

高纲 4077

江苏省高等教育自学考试大纲

# **14382 物联网大数据处理技术**

南京信息工程大学编（2024 年）

# I 课程性质与课程目标

## 一、课程性质和特点

本课程是针对物联网工程专业(专升本)开设的专业课,是为从事科学研究、理论研究、工程实践提供基本知识储备和基本训练的基础理论性课程。大数据是IT行业的颠覆性技术,将对人类社会的生产和生活产生重大而深远的影响,大数据处理有以下几个特点:(1)平台一定是大规模和可扩展的分布式系统;(2)处理数据的软件架构基于 MapReduce 分布式处理;(3)需要各种有效的存储技术和调度算法的支持。

本课程建立在《物联网工程导论》《数据通信基础》《宽带 IP 网络》等课程知识的基础上,是一门介绍大数据理论与实践相结合的课程,在注重大数据应用环境前提下,围绕存储、查询、安全等方面的技术,以经典的大数据处理流程、算法和实例对比性地介绍大数据常用的各种基础工具,由浅及深、由点及面地逐步介绍大数据处理技术,为物联网工程专业后继课程的学习打下坚实的基础。课程定位为大数据技术入门课程,为考生搭建起通向大数据知识空间的桥梁和纽带,帮助考生形成对大数据知识体系及其应用领域的轮廓性认识。通过本课程内容的学习,能够使考生对物联网大数据的发展、处理、分析、服务和应用的工程化方法有较为系统和完整的理解。以这些内容作为对智能数据处理技术的代表和延伸,旨在传递概念、方法和技术内涵与理论,为考生进一步学习和研究面向数据科学的方法和技术提供支撑。

## 二、课程目标

作为物联网工程专业的专业课,要求考生具备较好的数据素养和数理基础,扎实的大数据基础知识与技能,掌握大数据各个环节的技术,包括大数据采集、预处理、存储、处理、分析、应用技术。具体包括了解物联网大数据的概念和发展历程,各国的大数据发展战略及大数据的应用与产业发展;熟悉大数据与其他新兴技术如云计算、物联网、人工智能及区块链的关系;掌握大数据的基础知识,包括大数据安全、大数据思维、大数据伦理、数据共享、数据开放和大数据交易;熟悉大数据应用场景,包括大数据在互联网、生物医药、物流、城市管理、金融、汽车、零售、安全等各种领域的应用;掌握大数据处理的基本技术,包括数据采集与预处理、数据存储与管理技术;熟悉数据处理的进阶技术,包括机器学习和

数据挖掘算法、大数据处理与分析技术、大数据处理与分析的代表性产品；熟悉数据可视化技术与工具；通过大数据分析综合案例加深对大数据应用的理解，为开展物联网大数据的实际应用工作打下基础。本课程的目标大致包括：

1. 掌握物联网大数据的概念、特征、技术体系及相关关键技术，培养大数据思维，养成数据安全意识；
2. 了解大数据知识体系，熟悉物联网数据处理的应用背景，对数据处理对象所在的系统完整过程有较全面的认识，对数据处理之前的过程有基本的理解；
3. 熟悉大数据各个环节的相关技术，掌握大数据处理的基本技术，并对前沿的大数据处理与分析技术、数据可视化有较为清晰的认知；
4. 熟悉大数据处理、分析的应用案例，并形成对大数据处理和分析的全局性认知，了解大数据理论和技术的综合运用方法。

### 三、与相关课程的联系与区别

《物联网大数据处理技术》是江苏省高等教育自学考试物联网工程专业（专升本）必修的专业课程，与物联网工程专业的许多其他课程有着密切的关系。《物联网工程导论》《数据通信基础》《宽带 IP 网络》等课程是本课程的基础。

### 四、课程的重点和难点

本课程的重点之一是物联网大数据处理各个环节的技术，包括数据采集与预处理、数据存储与管理，数据处理与分析，数据可视化，尤其是数据存储管理中的大数据处理架构 Hadoop 与分布式文件系统 HDFS，以及大数据处理与分析的代表性产品 MapReduce，需要掌握其技术细节。

