

重庆数字资源集团数据运营分公司

2026 年招聘考试押题试卷

全卷共 68 题，考试时间 120 分钟

一、单选题（共 58 题） 本部分共 58 题。

1. 某数据运营团队计划在 10 天内完成一批数据清洗任务，前 4 天完成了总量的 40%，若按此效率，剩余任务还需多少天完成？
A. 5 天
B. 6 天
C. 7 天
D. 8 天
2. 在数据治理中，数据质量评估通常不包括以下哪个维度？
A. 准确性
B. 完整性
C. 安全性
D. 时效性
3. 数据治理是将数据视为组织资产，通过制定数据标准、明确数据责任、规范数据流程，对数据进行全生命周期管理，以提升数据质量、保障数据安全、促进数据共享与应用的过程。根据上述定义，下列属于数据治理的是：
A. 某公司采购高性能服务器以扩大数据存储容量
B. 某单位制定各部门数据录入格式和校验规则，并定期清理异常数据
C. 某企业开发移动端报表应用，实时展示经营情况
D. 某团队使用电子表格汇总各部门月度销售数据
4. 某数据产品原价 200 元，先提价 10%，再打八折出售，则现价为多少元？
A. 176 元
B. 180 元
C. 184 元
D. 188 元
5. SQL 中，用于从表中选取数据的关键字是：
A. SELECT
B. UPDATE
C. DELETE

D. INSERT

6. 下列哪组词语与“大数据：数据”的逻辑关系最为相似？
- A. 核能：能源
 - B. 云计算：计算机
 - C. 书籍：纸张
 - D. 网络：信息
7. 某数据平台每天处理的数据量以固定比例增长，若第一天处理 100 万条，第三天处理 144 万条，则第二天的处理量为多少万条？
- A. 110
 - B. 115
 - C. 120
 - D. 125
8. 在 Hadoop 生态系统中，用于分布式存储的组件是：
- A. MapReduce
 - B. HDFS
 - C. YARN
 - D. Spark
9. 如果不执行数据标准，那么数据质量无法保证。某单位的数据质量得到了保证，由此可以推出：
- A. 该单位执行了数据标准
 - B. 该单位没有执行数据标准
 - C. 该单位数据标准执行不到位
 - D. 该单位不需要数据标准
10. 某数据包从 A 地传输到 B 地，若传输速度提高 20%，则传输时间减少多少？
- A. 16.7%
 - B. 20%
 - C. 25%
 - D. 33.3%
11. 在数据预处理中，处理缺失值常用的方法不包括：
- A. 删除记录
 - B. 填充均值
 - C. 插值法
 - D. 增加噪声

12. 数据要素是指参与社会生产经营活动、为使用者带来经济效益的数据资源，是数字经济时代的关键生产要素。根据该定义，下列不属于数据要素的是：
- A. 电商平台记录的消费者购物偏好数据
 - B. 新能源汽车实时采集的行驶轨迹数据
 - C. 某企业内部员工考勤签到记录
 - D. 气象部门公开发布的实时气象数据
13. 某数据产品每次迭代后用户量增长 15%，若初始用户量为 1000，则两次迭代后用户量约为多少？
- A. 1322
 - B. 1325
 - C. 1330
 - D. 1335
14. 下列哪种算法属于监督学习？
- A. K-Means 聚类
 - B. 关联规则挖掘
 - C. 决策树分类
 - D. 主成分分析
15. 有观点认为，定期开展数据安全培训能够显著降低企业数据泄露事件的发生率。以下哪项如果为真，最能质疑这一观点？
- A. 已开展数据安全培训的企业普遍同步加强了防火墙等安全设备
 - B. 大多数数据泄露事件是由外部黑客攻击所致，而非内部人员操作失误
 - C. 数据安全培训需要投入时间、资金等成本，会增加企业负担
 - D. 一些企业虽然开展了培训，但仍发生过数据泄露事件
16. 某数据清洗任务，甲每小时处理 100 条，乙每小时处理 150 条，若两人合作 3 小时完成，则总数据量为多少条？
- A. 750
 - B. 800
 - C. 850
 - D. 900
17. 在数据生命周期管理中，最后一个阶段是：
- A. 数据采集
 - B. 数据存储
 - C. 数据销毁
 - D. 数据共享

18. “数据脱敏”与“数据安全”的关系，类似于：
- A. 杀菌消毒：无菌环境
 - B. 供水管道：城市用水
 - C. 排水系统：防洪排涝
 - D. 财务审计：财政监督
19. 在 DAMA-DMBOK 数据管理体系中，数据治理（Data Governance）的核心职责是_____
- A. 对数据管理的决策权与责任体系进行规划、监督和控制
 - B. 将业务系统中的数据通过 ETL 工具抽取到数据仓库
 - C. 使用 Spark 对大规模数据集进行分布式计算
 - D. 通过 BI 报表工具为管理层提供可视化驾驶舱
20. 某数据平台每日新增用户数服从正态分布，平均值为 500，标准差为 50，则某日新增用户数在 450 到 550 之间的概率约为多少？
- A. 68%
 - B. 75%
 - C. 80%
 - D. 95%
21. 根据《中华人民共和国数据安全法》，数据安全工作的首要原则是：
- A. 促进数据开发利用
 - B. 保障数据安全
 - C. 维护数据主权
 - D. 坚持安全与发展并重
22. 只有建立健全数据产权制度，才能充分激活数据要素价值。根据上述命题，以下哪项一定为真？
- A. 如果数据要素价值得到充分激活，说明数据产权制度已经建立健全
 - B. 如果数据产权制度建立健全了，数据要素价值就一定被充分激活
 - C. 数据产权制度未建立健全，但数据要素价值被充分激活
 - D. 建立健全数据产权制度是激活数据要素价值的充分条件
23. 在 SQL 查询中，若要消除查询结果中的重复行，应使用的关键字是_____
- A. DISTINCT
 - B. UNIQUE
 - C. GROUP BY
 - D. ORDER BY
24. 某数据产品定价为每份 50 元，成本为 30 元，若每月销售 1000 份，则月利润为多少元？

