# 大数据技术概论复习资料

## 一、客观题（20 题）

1. 数据在拉丁文里的意思是（ ）

A. 信息 B. 已知 C. 事实 D. 记录

1. 《牛津英语字典》中对 data 的定义不包括以下哪项（ ）

A. 由计算机执行操作的量、字符或符号

B. 以电信号的形式存储和传输

C. 记录在磁、光或机械记录介质上

D. 仅指数字和文字信息

1. 展示数据逐步转化为信息、知识及智慧过程的模型是（ ）

A. TCP/IP 模型 B. DIKW 模型 C. OSI 模型 D. 大数据处理模型

1. 数据集中，行通常表示（ ）

A. 特征 B. 属性 C. 样本 D. 维

1. 国家标准《信息技术大数据术语》（GB/T 35295—2017）中定义的大数据特征不包括（ ）

A. 体量巨大 B. 来源单一 C. 生成极快 D. 难以用传统数据体系结构有效处理

1. 图灵奖得主吉姆・格雷提出的第四范式是（ ）

A. 实验观测 B. 理论推演 C. 计算仿真 D. 数据驱动

1. 大数据在银行业应用中，不包括以下哪项（ ）

A. 客户画像 B. 精准营销 C. 车辆自动驾驶 D. 风险管理

1. 车联网技术的核心支撑是（ ）

A. 大数据技术 B. 传统数据库技术 C. 单机计算 D. 人工统计

1. 政府通过打造大数据施政平台的主要目的是（ ）

A. 增加政府办公成本 B. 打破政务大数据共享壁垒

C. 减少政务公开数据 D. 降低决策效率

1. 个性化医疗定制的基础是（ ）

A. 个人基因组信息 B. 医院规模 C. 医生经验 D. 医疗设备先进程度

1. 疫情防控中 “大数据＋网格化” 模式的核心作用是（ ）

A. 增加基层工作量 B. 实现疫情防控精细化、统一化

C. 限制人员流动 D. 仅统计感染人数

1. 数据采集的核心目的是（ ）

A. 从真实世界中采集和记录信息并数据化

B. 消耗存储资源 C. 增加数据冗余 D. 简化数据处理

1. 原始数据存在格式不一致、缺失值等问题时，需要进行（ ）

A. 数据采集 B. 数据预处理 C. 数据可视化 D. 数据存储

1. 关系数据库的基础是（ ）

A. 关系模型 B. 文档模型 C. 键值模型 D. 图形模型

1. HDFS 架构遵循的结构是（ ）

A. 主从结构 B. 星型结构 C. 环型结构 D. 总线结构

1. 将 SQL 的可靠性与 NoSQL 的速度和性能相结合的数据库是（ ）

A. 关系数据库 B. 文档数据库 C. NewSQL 数据库 D. 列存储数据库

1. 数据可视化的终极目的是（ ）

A. 绘制精美图表 B. 洞悉数据中的规律

C. 增加数据复杂度 D. 消耗计算资源

1. 用于展示两个数值变量之间关系的常用可视化图形是（ ）

A. 饼图 B. 散点图 C. 玫瑰图 D. 日历图

1. 从大量实际数据中提取隐含的、潜在价值信息的过程是（ ）

A. 数据采集 B. 数据预处理 C. 数据挖掘 D. 数据存储

1. 云计算的基本服务类型不包括（ ）

A. 基础设施即服务 B. 平台即服务 C. 软件即服务 D. 硬件即服务

## 二、判断题（20 题）

1. 数据仅指数字、文字，不包括图像、音频、视频等。（ ）
2. 数据化是释放信息潜在价值的基础。（ ）
3. 数据集的列通常表示样本，行表示特征。（ ）
4. 大数据的生成速度慢，且来源相对单一。（ ）
5. 第四范式的出现是科学发现和思维方式的革命性改变。（ ）
6. 社交网络舆情具有传播快、覆盖面广但信息均为积极健康的特点。（ ）
7. 大数据技术能帮助保险公司更准确地掌握客户真实风险水平。（ ）
8. 车联网不需要建立信息服务平台，仅依靠单一车辆数据即可运行。（ ）
9. 政务大数据共享不畅会造成信息资源闲置或浪费。（ ）
10. 个性化医疗仅依据患者的电子健康病例即可制定治疗方案。（ ）
11. 大数据在新冠肺炎疫情研究中，能帮助追踪病毒源起及传播路径。（ ）
12. 数据采集技术的好坏不影响后续的数据分析结果。（ ）
13. 数据预处理包括数据整理、变换、缺失值处理等操作。（ ）
14. 关系数据库以表格形式直观表示数据之间的关系。（ ）
15. HDFS 的主节点称为数据节点，从属节点称为名称节点。（ ）
16. NoSQL 数据库适用于处理大量非结构化数据。（ ）
17. 数据可视化的 “信” 是指图表设计要令人赏心悦目。（ ）
18. 直方图可用于查看数值型变量的分布特点。（ ）
19. 聚类分析中，同组内的数据差异大，不同组之间的数据相似度高。（ ）
20. MapReduce 编程模型主要包含 Mapping 和 Reducing 两个阶段。（ ）

## 三、简答题（8 题）

1. 简述数据的定义及常见类型划分角度。
2. 大数据的处理流程主要包括哪些步骤？
3. 简述大数据在政府领域的主要应用。
4. 数据预处理的目的是什么？主要包括哪些操作？
5. 简述 HDFS 的主要特点。
6. 数据可视化的基本原则是什么？
7. 简述分类问题的定义及分类算法的主要类型。
8. 大数据隐私保护的对策主要来自哪些方面？

## 四、综合题（2 题）

1. 请结合所学知识，分析大数据技术在电子商务领域的应用场景，并说明大数据技术如何帮助电子商务企业提升竞争力。
2. 请详细描述 MapReduce 的编程模型与执行流程，并举例说明其在实际场景中的应用