

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 67 题，考试时间 120 分钟

一、单选题（共 57 题） 本部分共 57 题。

1. 某数据公司从 A 地到 B 地传输数据，去程速度为 80 Mbps，回程速度为 120 Mbps，求全程平均速度（单位：Mbps）。
A. 96
B. 100
C. 104
D. 108
2. 近年来，随着数字经济的蓬勃发展，数据已成为重要的生产要素。重庆数字资源集团作为市属重点企业，正积极推进数据要素市场化配置改革。下列对“数据要素”理解不准确的是：
A. 数据要素是数字经济时代的基础性战略资源
B. 数据要素具有非竞争性和部分排他性，可被多次复制使用
C. 数据要素的价值实现依赖于数据的确权、流通和交易
D. 数据要素与其他传统生产要素一样，具有完全排他性
3. 数据治理是指对数据资产进行管理和控制的一系列活动，旨在提升数据的质量、可用性和安全性。根据上述定义，下列属于数据治理的是：
A. 某互联网公司利用用户浏览数据优化推荐算法
B. 某银行对客户交易数据进行脱敏处理并建立访问权限
C. 某研究团队对收集到的问卷数据进行统计分析
D. 某新媒体编辑对文章进行校对和排版
4. 在企业数据治理实践中，下列哪项是数据治理最核心的业务目标？
A. 降低数据存储硬件的采购成本
B. 提升数据的准确性、一致性与可用性
C. 扩大数据中心机房的机柜数量
D. 减少数据备份所需的磁带数量
5. 某数据产品原价 200 元，先提价 20%，再打八折出售，则现价为多少元？
A. 192
B. 200
C. 208

D. 224

6. 下列句子中，没有语病的一项是：

- A. 通过这次数据治理培训，使员工们深刻认识到数据质量的重要性。
- B. 集团要求各子公司必须建立健全数据安全管理体系，防止数据泄露事件不再发生。
- C. 数据运营分公司凭借其先进的技术和优质的服务，赢得了客户的广泛赞誉。
- D. 能否提升数据处理能力，是推动企业数字化转型的关键。

7. 某数据公司推出了一套智能客服系统，运行三个月后，客户投诉量下降了 30%。因此，该公司认为该套系统有效提升了客户服务质量。以下哪项如果为真，最能削弱上述结论？

- A. 该公司同期还优化了客服流程和人员培训
- B. 智能客服系统仅覆盖了少部分客户
- C. 客户投诉量的统计口径在系统推出前后发生了变化
- D. 竞争对手也在推广类似智能客服系统

8. 在数据采集与预处理流程中，ETL 工具中的“T”指的是哪种操作？

- A. Transfer（传输）
- B. Transform（转换）
- C. Test（测试）
- D. Trigger（触发）

9. 从 1 到 100 的整数中随机抽取一个数，抽到的数能被 3 或 5 整除的概率是多少？

- A. 0.32
- B. 0.47
- C. 0.53
- D. 0.68

10. 将下列句子重新排列，语序最连贯的一项是：（1）数据作为新型生产要素，正深刻改变着生产方式和社会治理方式。（2）重庆数字资源集团数据运营分公司应运而生，致力于挖掘数据价值。（3）近年来，大数据、人工智能等技术的飞速发展，使数据量呈爆炸式增长。（4）在此背景下，如何高效运营数据成为亟待解决的问题。