- A. 10000
- B. 15000
- C. 20000
- D. 25000

25. 在数据运营中，A/B 测试的主要作用是：

- A. 验证数据质量
- B. 评估算法性能
- C. 比较不同策略的效果
- D. 提升数据存储效率

26. 数据中台是将企业内分散、多源的数据进行统一采集、加工、存储、计算和共享，形成标准化、服务化的数据能力，为前台业务提供敏捷数据支持的平台。根据上述定义，下列不属于数据中台功能的是：

- A. 统一管理各业务系统产生的用户行为数据
- B. 按需生成销售报表并提供 API 接口供前台调用
- C. 搭建公司内部员工招聘与考勤网站
- D. 对来自不同渠道的数据进行清洗和标签化

27. 与经典 MapReduce 相比，Spark 在迭代计算和交互式查询场景下性能明显提升，其主要原因是_____

- A. 基于内存进行数据计算，减少磁盘 I/O
- B. 不支持 SQL 查询，仅支持 Scala 编程
- C. 只能运行在单机环境中
- D. 采用数据复制容错，而非血缘关系

28. 某数据团队有 5 名成员，随机选 2 人参加培训，则选中的 2 人恰好是其中 2 名指定成员的概率为多少？

- A. $1/5$
- B. $1/10$
- C. $1/20$
- D. $1/25$

29. 下列哪项不属于数据资产管理的内容？

- A. 数据标准管理
- B. 数据模型管理
- C. 数据血缘管理
- D. 数据备份管理

30. 某公司对近一年数据运营项目进行复盘，发现应用了机器学习算法的项目，用户留存率平均提升 12%；未应用机器学习算法的项目，用户留存率仅提升 3%。因此，该公司认为机器学习算法有效提升了用户留存率。以下哪项如果为真，最能支持上述结论？
- A. 应用机器学习算法的项目普遍投入了更多运营预算
 - B. 一部分未应用机器学习算法的项目用户留存率出现下滑
 - C. 应用与未应用机器学习算法项目的初始用户基础和市场环境基本一致
 - D. 该公司计划在下一年的所有数据运营项目中推广机器学习算法
31. 在数据预处理中，针对数据集中存在大量缺失值的情况，下列处理方式最符合数据运营实践的是_____
- A. 直接删除含缺失值的所有记录，且无需评估对业务的影响
 - B. 将缺失值全部填充为 0，不再做任何验证
 - C. 根据缺失比例与业务含义，选择删除记录、数值填充或建模预测等方法
 - D. 仅保留无缺失值的字段，删除全部含缺失值的字段
32. 某数据平台每年处理的数据量增长 20%，若今年处理 100 万条，则三年后处理量约为多少万条？
- A. 172.8
 - B. 173.2
 - C. 174.0
 - D. 175.0
33. 在 Spark 中，RDD 的转换操作是惰性的，这意味着：
- A. 转换操作立即执行
 - B. 转换操作不执行
 - C. 转换操作在行动操作时执行
 - D. 转换操作在缓存时执行
34. 下列选项与“数据库：索引：查询”的逻辑关系最为相似的是：
- A. 图书馆：书架：借阅
 - B. 文件：目录：检索
 - C. 网站：网页：浏览
 - D. 仓库：货架：存取
35. 在关联规则挖掘中，规则 A B 的支持度（Support）是指_____
- A. 同时包含 A 和 B 的事务数占事务总数的比例
 - B. 包含 A 的事务数占事务总数的比例
 - C. 包含 A 的事务中同时包含 B 的比例
 - D. 同时包含 A 和 B 的事务数与只包含 A 的事务数的比值

36. 某数据运营团队计划对一批数据进行清洗，若由甲组单独完成需 12 天，乙组单独完成需 18 天。现两队合作，但甲组中途因故停工 3 天，实际完成这批数据清洗共用了多少天？
- A. 7 天
 - B. 8 天
 - C. 9 天
 - D. 10 天
37. 在数据库设计中，将 E-R 图转换为关系模型时，一个多对多联系通常转换为：
- A. 一个独立的关系模式
 - B. 两个关系模式
 - C. 三个关系模式
 - D. 不转换
38. 某单位最终从张、王、李、赵四人中录用一名数据运营专员。招聘结束后，甲、乙、丙、丁四人进行了预测：甲：如果张被录用，那么王被录用。乙：张被录用且王不被录用。丙：李被录用。丁：赵被录用。已知只有一人的预测正确。根据以上信息，可以推出：
- A. 张被录用
 - B. 王被录用
 - C. 李未被录用
 - D. 赵被录用
39. 根据《个人信息保护法》，处理个人信息应当遵循最小必要原则。下列做法符合该原则的是_____
- A. 为提升营销效果，在未告知用户的情况下收集其浏览记录
 - B. 仅收集实现业务功能所必需的个人信息，并告知处理目的、方式和范围
 - C. 将用户个人信息无限期保存，以备后续营销活动使用
 - D. 因合作方要求，直接将其经营所需的个人信息提供给合作方
40. 某数据产品在推广期间，日活跃用户数每天比前一天增长 10%，若第 1 天日活跃用户为 1000 人，则第 5 天日活跃用户约为多少人？
- A. 1464 人
 - B. 1611 人
 - C. 1772 人
 - D. 1949 人
41. 在数据挖掘中，购物篮分析常用的算法是：
- A. Apriori
 - B. K-Means
 - C. SVM

