版

# 目录

|                            |    |
|----------------------------|----|
| 一、大数据法律法规、政策文件及相关标准 .....  | 1  |
| 二、计算机基础知识 .....            | 5  |
| 三、信息化基础知识 .....            | 6  |
| 四、大数据安全 .....              | 9  |
| 五、数据库系统(SQL + NoSQL) ..... | 10 |
| 六、数据仓库 .....               | 10 |
| 七、大数据技术与应用 .....           | 11 |
| 八、大数据分析模型 .....            | 12 |
| 九、数据科学 .....               | 12 |

# 第一部分：基础知识

## 一、大数据法律法规、政策文件及相关标准

### （一）法律法规

#### 1、《中华人民共和国民法典》

- （1）了解个人信息处理包括哪些方面，应遵循哪些原则和条件。
- （2）了解自然人和信息处理者的权利义务。
- （3）熟悉隐私和个人信息的定义及两个概念的关系。

#### 2、《中华人民共和国数据安全法》（中华人民共和国主席令第八十四号）

- （1）了解各级机关、各类组织在数据安全领域的职责，及开展数据处理活动应遵循的规范和原则。
- （2）了解数据安全制度、保护义务、法律责任、政务数据安全与开放等内容。
- （3）掌握数据、数据处理、数据安全等概念定义，及数据安全法适用范围。

#### 3、《中华人民共和国网络安全法》（中华人民共和国主席令第五十三号）

- （1）了解网络安全、网络运营者、网络数据、个人信息概念，及个人、组织和网络运营者在网络安全中的权利或义务。
- （2）了解网络安全支持与促进、网络安全运行、法律责任，及各级机关、各类组织在网络安全领域的职责。
- （3）掌握网络安全法的适用范围、关键基础设施的范围和运行安全，及监测预警与应急处置的相关制度和措施。

#### 4、《中华人民共和国电子签名法（2019 版）》

- （1）了解电子签名、数据电文及其认证证书、验证数据、制作数据等相关概念。
- （2）了解数据电文发送接收时间、地点等概念，及原件形式要求和文件保存要求。
- （3）了解可靠电子签名概念，及电子认证服务提供者的责任义务及其监管部门。

#### 5、《国务院关于在线政务服务的若干规定》（国务院令 第 716 号）

(1) 了解全国一体化在线政务服务平台、统一身份认证服务、政务服务事项、政务服务机构、行政相对人等概念及政务服务机构及其工作人员的责任义务。

(2) 了解国务院办公厅及各级各部门在政务服务平台体系建设中的职责分工。

(3) 掌握电子签名、电子印章、电子证照和电子档案的概念及其法律效力。

#### 6、《政务信息资源共享管理暂行办法》（国发〔2016〕51号）

(1) 了解政务信息资源的概念、分类、共享要求、共享原则及目录编制要求。

(2) 了解国家发改委、网信办等各级政务部门在数据共享、编制政务信息资源目录等工作中的职责分工。

(3) 了解共享信息的提供与使用、信息共享工作的监督保障。

(4) 掌握共享平台概念、分类及所依托网络，掌握信息资源的共享属性分类，及各政务部门共享信息维护更新、使用管理的原则和机制。

#### 7、《中华人民共和国密码法》（中华人民共和国主席令第三十五号）

(1) 了解密码概念分类、密码工作原则，及密码工作相关法律责任。

(2) 了解密码管理部门、密码工作机构的工作职责及相关监管审查制度。

(3) 了解商用密码检测认证体系、标准化体系、进出口许可，及国家密码管理部门、国家商务主管部门、商用密码行业协会在商用密码管理中的职责分工。

(4) 掌握核心密码、普通密码、商用密码的概念、用途。

#### 8、《山东省电子政务和政务数据管理办法》（省政府令第329号）

(1) 了解各级政府及大数据主管部门在政务数据开放网站、政务云平台、政务数据资源共享交换平台、政务网络、统一身份认证、电子印章、安全保障制度机制建设管理方面的职责。