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

11. 数据要素是指参与社会生产经营活动、能为使用者带来经济效益的数据资源。根据上述定义，下列属于数据要素的是：

- A. 某公司内部员工考勤记录

- B. 某景区拍摄的游客照片
C. 某电商平台的用户购买记录
D. 某学校保管的学生档案
12. Hadoop 生态中，HDFS（Hadoop 分布式文件系统）默认将大文件切分为多个数据块分布存储。在 Hadoop 3.x 版本中，默认数据块大小是？
A. 64MB
B. 128MB
C. 256MB
D. 512MB
13. 2, 3, 5, 9, 17, ?
A. 31
B. 33
C. 35
D. 37
14. 下列各句中，加点成语使用恰当的一项是：
A. 在数据安全问题，我们不能掉以轻心，必须未雨绸缪，防患于未然。
B. 他做事总是首鼠两端，果断决策，深受领导赏识。
C. 面对复杂的数据分析任务，他胸有成竹，结果却一败涂地。
D. 集团内部数据孤岛现象严重，各部门各自为政，导致资源浪费。
15. 只有建立数据安全管理制度，才能有效防范数据泄露风险。企业甲建立了数据安全管理制度。由此可以推出：
A. 甲企业一定不会发生数据泄露
B. 甲企业可能发生数据泄露
C. 甲企业必然能有效防范数据泄露
D. 甲企业的数据泄露风险一定降低
16. 某数据运营团队在 MySQL 中有一张用户订单表（order），需要查询每个用户的订单总金额（按 user_id 分组），并只显示总金额大于 1000 的用户。下列 SQL 语句正确的是？
A. `SELECT user_id, SUM(amount) FROM order WHERE SUM(amount) > 1000 GROUP BY user_id;`
B. `SELECT user_id, SUM(amount) FROM order GROUP BY user_id HAVING SUM(amount) > 1000;`
C. `SELECT user_id, SUM(amount) FROM order GROUP BY amount HAVING SUM(amount) > 1000;`

D. `SELECT user_id, SUM(amount) FROM order WHERE amount > 1000 GROUP BY user_id;`

17. 某数据中心的服务器数量每年增长 20%，若 2024 年底有 500 台，则 2027 年底有多少台？（按整年计算）

- A. 864
- B. 900
- C. 936
- D. 972

18. 下列句子中，表达最得体的一项是：

- A. 对于贵公司的数据合作项目，我们深感兴趣，希望尽快惠赐佳音。
- B. 感谢您的热情款待，我们一定不负众望，努力完成任务。
- C. 您提出的宝贵意见，我们将认真考虑，并尽快改进。
- D. 请您放心，我们保证按时完成数据迁移工作，绝不拖延。

19. 某公司数据部有甲、乙、丙、丁四个小组，针对今年集团数据治理专项检查结果，四个小组长有如下预测：甲：所有小组都通过了检查。乙：丙组通过了检查。丙：甲组没有通过检查。丁：有的小组没有通过检查。事后证明，四个预测中只有一项为真。根据以上信息，以下哪项一定为真？

- A. 甲组通过了
- B. 乙组通过了
- C. 丙组通过了
- D. 所有小组都通过了

20. 与 Hadoop MapReduce 相比，Apache Spark 在大数据处理中的主要优势是？

- A. 无需编程即可完成所有数据分析
- B. 基于内存的计算模型，减少磁盘读写开销
- C. 只能处理流式数据，不能处理批量数据
- D. 内置图形化操作界面，无需命令行

