北京信息科技大学
2027年硕士研究生入学考试初试自命题科目
考试大纲

考试科目名称：大数据技术基础　　考试科目代码：804

一、考试基本要求及适用范围概述

基本要求：1. 理解数据结构的基本概念、计算机内部数据对象的表示和特性。掌握线性表、树、图等数据逻辑结构、存储结构及其差异以及各种操作的实现。掌握算法时间和空间复杂性分析方法，能够针对实际问题选择合适的数据结构和方法设计出结构清晰、正确易读、复杂性较优的算法。

2. 理解数据库系统的基本概念、基本原理和基本结构，理解关系数据库、SQL 语言及关系数据库理论，掌握数据建模和数据库设计，掌握数据库体系结构及使用和管理方法，掌握 SQL 语言的运用，了解数据库的发展及一些前沿领域。

3. 理解大数据技术的基本概念、基本原理和基本方法。理解大数据分析处理流程，掌握常见的数据分析方法，了解大数据的发展趋势和研究前沿。

适用范围：本考试大纲适用于管理科学与工程学院（除125601工程管理外）所有专业的硕士生入学考试。

二、题型结构

（一）试卷满分及考试时间

本试卷满分为 150 分，考试时间为 180 分钟。

（二）答题方式

答题方式为闭卷、笔试。

（三）试题结构

主要试题类型包括：数据结构、数据库和大数据技术基础三部分的题目分数比例各占三分之一。主要采取客观题与主观题相结合的形式。

三、考试内容

第一部分　数据结构

（一）数据结构和算法简介

1. 什么是数据结构

2. 基本概念和术语

3. 算法和算法分析

（二）线性表

1. 线性表及其逻辑结构

2. 线性表的顺序表示与实现

3. 线性表的链式表示与实现

4. 线性表的应用

（三）栈与队列

1 栈

2. 队列

（四）树和二叉树

1. 树的定义和基本术语

2. 二叉树

3. 遍历二叉树和线索二叉树

4. 树和森林

5. 最优二叉树—哈夫曼树

（五）图

1. 图的定义和术语

2. 图的存储结构

3. 图的遍历

4. 生成树与最小生成树

5. 最短路径

第二部分　数据库

（一）数据库概论

1. 数据管理及其发展过程

2. 数据库系统的特点

3. 数据模型初步

4. 数据独立性与三层结构

5. 数据库管理系统和数据库系统

6. 数据库技术的研究和应用领域

（二）概念数据模型

1. 实体-联系方法

2. 概念数据模型设计

（三）关系数据库基础

1. 关系数据库系统概述

2. 关系数据模型

3. 关系模型的完整性约束

（四）数据定义与完整性约束的实现

1. SQL 的表定义和完整性定义功能

2. SQL 数据操作与完整性约束的作用

（五）SQL 语言

1. SQL 数据查询命令

2. 简单查询

3. 连接查询

4. 分组及汇总查询

5. 嵌套查询

6. 视图的概念

7. 存储过程的概念

（六）数据库设计

1. 数据库设计的步骤和方法

2. 逻辑数据模型设计

3. 物理数据模型设计

第三部分　大数据技术基础

（一）大数据概述

1. 数据的概念

2. 大数据的发展历程

3. 大数据的概念与影响

4. 大数据的应用以及大数据产业

（二）大数据与其他新兴技术的关系

1. 物联网的概念和应用

2. 云计算的概念和应用

3. 大数据和云计算以及物联网的相互关系

4. 人工智能的概念与关键技术

5. 大数据与人工智能的关系

6. 区块链的概念与技术原理

7. 大数据与区块链的关系

（三）大数据基础知识

1. 大数据安全

2. 大数据思维

3. 大数据伦理

4. 数据共享

5. 数据开放

6. 大数据交易

7. 大数据治理

（四）数据采集与预处理

1. 数据采集

2. 数据清洗

3. 数据转换

4. 数据脱敏

（五）数据存储与管理

1. 传统的数据存储与管理技术

2. 大数据时代的数据存储与管理技术

3. 大数据处理架构Hadoop

4. 分布式文件系统 HDFS

5. NoSQL 数据库

6. 云数据库

7. 分布式数据库 HBase

（六）数据处理与分析

1. 数据处理与分析的概念

2. 基于统计学方法的数据分析

3. 机器学习和数据挖掘算法

4. 大数据处理与分析技术

5. 大数据处理与分析领域具有代表性的产品

四、参考书目

序号	参考书目	编著者姓名	出版社名称	出版时间（第 X 版）
1	数据结构教程（第 6 版）	李春葆	清华大学出版社	2022 年（第 6 版）
2	数据库系统及应用（第 4 版）	崔巍	高等教育出版社	2017 年（第 4 版）
3	大数据导论（第 2 版）	林子雨	人民邮电出版社	2024 年（第 2 版）