本课程的重点之二是物联网大数据应用，包括大数据在互联网、生物医药、物流、城市管理、金融、汽车、零售、餐饮、电信、能源、体育娱乐、安全、政府、日常生活等各种领域的应用，需要熟悉各类应用中使用的算法、模型，并了解相关案例。

本课程的难点是物联网大数据处理与分析的基础知识和关键技术，物联网大数据处理与分析的基础知识包括数据清洗、数据转换和数据脱敏等数据预处理技术；物联网大数据处理与分析的关键技术包括如聚类、回归关联规则和协同过滤在内的机器学习和数据挖掘算法，流计算和图计算，大数据处理与分析的代表性产品如分布式计算框架 MapReduce、数据仓库 Hive 及 Impala，机器学习框架和

流计算框架、大数据编程框架和查询分析系统等，需要熟悉各种关键技术的原理和算法。

## II 考核目标

本大纲在考核目标中，按照识记、领会、简单应用和综合应用四个层次规定其应达到的能力层次要求。四个能力层次是递进关系，各能力层次的含义是：

识记：要求考生能够识别和记忆本课程中有关物联网大数据处理的相关名词、概念、知识的含义，并能够根据考核的不同要求，进行正确的表述和选择。

领会：要求考生能够领悟和理解本课程中有关物联网大数据的概念、理论和方法技巧的内涵及外延，能够鉴别关于概念和特性的似是而非的说法，理解相关物联网大数据知识的区别和联系，能根据考核的不同要求对相关物联网大数据处理知识进行论证，做出正确的解释和说明。此外，能根据考核的不同要求分析各种不同领域或背景下所使用的数据处理和分析技术。

简单应用：要求考生能够根据已知的物联网大数据处理对象的一个应用需求或应用背景，提出可以采用的大数据处理技术，比如数据预处理技术、大数据处理框架、分布式文件系统、数据库处理技术、大数据处理与分析技术、数据可视化技术等。能够分析和解决有关的物联网大数据的理论问题和实际问题，并得出正确的结论。

综合应用：要求考生能够根据已知的物联网大数据的多个知识点，分析和解决有关的物联网理论问题和实际问题，并得出解决问题的综合方案。对物联网大数据处理的相关应用具有一定的分析和设计能力，在相关产品的组成结构、功能特点及数据处理技术方面具有一些基本的见解。

## III 课程内容与考核要求

### 第1章 大数据概述

#### 一、学习目的与要求

通过本章学习，要求考生理解数据的基础知识，了解大数据时代，了解大数据的发展历程，了解各国的大数据发展战略，掌握大数据的概念，了解大数据的影响，理解大数据的应用和大数据产业。

## 二、考核知识点与考核要求

### （一）数据

识记：①数据的基础知识。

### （二）大数据时代

简单应用：①大数据时代。

### （三）大数据的发展历程

领会：①大数据的发展历程。

### （四）世界各国的大数据发展战略

领会：①美国、英国、法国、韩国、中国的大数据发展战略。

### （五）大数据的概念

识记：①大数据的概念、四个特点。

### （六）大数据的影响

领会：①大数据对科学研究、社会发展、就业市场、人才培养的影响。

### （七）大数据的应用

领会：①大数据在各个领域的应用总体情况。

### （八）大数据产业

领会：①大数据产业的各个层次及其包含内容。

### （九）高校大数据专业（本节内容不作考核要求）

## 三、本章的重点和难点

本章重点：①大数据的背景；②大数据的特点；③大数据的影响。

本章难点：①大数据的四个特点。

## 第2章 大数据与其他新兴技术的关系

### 一、学习目的与要求

通过本章学习，要求考生掌握大数据与其他新兴技术的关系。重点理解大数据与云计算、物联网、人工智能、区块链之间存在的紧密联系。要求对云计算和物联网，大数据与云计算、物联网的关系有清晰认知，熟悉人工智能的基础知识，并理解大数据与人工智能的关系，熟悉区块链的知识，并理解大数据与区块链的关系。