21. 一项数据标注任务，甲单独做需 12 天完成，乙单独做需 18 天完成，两人合作 3 天后，甲因故离开，剩余由乙单独完成，还需多少天？

- A. 10.5
- B. 11
- C. 11.5
- D. 12

22. 下列语句中，标点符号使用正确的一项是：

- A. 数据运营的核心包括：数据采集、数据清洗、数据存储、以及数据分析。

- B. 他问：“数据安全法》何时正式实施？”
- C. 重庆数字资源集团数据运营分公司的主要业务有：数据治理、数据服务、数据交易等。
- D. 数据要素的市场化配置，需要政府、企业、社会各方共同努力。
23. 数据脱敏是指对敏感数据进行变形处理，在不改变数据特征的前提下，消除其与真实身份或敏感信息的关联。根据上述定义，下列属于数据脱敏的是：
- A. 某系统将用户的身份证号统一替换为随机生成的编号
- B. 某统计部门将全国人口数据汇总后发布
- C. 某公司将员工薪资数据设置访问权限
- D. 某医院将患者姓名用‘患者 A’代替，但其病历号保留
24. 在数据预处理过程中，对于某连续型字段存在少量缺失值，下列哪种处理方法最不合理？
- A. 使用该字段的均值进行填充
- B. 删除存在缺失值的整条记录
- C. 使用回归模型基于其他字段预测填充
- D. 为缺失值添加一个远大于正常范围的常数
25. 某数据报表中，某指标本月值为 240，上月为 200，环比增长率是多少？
- A. 20%
- B. 16.7%
- C. 25%
- D. 40%
26. 下列句子中，没有歧义的一项是：
- A. 我们需要了解数据运营分公司今年的目标。
- B. 他借了集团的数据分析报告。
- C. 数据安全是每个员工的责任。
- D. 新来的数据管理员和经理谈了很久。
27. 某数据分析公司拟推行‘数据咖啡’项目，即在办公区设置自助咖啡机，通过分析员工咖啡消费数据来优化采购和排班。他们认为，只要收集员工购买记录，就能提升咖啡机使用效率。以下哪项是上述论证必须假设的？
- A. 员工愿意提供真实的个人信息
- B. 咖啡机使用效率可通过数据优化提升
- C. 公司有权收集员工消费数据
- D. 员工喝咖啡的数量与工作效率有关
28. 某 App 的 DAU 为 6000 万，MAU 为 1.2 亿，则该产品的用户粘性指标（DAU/MAU）约为？

- A. 20%
- B. 50%
- C. 100%
- D. 200%
29. 某公司数据存储成本为每月固定 5000 元，另加每 GB 0.5 元。若本月存储量为 8000 GB，则总成本为多少元？
- A. 9000
- B. 8500
- C. 8000
- D. 7500
30. 下列各句中，没有语病且句意明确的一项是：
- A. 通过实施数据资源普查，集团摸清了数据家底，为后续数据运营奠定了基础。
- B. 集团是否能够抓住数字经济发展机遇，关键在于人才队伍建设。
- C. 数据运营分公司成立以来，一直致力于提升数据服务能力，取得了显著成绩。
- D. 随着数据安全法的实施，企业对数据合规的重视程度有了明显提高。
31. 某研究机构调查了 1500 名大学生，发现每周运动 2 次以上者期末考试平均成绩高于运动不足者。因此，运动有助于提高学习成绩。以下哪项如果为真，最能支持上述结论？
- A. 调查对象涵盖不同专业和年级
- B. 运动后大脑血液循环加快，有助于记忆
- C. 成绩好的学生通常更有计划地安排运动
- D. 调查排除了因身体原因无法运动的学生
32. 在数据仓库维度建模中，星型模型由一张事实表和多张维度表组成。下列描述正确的是？
- A. 事实表存储描述性文字信息，维度表存储数值度量
- B. 维度表通过外键与事实表关联，事实表包含业务过程的度量值
- C. 星型模型中维度表可以继续规范化拆分为多层雪花结构
- D. 事实表通常不包含任何外键，仅包含度量列
33. 1, 4, 9, 16, 25, ?
- A. 36
- B. 49
- C. 64
- D. 81
34. 将下列句子组成一段连贯的话，排列最合理的一项是：（1）数据交易是数据要素市场化配置的重要环节。（2）通过数据交易，可以实现数据价值的变现和增值。（3）然而，数据交易也面临确权难、定价难、监管难等挑战。（4）因此，需要建立完善的数据交易规则和监管机制。