D. C4.5

42. 依据《中华人民共和国数据安全法》，国家建立数据分类分级保护制度。其中，数据分类是根据业务属性将数据划分为不同类别；数据分级是按数据遭到破坏后对国家安全、公共利益或者个人、组织合法权益造成的危害程度，将数据分为核心数据、重要数据和一般数据。根据上述定义，下列属于“数据分级”行为的是：
- A. 某银行将客户数据按“身份信息”“交易流水”“信用记录”三类分别存储
 - B. 某政务平台将数据标注为“可公开”“受限”“敏感”三个等级并实施不同权限管理
 - C. 某电商平台根据用户购物记录划分用户群体并推送个性化广告
 - D. 某医院对全部患者病历统一采用高强度加密存储
43. 在数据生命周期管理中，数据归档与数据销毁的主要区别在于_____
- A. 归档是将数据迁移至低成本存储介质长期保存；销毁是将数据彻底清除使其不可恢复
 - B. 归档是删除无效数据；销毁是将数据压缩存储
 - C. 归档适用于在线交易数据；销毁适用于历史分析数据
 - D. 归档不改变数据的存储位置；销毁改变数据的存储位置
44. 某数据运营团队有 5 名数据分析师和 3 名数据工程师，现需从中选出 3 人组成项目组，要求至少包含 1 名工程师，则不同的选法共有多少种？
- A. 40 种
 - B. 46 种
 - C. 50 种
 - D. 56 种
45. 在数据运营中，用户画像的构建通常不包含以下哪个步骤？
- A. 数据收集
 - B. 标签体系设计
 - C. 用户分群
 - D. 数据加密
46. 某数据运营团队中，甲、乙、丙三人分别负责数据分析、数据运营和数据安全三份工作，他们分别来自重庆、成都、贵阳。已知：（1）负责数据分析的人来自重庆；（2）来自成都的人负责数据运营；（3）甲不来自重庆；（4）乙不来自成都；（5）丙不来自贵阳，且不负责数据分析。根据以上信息，以下哪项一定为真？
- A. 乙负责数据分析
 - B. 乙来自成都
 - C. 丙来自贵阳
 - D. 甲负责数据分析

47. 在 Hadoop 生态中, Hive 的主要作用是_____
- A. 提供类 SQL 查询引擎, 将 SQL 转换为 MapReduce 或 Spark 作业
 - B. 实现分布式文件系统, 存储超大文件
 - C. 提供消息队列服务, 实现数据异步传输
 - D. 进行实时流式计算, 处理毫秒级延迟需求
48. 某数据运营团队计划对一批数据进行清洗, 原计划每天处理 120 万条, 实际每天多处理 30 万条, 结果提前 2 天完成任务。则这批数据共有多少万条?
- A. 1080
 - B. 1200
 - C. 1320
 - D. 1440
49. 在数字经济时代, 数据已成为重要的生产要素。重庆数字资源集团数据运营分公司在推进数据资源化、资产化过程中, 强调“以数据驱动决策, 以运营提升价值”。下列句子中, 没有语病且表意明确的一项是:
- A. 通过数据治理体系的建设, 使该公司数据质量得到了显著提升。
 - B. 该公司不仅注重数据采集的广度, 而且注重数据挖掘的深度。
 - C. 能否有效整合多源数据, 是提升数据运营效率的关键。
 - D. 数据安全是数据运营的底线, 必须防范不出现任何数据泄露事件。
50. 近年来, 某市大数据企业数量快速增长, 但该市核心商务区写字楼租金却保持稳定, 没有明显上涨。以下哪项如果为真, 最能解释这一现象?
- A. 大数据企业大多选择在租金较低的软件园区办公
 - B. 该市核心商务区写字楼空置率在逐年下降
 - C. 该市其他类型企业的办公空间需求也在快速增长
 - D. 大数据企业目前大多处于初创阶段, 员工人数较少
51. 关于 Spark RDD 的惰性求值 (Lazy Evaluation) 机制, 下列说法正确的是_____
- A. 转换算子 (Transformation) 不会立即执行, 只有遇到行动算子 (Action) 时才触发整个 DAG 计算
 - B. RDD 在创建时会立即将全量数据持久化到磁盘
 - C. 惰性求值会导致每个 RDD 被重复计算, 无法通过血缘恢复
 - D. 行动算子只能执行一次, 执行后 RDD 中的数据自动清空
52. 某数据产品每份售价 50 元, 成本为 30 元。若每日销量为 200 份, 每降价 1 元, 日销量增加 20 份。要使日利润最大, 售价应定为多少元?
- A. 40

- B. 42
- C. 45
- D. 48

53. 重庆数字资源集团数据运营分公司致力于构建“数据要素 ×”应用场景，推动数据赋能实体经济。下列语句排序最合理的一项是：（1）数据运营的核心在于挖掘数据价值，而价值实现离不开应用场景的落地。（2）因此，公司积极探索在智能制造、智慧农业等领域的融合应用。（3）通过场景牵引，数据才能从“资源”转变为“资产”，释放乘数效应。（4）这就要求数据运营必须紧扣产业需求，避免“为数据而数据”。

- A. （1）（3）（2）（4）
- B. （2）（1）（4）（3）
- C. （3）（1）（2）（4）
- D. （1）（4）（3）（2）

54. 所有数据运营专员都需要掌握 SQL；有些数据运营专员熟悉 Python；所有熟悉 Python 的人都了解机器学习基础。由此可以推出：

- A. 有些了解机器学习基础的人是数据运营专员
- B. 所有数据运营专员都了解机器学习基础
- C. 有些熟悉 Python 的人不是数据运营专员
- D. 掌握 SQL 的人一定熟悉 Python

55. 在订单表 orders（order_id, customer_id, amount）中，需要查询订单金额高于所有订单平均金额的订单号。下列 SQL 语句正确的是_____

- A. SELECT order_id FROM orders WHERE amount > (SELECT AVG(amount) FROM orders);
- B. SELECT order_id FROM orders HAVING amount > AVG(amount);
- C. SELECT order_id FROM orders WHERE amount > AVG(amount);
- D. SELECT order_id FROM orders GROUP BY order_id WHERE amount > AVG(amount);