(2) 掌握各级政府、有关部门在政务数据资源管理方面的职责，及政务数据采集、维护、开放、共享的规范和工作机制。

## （二）政策文件

### 1、《促进大数据发展行动纲要》国发〔2015〕50号

(1) 掌握数据共享的范围边界和使用方式。

### 2、《国务院办公厅关于促进电子政务协调发展的指导意见》国办发〔2014〕

66 号

(1) 了解电子政务的发展现状和存在的矛盾问题以及电子政务协调发展的目标和原则。

3、《国务院关于加快推进“互联网+政务服务”工作的指导意见》国发〔2016〕55 号

(1) 熟悉“互联网+政务服务”对激发市场活力和社会创造力的重要意义。

4、《国务院办公厅关于印发进一步深化“互联网+政务服务”推进政务服务“一网、一门、一次”改革实施方案的通知》国办发〔2018〕45 号

(1) 了解推进政务服务“一网、一门、一次”改革实施的基本原则和工作目标。

5、《山东省委办公厅、省政府办公厅印发加快数据融合应用深化流程再造实施方案》。

(1) 熟悉“一个平台、一个号（码）、一张网络、一朵云”所指的内容。

6、《山东省人民政府办公厅关于山东省数字基础设施建设的指导意见》鲁政办字〔2020〕34 号。

(1) 了解新型技术如何持续推动交通、能源、水利、市政等传统基础设施数字化升级。

7、《山东省数字政府建设实施方案（2019-2022）年》鲁政办发〔2019〕8 号

(1) 掌握集约化政务云平台“1+N”框架的含义以及打破信息孤岛，拔掉数据烟囱的基本举措。

8、《山东省“十四五”数字强省建设规划》

(1) 了解数字强省“十四五”规划发展目标及主要任务。

9、《关于促进大数据发展的意见》鲁政发〔2016〕25 号

(1) 了解山东省促进大数据发展的总体目标以及大数据应用生态体系。

10、《山东省人民政府办公厅关于贯彻国办发〔2014〕66 号文件加快推进电子政务协调发展的实施意见》鲁政办发〔2015〕59 号

(1) 熟悉云计算、大数据等新技术在电子政务建设中的应用。

11. 《关于印发山东省支持数字经济发展的意见的通知》鲁政办发〔2019〕

124 号

(1) 了解目标任务、政策措施等相关内容。

12. 《关于加快推进新型智慧城市建设的指导意见》鲁政办发〔2020〕136号

(1) 了解工作目标、总体架构、优化政务服务、拓展便民应用、五、推动精细治理等相关内容。

### (三) 相关标准

1、GB/T 35295-2017 信息技术大数据术语

(1) 熟悉本标准的范围、术语和定义，以及密切相关的通用术语。

2、GB/T 35589-2017 信息技术大数据技术参考模型

(1) 熟悉本标准的范围、术语和定义，熟悉大数据参考架构的目的、目标，熟悉大数据参考架构及其各组成部分。

3、GB/T 37973-2019 信息安全技术大数据安全管理指南

(1) 熟悉本标准的范围、术语和定义，了解大数据安全需求、大数据分类分级、大数据活动及安全要求、大数据安全风险评估等方面的内容。

4、GB/T 38667-2020 信息技术大数据数据分类指南

(1) 熟悉本标准的范围、术语和定义，熟悉数据分类的过程，以及分类视角、分类维度和分类方法。

5、GB/T 38673-2020 信息技术大数据大数据系统基本要求

(1) 熟悉本标准的范围、术语和定义，熟悉大数据系统框架、功能要求、非功能要求。

6、GB/T 38675-2020 信息技术大数据计算系统通用要求

(1) 熟悉本标准的范围、术语和定义，熟悉大数据计算系统的硬件要求、软件要求、软件组成、网络要求和安全要求。

DB37/T 3890.1-2020 新型智慧城市建设指标第1部分：市级指标

了解新型智慧城市建设市级指标

DB37/T 3890.2-2020 新型智慧城市建设指标第2部分：县级指标

了解新型智慧城市建设县级指标

DB37/T 3890.3-2020 新型智慧城市建设指标第3部分：智慧社区指标

了解新型智慧城市建设智慧社区指标

## **(四) 职业道德**

- 1、掌握专业技术人员职业道德的基本要求

## **二、计算机基础知识**

### **(一) 面向对象技术**

- 1、了解面向对象的发展历史，为什么引入面向对象。
- 2、掌握面向对象编程与非面向对象编程的优缺点。
- 3、掌握抽象的 5 个层次，抽象形式。
- 4、掌握类和方法。
- 5、掌握消息、实例和初始化。
- 6、掌握继承和替换。
- 7、掌握静态行为和动态行为。
- 8、掌握替换的本质。
- 9、掌握多重继承。
- 10、掌握多态及软件复用。
- 11、掌握面向对象设计原则和面向对象设计模式。

### **(二) 数据结构与算法**

- 1、掌握程序性能分析的概念和方法，包括时间复杂性与空间复杂性分析。
- 2、掌握线性表的概念，掌握堆栈、队列、跳表和散列的描述方法与应用。
- 3、掌握树的描述方法与应用。
- 4、掌握图的描述方法与应用。