## 二、考核知识点与考核要求

### （一）云计算

识记：①云计算的概念。

领会：①云计算的服务模式和类型；②云计算的应用；③云计算产业。

综合应用：①云计算数据中心。

### （二）物联网

识记：①物联网的概念。

领会：①物联网的关键技术。

综合应用：①物联网的应用；②物联网产业。

### （三）大数据与云计算、物联网的关系

领会：①大数据、云计算、物联网三者的关系。

### （四）大数据与人工智能

识记：①人工智能的概念。

领会：①人工智能的关键技术；②大数据与人工智能的关系。

综合应用：①人工智能的应用；②人工智能产业。

### （五）大数据与区块链

识记：①比特币概念；②区块链的定义。

领会：①区块链的原理；②大数据与区块链的关系。

综合应用：①区块链的应用。

## 三、本章的重点和难点

本章重点：①物联网关键技术；②大数据与云计算、物联网、人工智能、区块链之间存在的紧密联系。

本章难点：①物联网关键技术；②人工智能关键技术；③区块链的原理。

## 第3章 大数据基础知识

### 一、学习目的与要求

通过本章学习，要求考生掌握大数据基础知识，包括大数据安全、大数据思维、大数据伦理、数据共享、数据开放、大数据交易等，通过对相关理论和案例的学习、理解，提升考生的数据素养，培养对科学数据的采集与预处理、存储与管理、处理与分析、共享与协同创新利用等方面的能力。

### 二、考核知识点与考核要求

### （一）大数据安全

领会：①传统数据安全；②大数据安全与传统数据安全的不同；③大数据安全问题。

综合应用：①大数据安全典型案例。

### （二）大数据思维

领会：①传统的思维方式；②大数据时代新的思维方式的必要性；③大数据思维方式。

综合应用：①运用大数据思维的具体实例。

### （三）大数据伦理

识记：①大数据伦理的概念。

领会：①大数据的伦理问题。

综合应用：①大数据伦理典型案例。

### （四）数据共享

领会：①数据孤岛问题；②数据孤岛问题产生的原因；③实现数据共享所面临的挑战；④推进数据共享开放的举措。

综合应用：①数据共享案例。

### （五）数据开放

领会：①政府开放数据的理论基础；②政府信息公开与政府数据开放的联系与区别；③政府数据开放的重要意义。

### （六）大数据交易

识记：①大数据交易形式。

领会：①大数据交易发展现状。

简单应用：①大数据交易平台。

## 三、本章的重点和难点

本章重点：①大数据安全、大数据思维、大数据伦理、数据共享的应用案例；②大数据交易平台。

本章难点：①大数据伦理；②数据孤岛；③数据开放。

## 第4章 大数据应用

### 一、学习目的与要求

通过本章学习，要求考生掌握大数据在各大领域的典型应用，包括互联网、生物医药、物流、城市管理、金融、汽车、零售、餐饮、电信、能源、体育、娱乐、安全、政府和日常生活等领域。