56. 某数据运营团队计划对一批历史数据进行清洗，原计划每天清洗 200 万条，实际每天比原计划多清洗 50 万条，结果提前 2 天完成任务。则这批数据共有多少万条？

- A. 1800
- B. 2000
- C. 2200
- D. 2400

57. 国有企业推动数据资产入表时，判断数据资源能否确认为资产负债表中的“资产”，核心依据是_____

- A. 企业拥有或控制该数据资源，与其相关的经济利益很可能流入企业，且成本或价值能够可靠计量

- B. 数据资源已经通过公开渠道免费向公众开放
- C. 数据资源存储在境内自建的机房中
- D. 数据资源已通过第三方数据安全认证

58. 根据《中华人民共和国数据安全法》，重要数据的处理者应当明确数据安全负责人和管理机构，并落实_____

- A. 数据安全保护责任
- B. 数据跨境流动自由
- C. 免费数据开放义务
- D. 数据交易备案制度

二、多选题（共 3 题） 本部分共 3 题。

1. 关于数据挖掘算法的应用场景，下列说法正确的有_____
 - A. 关联规则可用于购物篮分析，发现商品之间的共现关系
 - B. 聚类分析适用于没有类别标签的数据，自动发现数据的内在分组
 - C. 决策树算法主要用于关联规则挖掘
 - D. 时间序列分析可用于识别订单量的趋势、周期和异常波动
2. 在 DAMA-DMBOK 数据管理框架中，下列属于数据管理职能域的有_____
 - A. 数据架构管理
 - B. 元数据管理
 - C. 数据质量管理
 - D. 网络爬虫管理
3. 根据《数据安全法》和《个人信息保护法》，企业开展数据运营时，下列合规做法有_____
 - A. 对数据实行分类分级保护，重要数据加强安全管理
 - B. 处理个人信息前以显著方式、清晰易懂的语言向个人告知并取得同意
 - C. 向境外提供重要数据时，依法进行安全评估
 - D. 因内部管理需要，将个人信息用于与先前告知无关的用途

三、判断题（共 5 题） 本部分共 5 题。

1. 数据生命周期管理仅包括数据采集、存储和删除三个阶段。（ ）
2. 根据《中华人民共和国数据安全法》，国家建立数据分类分级保护制度，对数据实行分类分级保护。（ ）
3. 数据中心电能利用效率（PUE）的计算公式为数据中心总能耗与 IT 设备能耗的比值，PUE 值越接近 1，说明数据中心能效水平越高。（ ）

4. 根据“数据二十条”，我国数据产权制度框架确立为“数据资源持有权、数据加工使用权、数据产品经营权”三权分置。（ ）
5. 数据中心运维中，变更操作必须经过变更申请、影响评估、审批授权、实施与验证等规范化流程，严禁未经评估的随意变更。（ ）

四、资料分析（共 2 题） 本部分共 2 题。

1. 国家统计局与国家能源局发布的电力行业统计数据显示，本次数据采集以全国全口径发电企业为统计对象，涵盖火电、水电、核电、风电及光伏发电等类型，全社会用电量由电网企业抄表汇总，发电装机容量为期末累计装机，非化石能源装机占比指水电、核电、风电、光伏等装机合计占总装机的比例，火电发电量包括燃煤、燃气等机组发电量。统计口径历年保持一致，数据真实可靠，适用于行业趋势分析和政策效果评估。在“双碳”目标引领下，国家先后发布《加快构建新型电力系统行动方案（2024—2027 年）》和《“十四五”现代能源体系规划》，明确要加快能源清洁低碳转型，提高非化石能源消费比重，推动现货市场建设，完善辅助服务补偿机制。2025 年国务院进一步提出深化电力体制改革，促进新能源消纳，这些政策对电力生产结构和消费模式产生深远影响，也为数据运营企业提供了海量电力运行数据的挖掘场景。从总体趋势观察，2020—2025 年全社会用电量由 70000 亿千瓦时增至 98000 亿千瓦时，累计增长 40%，年均增速约 7%，其中 2021 年增长 7.1%，2022 年增长 6.7%，2023 年增长 7.5%，2024 年增长 7.0%，2025 年增长 6.5%，增速保持稳健。发电装机容量从 200000 万千瓦增至 300000 万千瓦，增长 50%，高于用电量增速，为电力保供提供充裕容量。火电发电量从 50000 亿千瓦时增至 57000 亿千瓦时，仅增长 14%，增量贡献明显低于非化石能源，显示电源结构持续优化。结构变化方面，非化石能源装机占比从 40% 提升至 50%，年均提升 2 个百分点，其中风电和光伏贡献最大；火电发电量占比由 71.4% 降至 58.2%，下降 13.2 个百分点。用电侧，第二产业用电量占比由 68% 降至 55%，第三产业和居民生活用电合计占比从 29% 提升至 45%，反映经济结构向服务业转型以及居民电气化水平提高。影响因素包括宏观经济增速放缓后高耗能行业用电低迷、新能源发电出力波动增大系统调峰需求、电动汽车和电采暖等新型负荷快速增长等。表 1 列示 2020—2025 年全社会用电量、发电装机容量、非化石能源装机占比和火电发电量等指标，供分析电力行业运行趋势使用（具体数据见下表）。

年份	全社会用电量 (亿千瓦时)	发电装机容量 (万千瓦)	非化石能源装机占比 (%)	火电发电量 (亿千瓦时)
2020	70000	200000	40	50000
2021	75000	220000	42	52000
2022	80000	240000	44	54000
2023	86000	260000	46	55000
2024	92000	280000	48	56000
2025	98000	300000	50	57000

(1) 2025 年全社会用电量比 2023 年增长约 ()？

- A. 12.5%
- B. 14.0%
- C. 16.2%
- D. 10.8%

(2) 2020—2025 年，发电装机容量年均增加 () 万千瓦？

- A. 18000
- B. 22000
- C. 20000
- D. 24000

(3) 2025 年火电发电量占全社会用电量的比重比上年 ()?