### **(三) 操作系统**

- 1、掌握操作系统的概念和操作系统结构。
- 2、掌握操作系统的进程管理，包括进程概念、进程调度、同步及死锁处理。
- 3、掌握内存管理，包括内存管理策略和虚拟内存管理。
- 4、掌握存储管理，包括文件系统、文件系统实现、大容量存储结构和 I/O 系统。
- 5、掌握系统保护与安全。

## （四）计算机网络

1、了解计算机网络在信息时代的核心作用以及计算机网络的发展历史；理解计算机网络的分类，网络标准化，网络参考模型，网络体系结构。

2、掌握物理层的基本概念，理解信道极限容量的概念以及信道最大传输速率的公式，模拟传输和数字化传输的物理层标准。

3、掌握数据链路层的基本概念，理解停等协议和连续 ARQ 协议，滑动窗口协议，检错和纠错机制。

4、掌握介质访问控制子层的基本概念，理解动态多路访问控制协议，以太网，无线局域网，数据链路层的交换技术。

5、掌握网络层的基本概念，理解路由协议，拥塞控制算法，服务质量，网络互连，IP 协议，子网掩码。

6、掌握传输层的基本概念，理解传输层路由协议，TCP，UDP，拥塞控制算法。

7、掌握应用层的基本概念，理解 DNS，邮件系统，www，流音频与视频。

## 三、信息化基础知识

### （一）电子政务

1、电子政务的概念、内容和技术形式

（1）掌握电子政务的概念

（2）掌握电子政务的内容

（3）熟悉电子政务的技术形式

2、中国政府信息化的策略和历程

3、电子政务应用模式

（1）掌握政府对政府（G2G）

（2）掌握政府对企业（G2B）

（3）掌握政府对公众（G2C）

（4）掌握政府对公务员（G2E）

### （二）电子商务

1、电子商务的概念及基本特征

- (1) 掌握电子商务的概念
- (2) 掌握电子商务的基本特征

## 2、电子商务的类型

- (1) 熟悉企业与企业之间的电子商务（B2B）
- (2) 熟悉商业企业与消费者之间的电子商务（B2C）
- (3) 熟悉消费者与消费者之间的电子商务（C2C）
- (4) 熟悉电子商务与线下实体店有机结合向消费者提供商品和服务（O2O）

## （三）新一代信息技术

### 1、物联网

- (1) 掌握物联网的概念
- (2) 熟悉物联网架构
- (3) 熟悉物联网技术的应用

### 2、云计算

- (1) 掌握云计算的概念
- (2) 掌握云计算的架构及关键技术
- (3) 熟悉云计算服务的类型
- (4) 熟悉发展云计算的主要任务

### 3、移动互联网

- (1) 掌握移动互联网的概念
- (2) 熟悉移动互联网的特点
- (3) 熟悉移动互联网技术的应用

### 4、人工智能

- (1) 掌握人工智能的概念
- (2) 熟悉人工智能研究范畴
- (3) 熟悉人工智能实际应用

### 5、区块链

- (1) 掌握区块链的概念
- (2) 熟悉区块链的系统架构
- (3) 熟悉区块链的分类

(4) 熟悉区块链的应用

## 6、量子技术

(1) 掌握量子技术的概念

(2) 熟悉量子技术的特点

(3) 熟悉量子技术的应用

## (四) 项目管理

### 1、项目管理的理论与体系

(1) 了解项目管理基础知识

(2) 掌握项目管理知识体系的构成

(3) 了解 IPMP/PMP、PRINCE2 等项目管理认证体系

(4) 掌握项目管理成熟度模型

(5) 了解项目管理的量化方法

### 2、组织结构对项目的影响

(1) 了解组织结构对项目的影响

### 3、信息系统项目典型生命周期模型

(1) 掌握瀑布模型及其瀑布模型基础上改进的模型

(2) 掌握原型化模型

(3) 掌握敏捷开发模型

(4) 了解软件生命周期其它模型

### 4、单个项目的管理过程

(1) 掌握项目整体管理的含义、作用

(2) 了解项目整体管理过程

(3) 掌握项目进度管理

(4) 掌握项目成本管理

(5) 掌握项目质量管理

(6) 掌握项目人力资源管理

(7) 掌握项目沟通管理和干系人管理

(8) 了解项目风险管理

(9) 了解项目单个项目管理的其它方法

## 5、项目集（大型项目）管理

- (1) 掌握项目集管理基础
- (2) 了解项目集与战略一致性
- (3) 了解项目集生命周期和收益管理
- (4) 了解项目集管理中的治理、管理支持过程
- (5) 了解项目集管理的其它方法