## 二、考核知识点与考核要求

### （一）大数据在互联网领域的应用

识记：①推荐系统概念。

领会：①长尾理论；②推荐方法；③推荐系统模型。

综合应用：①推荐系统的应用。

### （二）大数据在生物医学领域的应用

领会：①流行病预测；②智慧医疗；③生物信息学。

综合应用：①基于大数据的综合健康服务平台。

### （三）大数据在物流领域的应用

识记：①智能物流的概念。

领会：①智能物流的作用；②智能物流的应用；③大数据是智能物流的关键。

综合应用：①中国智能物流骨干网——菜鸟。

### （四）大数据在城市管理领域的应用

领会：①智能交通；②环保监测；③城市规划；④安防；⑤疫情防控。

### （五）大数据在金融领域的应用

简单应用：①高频交易；②市场情绪分析；③信贷风险分析；④大数据征信。

### （六）大数据在汽车领域的应用

综合应用：①大数据在汽车领域的应用；②发现关联购买行为；③客户群体细分；④供应链管理。

### （七）大数据在零售领域的应用

领会：①发现关联购买行为；②供应链管理。

综合应用：①客户群体细分。

### （八）大数据在餐饮领域的应用

综合应用：①餐饮领域拥抱大数据；②餐饮 O2O。

### （九）大数据在电信领域的应用

领会：①电信客户离网分析；②智能电网。



综合应用：①国内大数据在电信领域的具体应用；②大数据在能源领域的应用。

#### （十）大数据在能源领域的应用

领会：①智能电网和大数据的关系。

#### （十一）大数据在体育和娱乐领域的应用

简单应用：①训练球队；②投拍影视作品；③预测比赛结果。

#### （十二）大数据在安全领域的应用

领会：①大数据与国家安全。

简单应用：①应用大数据技术防御网络攻击；②警察应用大数据工具预防犯罪。

#### （十三）大数据在政府领域的应用

简单应用：①大数据在政府领域的应用。

#### （十四）大数据在日常生活中的应用

简单应用：①大数据在日常生活中的应用。

### 三、本章的重点和难点

本章重点：①大数据在互联网领域的应用；②大数据在生物医学领域的应用；③大数据在物流领域的应用；④大数据在金融领域的应用。

本章难点：①大数据在互联网领域的应用；②大数据在生物医学领域的应用；③大数据在物流领域的应用。

## 第5章 数据采集与预处理

### 一、学习目的与要求

通过本章学习，要求考生掌握大数据技术的第一道环节，即数据采集与预处理，熟悉通过数据采集能够得到的数据，包括传感器数据、互联网数据、日志文件、企业业务系统数据等，以及数据预处理所包含的数据清洗、数据转换和数据脱敏。

### 二、考核知识点与考核要求

#### （一）数据采集

识记：①数据采集的概念；②数据采集的三大要点。

领会：①数据采集的数据源；②数据采集方法。

综合应用：①网络爬虫。

## （二）数据清洗

领会：①数据清洗的内容；②数据清洗的注意事项。

## （三）数据转换

识记：①数据转换策略。

领会：①平滑处理；②规范化处理。

## （四）数据脱敏

领会：①数据脱敏原则；②数据脱敏方法。

## 三、本章的重点和难点

本章重点：①数据采集方法；②数据清洗；③数据转换；④数据脱敏。

本章难点：①数据清洗；②数据转换；③数据脱敏。

# 第6章 数据存储与管理

## 一、学习目的与要求

通过本章学习，要求考生掌握大数据分析流程中的数据存储与管理环节，清楚其对于大数据分析的意义。熟悉传统的数据存储与管理技术，掌握大数据时代的数据存储与管理技术，熟悉其代表性产品。