- A. 上升 2.7 个百分点
- B. 下降 2.7 个百分点
- C. 上升 3.2 个百分点
- D. 下降 3.2 个百分点

(4) 2024 年非化石能源装机容量约是 2020 年的 () 倍?

- A. 1.58
- B. 1.78
- C. 1.88
- D. 1.68

(5) 若 2025 年全社会用电量同比增长 6.5%，2024 年同比增长 7.0%，则 2023 年全社会用电量为 () 亿千瓦时?

- A. 86000
- B. 85600
- C. 85800
- D. 86200

2. 国家能源局数据显示，全国全社会用电量统计范围覆盖全国除港澳台以外的所有省份，数据口径为全社会各行业用电量与城乡居民生活用电量之和，不含企业自备电厂自发自用部分，也不包括线损和厂用电。统计周期为公历年度，原始数据来源于各省电网企业月度报送及用电信息采集系统，经国家统计局审核后发布。受宏观经济运行、产业结构调整、极端天气频发等因素影响，用电量波动总体处于合理区间。近年来，在‘双碳’目标引领下，电力行业加快绿色转型，但第二产业仍是用电主体。2024 年国务院印发《2024—2025 年节能降碳行动方案》，明确要求提升非化石能源消费比重，推动重点行业节能增效，并鼓励数据中心、5G 基站等新型基础设施参与绿电交易。2025 年是‘十四五’收官之年，电力需求保持韧性增长，第三产业用电量增速持续快于第二产业，反映服务业数字化、智能化趋势加速。同时，新能源汽车保有量突破 3000 万辆，带动充换电服务业用电量快速增长；人工智能大模型训练、数据中心等算力基础设施用电需求成为新的增长极。从总量看，2023 年全社会用电量为 92241 亿千瓦时，2024 年增至 98521 亿千瓦时，2025 年进一步升至 104600 亿千瓦时，三年间年均增量超过 6000 亿千瓦时。同比增速由 2023 年的 6.7% 小幅升至 2024 年的 6.8%，2025 年回落至 6.2% 左右，整体呈现‘前高后稳’态势，与同期 GDP 增速走势基本吻合，电力消费弹性系数维持在 1.0 以上。结构方面，第二产业用电量占比由 2023 年的 65.85% 逐年下降至 2024 年的 64.83% 和 2025 年的 64.05%；第三产业用电量占比由 18.03% 升至 19.12%，首次接近 20%；城乡居民生活用电量占比稳定在 15% 左右。高技术及装备制造业用电量增速领先，光伏设备及元器件制造用电量增长超 30%，成为新的用电增长点。分区域看，东、中部地区用电量占全国比重略有下降，西部地区因大数据中心集聚用电量占比上升。下表给出了 2023—2025 年全社会用电量及分产业绝对量，供分析使用。

年份	全社会用电量 (亿千瓦时)	第一产业用电 量 (亿千瓦时)	第二产业用电 量 (亿千瓦时)	第三产业用电 量 (亿千瓦时)	城乡居民生活 用电量 (亿千瓦 时)
2023	92241	1278	60745	16635	13583
2024	98521	1357	63874	18397	14893
2025	104600	1400	67000	20000	16200