## 6、信息系统安全管理

- (1) 掌握信息系统安全策略
- (2) 了解信息安全系统工程
- (3) 掌握 PKI 公开密钥基础设施
- (4) 掌握 PMI 权限（授权）管理基础
- (5) 了解信息安全审计
- (6) 了解信息系统安全管理其它方法

## 7、知识管理

- (1) 了解知识和知识管理概念
- (2) 掌握知识管理常用的方法和工具
- (3) 了解知识产权保护
- (4) 了解知识产权保护相关法律与政策

# 四、大数据安全

## （一）网络安全

1、熟悉网络安全的一些基本概念、基本属性及网络安全模型等；了解网络安全评估；掌握安全策略的功能；掌握各种类型的网络安全服务。

2、了解网络安全体系的安全服务、安全机制、安全服务配置和安全管理等；掌握 TCP/IP 协议安全体系。

3、熟悉各种网络安全技术的概念，了解物理安全、网络安全设备、系统安全、应用安全的技术知识及设置方法

4、了解网络安全运维与保障的框架、基本原理等知识和方法，掌握安全管理的技术、设施、工具和方法。

## (二) 数据安全

1、了解大数据安全管理及安全需求的原则，掌握数据安全分类、定级的原则、分类和方法。

2、了解数据采集、存储、处理、分发、删除等活动的概念，了解相关安全技术与安全要求。

3、熟悉云存储安全体系、数据生命周期中的安全风险、保障云存储安全的原则及云计算边界安全相关的概念；了解云计算安全架构和应对策略，掌握IaaS, PaaS, SaaS层云计算安全及边界安全技术。

4、了解大数据中心的安全保障机制、Hadoop平台的安全机制、大数据业务流程安全防护机制、大数据采集与存储的安全防护机制、大数据挖掘与应用的安全防护机制、大数据安全审计、大数据安全评估与安全管理系统。

## 第二部分：专业知识

### 五、数据库系统(SQL + NoSQL)

1、掌握关系模型概念和 SQL 语言。

2、掌握关系数据库设计方法。

3、掌握关系数据库索引的概念和使用方法。

4、掌握关系数据库查询处理与查询优化方法。

5、掌握关系数据库事务概念和事务调度方法。

6、掌握数据库并发控制技术。

7、了解非关系数据库背景、特点和分类。

8、了解分布式系统数据一致性理论，包括强一致性模型、弱一致性模型、最终一致性模型、CAP理论和BASE模型。

### 六、数据仓库

1、了解数据挖掘的定义、功能、常用方法。

2、了解数据仓库的产生与发展，掌握数据仓库的定义。

3、了解数据仓库与数据挖掘的联系与区别。

4、了解常用数据挖掘工具。

5、了解数据仓库的体系结构，掌握元数据、粒度、分割的概念，了解数据

仓库中的数据组织形式。

6、掌握数据仓库的数据模型，包括概念数据模型、逻辑数据模型、物理数据模型。

7、了解高层数据模型、中间层数据模型和低层数据模型。

8、掌握数据仓库的设计，包括概念模型设计、逻辑模型设计、物理模型设计。

9、了解数据仓库的生成、使用和维护方法。

10、掌握联机分析处理技术（OLAP）的概念、特征。

11、掌握 OLAP 中的多维分析操作，包括钻取、切片和切块、旋转。

12、了解多维联机分析处理、关系联机分析处理、MOLAP 和 ROLAP 的差异、混合型联机分析处理；OLAP 的衡量标准。

13、了解数据预处理的原因、数据预处理的方法。

14、掌握数据清洗、数据集成和变换、数据归约的概念与方法，具有应用上述方法进行数据清洗的能力。

15、了解关联规则的概念和分类。

16、了解文本数据挖掘、Web 数据挖掘、多媒体数据挖掘的概念，具有应用数据挖掘方法进行上述数据挖掘的能力。

## 七、大数据技术与应用

### （一）数据可视化

1、了解如下各类可视化图形的概念和实现方法，具有可视化图形展示数据的能力。

（1）饼图、环图、双层环图、柱状图、堆积柱状图、分组柱状图、正负条形图、组成瀑布图、折线图、散点图、层叠面积图

2、了解如下地理坐标/地图的数据可视化概念。

（1）k 线图、盒须图、雷达图、旭日图、南丁格尔玫瑰图、漏斗图、立体气泡图、矩形树图、热力图、日历坐标系、矩形树图、弦图、关系图、路径图、树图、桑葚图、平行坐标系、漏斗图、主题河流图、3D 曲面、3D 散点图、3D 柱状图