## 二、考核知识点与考核要求

### （一）传统的数据存储与管理技术

领会：①文件系统；②关系数据库；③数据仓库；④并行数据库。

### （二）大数据时代的数据存储与管理技术

领会：①分布式文件系统；②NewSQL 和 NoSQL 数据库；③云数据库。

### （三）大数据处理架构 Hadoop

识记：①Hadoop 特性。

领会：①Hadoop 生态系统。

### （四）分布式文件系统 HDFS

识记：①HDFS 的设计目标。

领会：①HDFS 体系结构。

### （五）NoSQL 数据库

领会：①键值数据库；②列族数据库；③文档数据库；④图数据库。

### （六）云数据库

识记：①云数据库的概念。

领会：①云数据库的特性；②云数据库与其他数据库的关系；③代表性云数据库产品。

### （七）分布式数据库 HBase

识记：①Bigtable；②HBase 简介。

领会：①HBase 数据模型；②HBase 系统架构。

### （八）Google Spanner

识记：①Spanner 概念；②特性。

领会：①Spanner 部署；②Spanner 服务器的组织方式。

## 三、本章的重点和难点

本章重点：①NewSQL 和 NoSQL 数据库；②云数据库；③大数据处理架构 Hadoop 和分布式文件系统 HDFS。

本章难点：①关系数据库；②数据仓库；③NewSQL 和 NoSQL 数据库；④云数据库。

## 第 7 章 数据处理与分析

### 一、学习目的与要求

通过本章学习，要求考生掌握数据处理与分析的概念，熟悉机器学习和数据挖掘算法，掌握包括流计算和图计算等大数据处理与分析技术，熟悉大数据处理与分析的代表性产品。本章理论性较强，技术前沿，属于课程较为核心和复杂的部分，要求考生能够综合把握。

### 二、考核知识点与考核要求

#### （一）数据处理与分析的概念

识记：①数据分析与数据挖掘概念；②数据分析与数据处理概念；③大数据处理与分析概念。

#### （二）机器学习和数据挖掘算法

识记：①机器学习概念；②数据挖掘概念；③典型机器学习和数据挖掘算法（包括分类、聚类、回归分析和关联规则）的概念。

领会：①分类方法；②聚类；③回归分析方法；④关联规则方法；⑤协同过

滤方法。

### （三）大数据处理与分析技术

识记：①大数据处理与分析技术分类；②流计算概念。

领会：①流计算处理流程；②图计算。

### （四）大数据处理与分析代表性产品

识记：①分布式计算框架 MapReduce 概念；②基于内存的分布式计算框架 Spark 概念。

领会：①分布式计算框架 MapReduce 工作流程与不足之处；②数据仓库 Hive；③数据仓库 Impala；④基于内存的分布式计算框架 Spark；⑤机器学习框架 TensorFlowOnSpark；⑥流计算框架 Storm；⑦流计算框架 Flink；⑧大数据编程框架 Beam；⑨查询分析系统 Dremel。

## 三、本章的重点和难点

本章重点：①机器学习和数据挖掘算法；②大数据处理与分析代表性产品（Hive，Spark 和 Storm）。

本章难点：①大数据处理与分析代表性产品（Hive，Spark 和 Storm）。

# 第 8 章 数据可视化

## 一、学习目的与要求

通过本章学习，要求考生掌握大数据分析的最后环节数据可视化，掌握可视化的概念，熟悉可视化图表，了解可视化工具，熟悉典型的可视化案例。

## 二、考核知识点与考核要求

### （一）可视化概述

识记：①数据可视化概念；②数据可视化发展历程。

领会：①可视化的作用。

### （二）可视化图表

识记：①常用的可视化图表，包括统计类图表及其他图表。

### （三）可视化工具

识记：①可视化入门级工具。

领会：①信息图表工具；②地图工具；③时间线工具；④高级分析工具。

### （四）可视化典型案例

综合应用：①全球黑客活动；②互联网地图；③编程语言之间的影响力关系图；④世界国家健康与财富之间的关系；⑤3D 可视化互联网地图 App。

### 三、本章的重点和难点

本章重点：①可视化图表和可视化工具。

本章难点：①可视化工具（信息图表工具）。

## 第9章 大数据分析综合案例

### 一、学习目的与要求

通过本章学习，要求考生熟悉大数据分析案例中的系统设计，包括系统总体设计、数据库设计、系统网站设计和算法设计，系统设计中的技术选择，系统实现以及案例所需知识和技能，要求考生掌握大数据分析全流程的相关理论和技术及其如何有机融合在一起来解决实际应用中的具体问题，形成对大数据分析的全局性的轮廓认识，连接大数据理论和技术的综合运用方法。