(1) 2023 年第一产业用电量占全社会用电量的比重约为多少？

- A. 1.4%
- B. 1.6%
- C. 1.2%
- D. 1.8%

(2) 2024 年全社会用电量较上年增长约多少？

- A. 5.2%
- B. 7.6%
- C. 6.8%
- D. 8.3%

(3) 2024 年第二产业用电量占全社会用电量的比重比上年同期约（ ）

- A. 上升 0.5 个百分点
- B. 下降 1.0 个百分点
- C. 下降 0.5 个百分点
- D. 上升 1.0 个百分点

(4) 2025 年第三产业用电量是城乡居民生活用电量的多少倍？

- A. 1.15 倍
- B. 1.23 倍
- C. 1.31 倍
- D. 1.38 倍

(5) 2023—2025 年全社会用电量的年均增长率约为多少？

- A. 5.2%
- B. 6.5%
- C. 7.3%
- D. 8.1%

参考答案与解析

一、单选题（共 58 题）

1. 答案：B。解析：前 4 天完成 40%，则每天完成 10%，剩余 60% 需要 $60\% \div 10\% = 6$ 天。
2. 答案：C。解析：数据质量评估的常见维度包括准确性、完整性、一致性、时效性、唯一性等。安全性虽然重要，但通常属于数据安全管理的范畴，而非数据质量评估的核心维度。
3. 答案：B。解析：题干定义的关键是：制定标准、明确责任、规范流程、全生命周期管理。B 项“制定录入格式和校验规则”体现标准与规范，“定期清理异常数据”体现治理过程，符合定义。A 项只是硬件扩容，C 项是应用开发，D 项是基础汇总，均不属于数据治理。
4. 答案：A。解析：提价后价格为 $200 \times 1.1 = 220$ 元，打八折后为 $220 \times 0.8 = 176$ 元。
5. 答案：A。解析：SQL 中的 SELECT 语句用于从数据库表中查询数据。UPDATE 用于更新数据，DELETE 用于删除数据，INSERT 用于插入数据。
6. 答案：A。解析：大数据是数据的一种，二者是包含关系（种属关系）。A 项核能是能源的一种，也是包含关系，与题干一致。B、C、D 均不是严格的包含关系。
7. 答案：C。解析：设增长率为 r ，则 $100 \times (1+r)^2 = 144$ ，解得 $1+r = 1.2$ ，第二天为 $100 \times 1.2 = 120$ 。
8. 答案：B。解析：HDFS（Hadoop 分布式文件系统）是 Hadoop 的核心存储组件，提供高吞吐量的数据访问。MapReduce 是计算模型，YARN 是资源管理器，Spark 是快速计算引擎。
9. 答案：A。解析：题干给出充分条件假言命题：不执行标准 $\rightarrow$ 质量无法保证。已知“质量得到保证”即否定了后件“质量无法保证”，根据逆否推理，可否定前件，推出“并非不执行标准”，即“执行了数据标准”。故选 A。
10. 答案：A。解析：设原速度为 v ，时间为 t ，距离为 vt 。提速后速度为 $1.2v$ ，新时间为 $vt/(1.2v) = t/1.2 \approx 0.833t$ ，减少 $(1 - 0.833)/1 = 16.7\%$ 。
11. 答案：D。解析：处理缺失值的方法包括删除记录、填充均值/中位数/众数、插值法等。增加噪声通常用于数据增强，而非处理缺失值。
12. 答案：C。解析：数据要素必须参与生产经营并能带来经济效益。A、B、D 均可用于商业分析、服务优化或决策支持，从而创造价值；C 项考勤记录主要用于内部人事管理，通常不直接对外产生经济效益，故不属于数据要素。
13. 答案：A。解析：一次后 $1000 \times 1.15 = 1150$ ，二次后 $1150 \times 1.15 = 1322.5 \approx 1322$ 。
14. 答案：C。解析：决策树分类属于监督学习，需要有标签的训练数据。K-Means 聚类和关联规则挖掘属于无监督学习，主成分分析属于降维技术。
15. 答案：B。解析：题干观点：培训 $\rightarrow$ 降低泄露发生率。B 项说明数据泄露的主要原因是外部攻击，并非内部人员失误，而培训主要针对内部人员，因此培训难以有效降低发生率，直接削弱论点。A 项指出是设备投入而非培训作用，也有削弱效果，但属于“另有他因”，强度弱于 B。C 项无关，D 项不必然否定整体效果。
16. 答案：A。解析：合作效率 $100 + 150 = 250$ 条/小时，3 小时完成 $250 \times 3 = 750$ 条。
17. 答案：C。解析：数据生命周期包括数据采集、存储、使用、共享、归档和销毁等阶段，销毁是数据生命周期的最后环节，确保数据安全合规。
18. 答案：A。解析：数据脱敏是保障数据安全的一种手段，通过脱敏去除敏感信息，达到保护数据安全的目的。A 项杀菌消毒是营造无菌环境的手段，二者都是“行动 $\rightarrow$ 目的”的对应关系。B、C、D 均不属于手段与目的的直接对应。
19. 答案：A。解析：DAMA-DMBOK 将数据治理定义为对数据资产管理的决策权与责任体系进行规划、监督和控制，属于管理性职能，因此选 A。选项 B 属于数据集成技术，选项 C 属于大数据处理技术，选项 D 属于数据可视化应用，均不属于数据治理的核心职责。

20. 答案：A。解析：根据正态分布经验法则，约 68% 的数据落在均值 ± 1 个标准差内，即 450-550。
21. 答案：D。解析：《数据安全法》规定，数据安全工作应当坚持安全与发展并重，既要保障数据安全，又要促进数据开发利用。其他选项均不全面。
22. 答案：A。解析：“只有 P，才 Q”是必要条件假言命题，即 $Q \rightarrow P$ 。价值激活 (Q) $\rightarrow$ 产权制度建立 (P)。A 项符合；B 项将条件关系颠倒 (将“必要条件”误作“充分条件”)，错误；C 项否定前件肯定后件，错误；D 项说法与题干矛盾，题干中 P 是 Q 的必要条件而非充分条件。
23. 答案：A。解析：DISTINCT 用于返回唯一不重复的记录；UNIQUE 是建表时字段唯一性约束，不是查询关键字；GROUP BY 用于分组聚合；ORDER BY 用于排序。
24. 答案：C。解析：每份利润 $50 - 30 = 20$ 元，月利润 $20 \times 1000 = 20000$ 元。
25. 答案：C。解析：A/B 测试通过将用户随机分为两组，分别应用不同策略，以科学评估哪种策略更优，常用于运营策略优化。
26. 答案：C。解析：A、B、D 分别体现了数据中台的数据统一管理、服务化输出、数据加工等核心功能；C 项是人力资源管理信息化，与数据中台无关。
27. 答案：A。解析：Spark 使用 RDD 并优先将数据缓存到内存，迭代时避免反复读写磁盘。B、C 错误；Spark 支持 SQL、Scala、Python、Java 等，可运行在集群；D 错误，Spark 通过 RDD 血缘 (Lineage) 实现容错，而非仅靠复制。
28. 答案：B。解析：总选法 $C(5,2) = 10$ ，指定 2 人只有 1 种选法，概率为 $1/10$ 。
29. 答案：D。解析：数据资产管理包括数据标准、数据模型、数据血缘、数据质量等管理内容。数据备份属于数据存储和容灾范畴，不属于资产管理。
30. 答案：C。解析：题干通过比较得出结论。C 项说明两组项目在初始条件上无差异，排除了“初始用户基础和市场环境”这一混淆因素，使留存率差异更可归因于机器学习算法，从而加强结论。A 项属于另有他因，反而削弱；B 项有一定支持作用，但只涉及部分项目，力度弱于 C；D 项无关。
31. 答案：C。解析：缺失值处理需结合缺失比例、字段重要性和业务场景，可删除、统计填充、插值或建模预测。
32. 答案：A。解析：三年后为 $100 \times (1.2)^3 = 100 \times 1.728 = 172.8$ 万条。
33. 答案：C。解析：RDD 的转换操作是惰性的，即不会立即执行，只有在遇到行动操作 (如 count、collect) 时才会真正触发计算。
34. 答案：B。解析：题干中，索引是数据库中用于加快查询的工具，借助索引可以执行查询。B 项，目录是文件中用于定位的检索工具，借助目录可以检索文件，与题干结构一致。A、D 虽也有对应，但不像 B 那样是“工具-检索目的”的紧密对应。
35. 答案：A。解析：支持度 $\text{Support}(A B)$ 定义为同时包含 A 和 B 的事务数占事务总数的比例，即 $\text{Support}(A B) = P(A \cap B)$ ，因此选 A。选项 C 是置信度 $\text{Confidence}(A B) = P(B|A) = \text{包含 A 和 B 的事务数} / \text{包含 A 的事务数}$ ；选项 B 是 A 的支持度；选项 D 无对应定义。
36. 答案：C。解析：设工作总量为 36 (12 和 18 的最小公倍数)，则甲效率为 $36 \div 12 = 3$ ，乙效率为 $36 \div 18 = 2$ 。甲停工 3 天，相当于乙单独工作 3 天，完成 $2 \times 3 = 6$ ，剩余工作量为 $36 - 6 = 30$ 。剩余工作量由两团队合作完成，合作效率为 $3 + 2 = 5$ ，所需时间为 $30 \div 5 = 6$ 天。总时间 $= 3 + 6 = 9$ 天。
37. 答案：A。解析：E-R 图中多对多联系转换为关系模型时，需要创建一个独立的关系模式，包含两个实体的主键以及联系自身的属性。
38. 答案：C。解析：甲的话 (张 $\rightarrow$ 王) 与乙的话 (张 $\neg$ 王) 互为矛盾关系，必有一真一假。由于只有一人的预测正确，则丙、丁的话均为假。丙说“李被录用”为假，所以李未被录用；丁说“赵被录用”为假，所以赵未被录用。故 C 项一定为真。
39. 答案：B。解析：最小必要原则要求处理个人信息应限于实现目的的最小范围，不得过度收集，