### （二）大数据处理技术

- 1、掌握云计算的概念、云计算的特征、云计算的三层 SPI 模型（IaaS、PaaS、SaaS）。
- 2、掌握虚拟化技术的概念、常用虚拟化方法，了解常用虚拟化软件。
- 3、具有应用虚拟化技术搭建虚拟化平台的能力。
- 4、了解分布式计算、高性能计算、边缘计算、工业换联网、5G 的概念。
- 5、了解有监督学习、无监督学习、强化学习的概念和区别。
- 6、掌握数据集市、数据仓库、数据中台、数据平台的概念。
- 7、掌握数据采集与预处理过程，掌握数据采集、数据清洗、数据转换、数据脱敏的方法。
- 8、了解数据存储与管理技术、大数据处理架构 Hadoop、分布式文件系统 HDFS、NoSQL 数据库、分布式数据库 HBase 的相关概念。
- 9、掌握如下大数据处理环节的主要思想：大数据采集、大数据预处理、大数据存储及管理、大数据分析及挖掘、大数据展现和应用（包括大数据检索、大数据可视化、大数据应用、大数据安全等）。
- 10、了解大数据处理与分析的代表性流行产品。
- 11、具有应用大数据技术分析综合案例的能力，包括案例需求分析、系统设计、技术选择、系统实现等环节。

## 八、大数据分析模型

- 1、了解常用的大数据分析模型，例如行为事件分析模型、漏斗分析模型、分布分析模型、点击分析模型、用户行为路径分析模型、用户分群分析模型等，具有应用大数据技术进行大数据分析能力。

## 九、数据科学

### （一）概率统计与统计推断

- 1、掌握概率、条件概率的概念，并会计算简单的概率、条件概率。
- 2、理解随机事件的独立性和随机变量的独立性。
- 2、掌握乘法公式、全概率公式、贝叶斯公式，并会简单应用。
- 3、掌握随机变量分布函数的概念，掌握连续型随机变量的密度函数和离散型随机变量的分布列。

4、掌握常见分布（两点分布、二项分布、泊松分布、均匀分布、正态分布、指数分布）及其简单性质。

5、理解数学期望、方差（标准差）、相关系数的概念，并会简单计算。

6、了解多元随机变量的概念，掌握常见多元分布。

7、理解统计量的概念，理解样本均值、样本方差（标准差）、样本相关系数的概念，并会简单计算。

8、了解点估计的概念，了解点估计的无偏性、相合性、有效性。

9、了解参数的矩估计、最大似然估计、区间估计。

10、了解假设检验的基本步骤，会单正态总体参数的假设检验。

11、了解线性回归的基本概念，会简单应用。

12、了解方差分析（ANOVA）的基本概念。

## （二）统计学习与数据挖掘

1、掌握感知机的定义和基本原理。

2、掌握 Logistic 回归算法原理与特点，能够使用 Logistic 回归进行数据的分类建模与参数解释。

3、掌握朴素贝叶斯算法的定义与基本原理。

4、掌握 k 近邻算法算法定义与原理。

5、掌握支持向量机（SVM）算法的思想与原理。

6、掌握决策树的算法的定义与原理，了解决策树的剪枝理论。

7、了解常见集成方法如 boosting、bagging 等。

8、掌握聚类分析的相关概念。

9、掌握主成分分析的原理。

10、了解生成式半监督学习方法、半监督 SVM、图半监督学习方法的思想与原理。

## （三）深度学习与强化学习

1、掌握全连接网络（MLP）的定义与算法原理、并能够应用。

2、掌握卷积神经网络（CNN）基本概念，了解几种经典的卷积神经网络，如 AlexNet、VGG、GoogLeNet、ResNet。

3、掌握循环神经网络（RNN）原理与基本概念，了解几种常见的模型，如长

短期记忆网络 LSTM、GRU。

4、掌握生成对抗网络（GAN、WGAN）的基本定义与原理，能够完成基本应用

5、掌握马尔可夫决策过程的定义，了解其算法应用范围，并能够实际应用

6、掌握经典强化学习算法（Q-learning、SARMA）的定义与原理。

7、掌握经典深度强化学习算法（DQN、DDPG、A3C、TRPO、PPO）的定义与原理，了解其算法特性。