### 二、考核知识点与考核要求

#### （一）案例任务

识记：①电影推荐系统功能。

#### （二）系统设计

领会：①系统总体设计；②数据库设计；③系统网站设计；④算法设计。

#### （三）技术选择

识记：①案例中所用到的实现技术。

领会：①案例中的数据分析过程。

#### （四）系统实现

识记：①系统实现所涉及到的主要任务。

#### （五）案例所需知识和技能

识记：①案例所需知识和专业技能。

### 三、本章的重点和难点

本章重点：①系统涉及的步骤。

本章难点：①大数据分析案例中的数据分析过程。

## IV 关于大纲的说明与考核实施要求

### 一、自学考试大纲的目的和作用

课程自学考试大纲是根据专业考试计划的要求，结合自学考试的特点而确定。其目的是对个人自学、社会助学和课程考试命题进行指导和规定。

课程自学考试大纲明确了课程学习的内容以及深广度，规定了课程自学考试的范围和标准。因此，它是编写自学考试教材和辅导书的依据，是社会助学组织进行自学辅导的依据，是考生学习教材、掌握课程内容知识范围和程度的依据，也是进行自学考试命题的依据。

### 二、课程自学考试大纲与教材的关系

课程自学考试大纲是进行学习和考核的依据，教材是学习掌握课程知识的基本内容与范围，教材的内容是大纲所规定的课程知识和内容的扩展与发挥。课程内容在教材中可以体现一定的深度或难度，但在大纲中对考核的要求一定要适当。

大纲与教材所体现的课程内容应基本一致；大纲里面的课程内容和考核知识点，教材里一般也要有。反过来教材里有的内容，大纲里就不一定体现。

### 三、关于自学教材

本课程使用教材为：《大数据导论》，林子雨编著，人民邮电出版社，2020年版。

### 四、关于自学要求和自学方法的指导

本大纲的课程基本要求是依据专业考试计划和专业培养目标而确定的。课程基本要求还明确了课程的基本内容，以及对基本内容掌握的程度。基本要求中的知识点构成了课程内容的主体部分。因此，课程基本内容掌握程度、课程考核知识点是高等教育自学考试考核的主要内容。

为有效地指导个人自学和社会助学，本大纲已指明了课程的重点和难点，在章节的基本要求中一般也指明了章节内容的重点和难点。

根据物联网工程专业（专升本）的要求，以及本课程的特点，考生在进行自学时应注意以下几点：

1. 在学习本课程前，应仔细阅读课程大纲的第一部分，了解课程的性质、地位和任务，熟知课程的基本要求以及与本课程有关的课程的联系，以便使以后的学习能紧紧围绕课程的基本要求。

2. 在阅读某一章教材内容前，考生应先认真阅读大纲中关于该章的考核知识点、自学要求和考核要求，注意对各知识点的能力层次要求，以便在阅读教材时做到心中有数，有的放矢。

3. 阅读教材时，考生应根据大纲要求，要逐段细读，逐句推敲，集中精力，吃透每个知识点，对基本概念必须深刻理解，基本原理必须牢固掌握，在阅读中遇到个别细节问题不清楚，在不影响继续学习的前提下，可暂时搁置。

4. 学完教材的每一章内容后，考生应针对考试大纲列出的重点、难点认真复习，以便进一步理解、消化和巩固所学知识，增强分析问题、解决问题的能力。

## **五、应考指导**

### **1. 如何学习**

很好的计划和组织是你学习成功的法宝。如果考生正在接受培训学习，一定要跟紧课程并完成作业。为了在考试中做出满意的回答，考生必须对所学课程内容有很好的理解，如使用“行动计划表”来监控你的学习进展。考生阅读课本时可以做读书笔记，可以用彩笔来标注需要重点注意的内容，如红色代表重点、绿色代表需要深入研究的领域、黄色代表可以运用在工作之中。此外，还可以在空白处记录相关网站与文章。

### **2. 如何考试**

卷面整洁非常重要。书写工整，段落与间距合理，卷面赏心悦目有助于教师评分，教师只能为他能看懂的内容打分。回答所提出的问题。要回答所问的问题，而不是回答你自己乐意回答的问题！避免超过问题的范围。