且需告知同意。A 未告知；C 超期保存；D 未经同意向第三方提供均不符合规定。

40. 答案：A。解析：第 5 天日活跃用户为 $1000 \times (1+10\%)^4 = 1000 \times 1.1^4 = 1000 \times 1.4641 = 1464.1$ ，约 1464 人。

41. 答案：A。解析：购物篮分析用于发现商品之间的关联规则，Apriori 算法是经典的关联规则挖掘算法。K-Means 是聚类算法，SVM 是分类算法，C4.5 是决策树算法。

42. 答案：B。解析：A 项是按业务属性分类，属于数据分类；B 项按敏感性和危害程度划分不同等级并实施差异权限，属于数据分级；C 项是用户画像应用；D 项是统一安全加固措施，未体现分级。

43. 答案：A。解析：归档针对不常用但需长期保存的数据，迁移到磁带、冷存储等低成本介质；销毁是物理或逻辑删除，确保数据不可恢复。B、C、D 表述不正确。

44. 答案：B。解析：从 8 人中任选 3 人的总选法为 $C(8,3) = 56$ 种。不含工程师即全为数据分析师的选法为 $C(5,3) = 10$ 种。因此至少包含 1 名工程师的选法为 $56 - 10 = 46$ 种，选 B。

45. 答案：D。解析：用户画像构建包括数据收集、标签体系设计、用户分群等步骤。数据加密属于数据安全措施，不属于用户画像构建的步骤。

46. 答案：A。解析：由（1）（2）可知，数据分析对丙不负责数据分析且不来自贵阳，所以丙只能来自成都，负责数据运营。成都已被丙占据，乙不来自成都，则乙只能来自重庆或贵阳；又甲不来自重庆，所以甲来自贵阳，负责数据安全，乙来自重庆，负责数据分析。故 A 项一定为真。

47. 答案：A。解析：Hive 是数据仓库工具，通过 HQL 解析为分布式计算作业；HDFS 负责存储；Kafka 是消息队列；Flink 等适合实时流计算。

48. 答案：B。解析：设计划用 x 天完成，则数据总量为 $120x$ 万条。实际每天处理 150 万条，用了 $(x-2)$ 天，故 $120x = 150(x-2)$ ，解得 $120x = 150x - 300$ ， $30x = 300$ ， $x = 10$ 。所以数据总量为 $120 \times 10 = 1200$ 万条。

49. 答案：B。解析：A 项成分残缺，缺少主语，应删去“通过”或“使”；C 项两面对一面，应删去“能否”；D 项否定不当，“防范不出现”应改为“防范出现”或“确保不出现”；B 项关联词使用正确，语序得当，无语病。

50. 答案：A。解析：题干矛盾在于：大数据企业增多（理应推高写字楼需求）但核心商务区租金未涨。A 项说明大数据企业的办公需求并不集中于核心商务区，而是流向软件园区，因此不会对核心商务区租金产生明显影响，解释了矛盾。B 项空置率下降意味着需求上升，应推动租金上涨，反而加剧矛盾；C 项其他企业需求增长也应推高租金；D 项员工工少不足以完全说明企业数量增长带来的影响，解释力度较弱。

51. 答案：A。解析：Spark 的 Transformation 记录血缘，惰性执行；Action 触发计算。B 错误，RDD 创建时不立即持久化；C 错误，血缘用于容错重算；D 错误，RDD 可被多个作业重复使用。

52. 答案：C。解析：设降价 x 元，则售价为 $(50-x)$ 元，销量为 $(200+20x)$ 份。日利润 $y = (50-x-30)(200+20x) = (20-x)(200+20x) = 4000 + 200x - 20x^2$ ，二次函数开口向下，顶点 $x = -200/(2 \times (-20)) = 5$ 。故降价 5 元，售价为 45 元。

53. 答案：D。解析：首句应提出核心观点，（1）句提出“数据运营核心在于挖掘数据价值，价值实现离不开应用场景”，适合作为首句。（4）句承接（1）句，说明“这就要求数据运营紧扣产业需求”，（3）句进一步阐述“通过场景牵引，数据才能从资源转变为资产”，（2）句以“因此”总结，提出公司探索具体应用。故正确排序为（1）（4）（3）（2），对应 D 项。

54. 答案：A。解析：由“所有熟悉 Python 的人都了解机器学习基础”和“有些数据运营专员熟悉 Python”，可推出“有些数据运营专员了解机器学习基础”。根据直言命题的换位推理，“有些 S 是 P”等价于“有些 P 是 S”，因此可推出“有些了解机器学习基础的人是数据运营专员”。选项 A 正确。B 项“所有”过于绝对，无法推出；C 项“有些熟悉 Python 的人不是数据运营专员”与题干不矛盾但无法推

出；D 项将“所有数据运营专员都需要掌握 SQL”错误地反推为“掌握 SQL 的人一定熟悉 Python”，不符合三段论规则。

55. 答案：A。解析：子查询求全体平均金额后再比较，语法正确。聚合函数不能直接出现在 WHERE 中，除非使用 HAVING 与 GROUP BY 组合，但 B、D 语法或逻辑均不正确。

56. 答案：B。解析：设原计划需要 x 天完成，则数据总量为 $200x$ 万条。实际每天清洗 250 万条，用了 $(x-2)$ 天，所以 $200x = 250(x-2)$ 。解得 $200x = 250x - 500$ ，即 $50x = 500$ ， $x = 10$ 。因此数据总量为 $200 \times 10 = 2000$ 万条。

57. 答案：A。解析：资产确认需满足可控制、经济利益很可能流入、成本价值可可靠计量三个基本条件（参照《企业数据资源相关会计处理暂行规定》基本思路）。开放共享、存储位置、安全认证都不是充分条件。

58. 答案：A。解析：《数据安全法》明确，重要数据处理者应当明确数据安全负责人和管理机构，落实数据安全保护责任。B、C、D 与法条内容不符。

二、多选题（共 3 题）

1. 答案：ABD。解析：A 正确，关联规则典型应用是购物篮分析；B 正确，聚类是无监督学习；C 错误，决策树主要用于分类和回归，而非关联规则挖掘；D 正确，时间序列分析适合趋势、季节性与异常检测。

2. 答案：ABC。解析：DAMA-DMBOK 包括数据架构、数据质量、元数据、主数据、数据仓库等职能域；网络爬虫是数据采集技术工具，不属于 DAMA 管理职能域。

3. 答案：ABC。解析：A 符合数据分类分级要求；B 符合告知同意规则；C 符合数据出境安全评估规定；D 违反个人信息处理目的限制，

三、判断题（共 5 题）

1. 答案：错误。解析：数据生命周期管理涵盖数据采集、存储、使用、共享、归档、销毁等全流程环节，而不仅仅是采集、存储和删除。数据在使用和共享阶段同样需要管控，因此题干表述不完整。

2. 答案：正确。解析：《数据安全法》第二十一条明确规定，国家建立数据分类分级保护制度，根据数据在经济社会发展中的重要程度，以及一旦遭到篡改、破坏、泄露或者非法获取、非法利用，对国家安全、公共利益或者个人、组织合法权益造成的危害程度，对数据实行分类分级保护。

3. 答案：正确。解析： $PUE = \text{数据中心总能耗} \div \text{IT 设备能耗}$ ，是衡量数据中心能效的核心指标。PUE 值越接近 1，表示非 IT 设备（如制冷、配电等）的能耗占比越低，能效水平越高，因此题干表述正确。

4. 答案：正确。解析：《中共中央国务院关于构建数据基础制度更好发挥数据要素作用的意见》（即“数据二十条”）提出，建立数据资源持有权、数据加工使用权、数据产品经营权三权分置的产权运行机制，这是我国数据要素基础制度的重要创新。

5. 答案：正确。解析：数据中心基础设施与 IT 系统变更可能影响业务连续性，必须遵循变更管理流程，通过评估影响、审批授权、实施验证等环节控制风险，确保运维安全。题干表述符合行业最佳实践。

四、资料分析（共 2 题）

1. 答案：B；C；B；D；A

(1) B。 $(98000 - 86000) \div 86000 = 12000 \div 86000 \approx 13.95\% \approx 14.0\%$ 。

(2) C。 $(300000 - 200000) \div 5 = 100000 \div 5 = 20000$ 万千瓦。

(3) B。2025 年比重 = $57000 \div 98000 \approx 58.16\%$ ，2024 年比重 = $56000 \div 92000 \approx 60.87\%$ ， $58.16\% - 60.87\% = -2.71$ 个百分点，即下降 2.7 个百分点。

(4) D。2024 年非化石能源装机容量 = $280000 \times 48\% = 134400$ 万千瓦；2020 年 = $200000 \times 40\% =$

80000 万千瓦； $134400 \div 80000 = 1.68$ 。

(5) A。间隔增长率 $= 6.5\% + 7.0\% + 6.5\% \times 7.0\% = 13.5\% + 0.455\% = 13.955\%$ ；2023 年 $= 98000 \div (1 + 13.955\%) \approx 98000 \div 1.13955 \approx 86002$ 亿千瓦时，约 86000。

2. 答案：A；C；B；B；B

(1) A。2023 年第一产业用电量为 1278 亿千瓦时，全社会用电量为 92241 亿千瓦时，比重 $= 1278 \div 92241 \approx 0.01386 = 1.386\%$ ，与 1.4% 最接近，选 A。

(2) C。增长率 $= (98521 - 92241) \div 92241 = 6280 \div 92241 \approx 0.0681 = 6.8\%$ ，选 C。

(3) B。2023 年第二产业占比 $= 60745 \div 92241 \approx 65.85\%$ ，2024 年第二产业占比 $= 63874 \div 98521 \approx 64.83\%$ ，比重变化 $= 64.83\% - 65.85\% = -1.02$ 个百分点，即下降约 1.0 个百分点，选 B。

(4) B。倍数 $= 20000 \div 16200 \approx 1.2346$ ，即约 1.23 倍，选 B。

(5) B。设年均增长率为 r ，则 $(1 + r)^2 = 104600 \div 92241 \approx 1.1340$ ，开平方得 $1 + r \approx 1.0650$ ，故 $r \approx 6.50\%$ ，选 B。