### **3. 如何处理紧张情绪**

正确处理对失败的惧怕，要正面思考。如果可能，请教已经通过该科目考试的人，问他们一些问题。做深呼吸放松，这有助于使头脑清醒，缓解紧张情绪。考试前合理膳食，保持旺盛精力，保持冷静。

### **4. 如何克服心理障碍**

这是一个普遍存在的问题！如果你在考试中出现这种情况，可以试试下列方法：使用“线索”纸条。进入考场之前，将记忆“线索”记在纸条上。但你不能将纸条带进考场，因此当你阅读考卷时，一旦有了思路就快速记下。按自己的步调进行答卷。为每个考题或部分合理分配时间，并按此时间安排进行。要做到心理障碍的排除，还可以试着在考试的时候一边浏览试题一边回顾书本大纲，梳理出一个题目分布的框架，这样你就会对答题有进一步的把握。

## 六、对社会助学的要求

1. 社会助学者应根据本大纲规定的考试内容和考核目标，以及各章的难点和重点，认真钻研指定教材，明确本课程的特点和学习要求，对考生进行切实有效的辅导，避免考生在自学时可能出现的各种偏向，把握社会助学的正确方向。

2. 社会助学者应对考生进行学习方法的指导，向考生提倡“认真阅读教材，刻苦钻研教材，主动提出问题，依靠自己学懂”的学习方法。

3. 社会助学者应注意对考生自学能力的培养，使考生逐步学会独立学习，在自学过程中善于提出问题、分析问题、作出判断和解决问题。对考生提出的问题，社会助学者应以启发引导为主。

4. 社会助学者应努力引导考生将识记、领会、简单应用和综合应用联系起来，将基本知识转化为识记工作能力，全面培养和提升考生的综合素质。

5. 社会助学者应指导考生正确处理重点和一般的关系，帮助考生掌握全部考试内容和考核知识点，切勿孤立地抓重点，将考生引向猜题和押题。

## 七、对考核内容的说明

本课程要求考生学习和掌握的知识点内容都作为考核的内容。课程中各章的内容均由若干知识点组成，在自学考试成为考核知识点。因此，课程自学考试大纲中所规定的考试内容是以分解为考核知识点的方式给出的。由于各知识点在课程中的地位、作用以及知识自身的特点不同，自学考试将对各知识点分别按四个能力层次确定其考核要求。

## 八、关于考试命题的若干规定

1. 本课程考试方式为闭卷、笔试，考试时间为 150 分钟。评分采用百分制，60 分为及格。考生只准携带 0.5 毫米黑色墨水的签字笔、铅笔、圆规、直尺、三角板、橡皮等必需的文具用品。不可携带计算器。



2. 本大纲各章所规定的基本要求、知识点及知识点下的知识细目，都属于考核的内容。考试命题既要覆盖到章，又要避免面面俱到。要注意突出课程的重点、章节重点，加大重点内容的覆盖度。

3. 命题不应有超出大纲中考核知识点范围的题，考核目标不得高于大纲中所规定的相应的最高能力层次要求。命题应着重考核考生对基本概念、基本知识和基本理论是否了解或掌握，对基本方法是否会用或熟练。不应出与基本要求不符的偏题或怪题。

4. 本课程在试卷中对不同能力层次要求的分数比例大致为：识记占 20%，领会占 30%，简单应用占 30%，综合应用占 20%。

5. 要合理安排试题的难易程度，试题的难度可分为：易、较易、较难和难四个等级。每份试卷中不同难度试题的分数比例一般为 2:3:3:2。

必须注意，试题的难易程度与能力层次有一定的联系，但二者不是等同的概念，在各个能力层次中都存在着不同难度的试题。

6. 本课程考试试卷中可能采用的题型有：单项选择题、判断改错题、名词解释题、简答题、论述题，案例分析题等。

## 附录 题型举例

### 一、单项选择题

1. 关系数据库对数据库进行查询使用的语言是（ ）

A. HTML                  B. XML                  C. SQL                  D. NoSQL

参考答案：C

2. 下列选项中属于列族数据库产品的是（ ）

A. Redis                  B. HBase                  C. MongoDB                  D. Neo4j

参考答案：B

### 二、判断改错题

1. TwitterStorm 是一个免费、开源的集中式实时计算系统。（ ）

参考答案：×，“集中”改为“开放”。

### 三、名词解释题

1. 数据脱敏

参考答案：数据脱敏是在给定的规则、策略下对敏感数据进行变换、修改的技术。

## 2. 日历图

参考答案：日历图是以日历为基本维度的、对单元格加以修饰的图表。

## 四、简答题

### 1. 简述完整的物联网产业链主要包括的内容。

参考答案：

完整的物联网产业链主要包括核心感应器件提供商、感知层末端设备运营商、网络提供商、软件与行业解决方案提供商、系统集成商、运营及服务提供商等。

### 2. 简述智慧医疗的 3 个优点。

参考答案：

（1）促进优质医疗资源的共享；（2）避免患者重复检查；（3）促进医疗智能化。

## 五、论述题

### 1. 试述 MapReduce 存在的缺点。

参考答案：

（1）表达能力有限。计算都必须转化成 Map 和 Reduce 两种操作，但这并不适合所有的情况，难以描述复杂的数据处理过程。

（2）磁盘 IO 开销大。每次执行时都需要从磁盘读取数据，并且在计算完成后需要将中间结果写入磁盘中，IO 开销较大。

（3）延迟高。一次计算可能需要分解成一系列按顺序执行的 MapReduce 任务，任务之间的衔接由于涉及 IO 开销，会产生较高延迟。而且，在前一个任务执行完成之前，其他任务无法开始，因此难以胜任复杂、多阶段的计算任务。

### 2. 试述反爬机制产生的原因。

参考答案：

反爬机制产生的原因主要有两点：

第一，在大数据时代，数据是十分宝贵的财富，很多企业不愿意让自己的数据被别人免费获取，因此，它们为自己的网站设计反爬机制，防止网页上的数据被抓取；

第二，简单低级的网络爬虫，数据采集速度快，伪装度低，如果没有反爬机制，它们可以很快地抓取大量数据，甚至因为请求过多，造成网站服务器不能正常工作，影响企业的业务开展。

## 六、案例分析题

### 1. Google 的关键词广告系统 AdWords 不仅是世界上最赚钱的产品，对广告商来说也是

广告效果最好的平台。试分析 Google 是怎么兼顾自己和广告商的利益。

参考答案：

它巧妙地利用数据来形成双赢甚至多赢的格局，它的做法是收集大量的数据然后利用这些数据。比如说它掌握了广告被点击的数据，那么它在展示广告的时候，如果一个广告很少被点击，Google 就会尽量少地展示这个广告。带来的结果是什么？对广告商来说省钱了，因为我们不用花钱在无用的广告上面；对 Google 来说，不展示这些广告那就可以把有限而宝贵的搜索流量留给那些可能被点击的广告，从而增加自己的收入；对用户来说，我们也不会看到自己不想看并且跟自己没关系的广告，从而提升了用户的体验。这就是用数据来获得智能。

2. 2015 年 12 月 16 日，华东江苏大数据交易中心平台在盐城上线运营，这是国家批准设立的、华东地区首家也是唯一的一家跨区域、标准化、综合性的大数据交易平台。试从大数据应用的角度分析华东江苏大数据交易中心的业务发展模式。

参考答案：

华东江苏大数据交易中心以“大数据+产业+金融”为业务发展模式，将分散在各个信息孤岛的数据汇聚后通过关联、交叉、分析、挖掘，为全社会提供数据应用服务。同时，该中心基于数据金融资产证券化方向提供数据资产的典当、融资、抵押、贷款等多种业务模式，为各类经济主体，如政府、机构、企业以及个人，盘活数据存量资源、提供全面的综合解决方案。