- A. (1) (2) (3) (4)
B. (1) (3) (2) (4)
C. (2) (1) (3) (4)
D. (2) (3) (1) (4)
35. 数据运营是指通过数据采集、分析和应用，对产品或业务进行持续优化，以实现业务目标的过程。根据上述定义，下列属于数据运营活动的是：
- A. 某公司为保障数据安全，定期对服务器进行漏洞扫描
B. 某电商平台通过分析用户行为数据，优化商品推荐策略
C. 某企业按照监管部门要求，编制数据合规报告
D. 某高校开发了一套数据仓库系统用于存储科研数据
36. 下列哪一项通常不属于数据质量评估的常用维度？
- A. 准确性
B. 完整性
C. 可扩展性
D. 一致性
37. 某数据团队有 10 人，其中 6 人会 Python，5 人会 SQL，3 人两者都会，则两者都不会的有几人？
- A. 2
B. 3
C. 4
D. 5
38. 下列句子中，表达最准确、简洁的一项是：
- A. 由于数据泄露事件的发生，导致公司声誉受到了严重影响。
B. 公司决定对数据安全管理制度进行修订和完善，以适应当前形势。
C. 关于数据运营的课程，员工们普遍反映良好。
D. 集团要求各子公司必须建立健全数据安全应急响应机制，确保数据安全。
39. 某城市推行‘数字治堵’后，交通拥堵指数下降，但市民平均通勤时间却增加了。以下哪项最能解释这一矛盾？
- A. 市民更倾向于自驾出行，导致道路车流量增加
B. 该城市通勤距离普遍较长
C. 数据治堵系统优化了信号灯，车辆行驶速度提高，但市民因寻找停车位而增加时间
D. 交通拥堵指数统计范围不包括远郊
40. 在数据生命周期管理中，企业应遵循“最小够用、最短留存”原则。数据生命周期通常不包括下列哪个阶段？

- A. 数据采集与录入
 - B. 数据存储与维护
 - C. 数据无限期永久保存
 - D. 数据共享与销毁
41. 某数据产品利润率为 25%，若成本降低 20%，售价不变，则新的利润率是多少？
- A. 50%
 - B. 56.25%
 - C. 60%
 - D. 66.7%
42. 下列各句中，没有语病的一项是：
- A. 数据运营分公司通过技术创新，提高了数据处理效率，降低了运营成本。
 - B. 能否有效保护数据隐私，是数据产业健康发展的关键。
 - C. 集团领导深入基层，了解数据运营情况，解决存在的问题。
 - D. 随着数字化进程的加快，使传统行业面临着转型升级的压力。
43. 某公司计划投入一笔经费用于新建数据中心，因为公司高管认为，增加服务器数量就能提升数据处理能力。以下哪项指出了上述论证的缺陷？
- A. 将数据处理能力等同于服务器数量
 - B. 忽略了其他影响数据处理能力的因素
 - C. 没有说明经费的具体数额
 - D. 对未来数据处理需求估计不足
44. 根据财政部 2023 年发布的《企业数据资源相关会计处理暂行规定》，企业将内部使用的数据资源确认为无形资产时，应满足的条件之一是？
- A. 该数据资源与生产经营活动无关
 - B. 该数据资源的使用寿命可以可靠估计
 - C. 该数据资源在当期已全部计入费用
 - D. 该数据资源无任何实际应用场景
45. 某数据包传输需要经过 3 个节点，每个节点传输成功的概率分别为 0.9、0.8、0.7，且相互独立，则整个传输成功的概率是多少？
- A. 0.504
 - B. 0.576
 - C. 0.648
 - D. 0.72
46. 下列句子中，标点符号使用正确的一项是：

- A. 数据运营涉及多个环节：采集、清洗、存储、分析、应用等。
- B. 他问我：“数据资产入表了吗？”我说：“还没有。”
- C. 重庆数字资源集团数据运营分公司成立于 2024 年，是集团的核心子公司。
- D. 数据安全法、个人信息保护法等法律法规，为数据运营提供了法律保障。
47. 数据资产是指由企业拥有或者控制的、能够为企业带来未来经济利益的、以物理或电子的方式记录的数据资源。根据上述定义，下列不属于数据资产的是：
- A. 某电商平台收集的用户浏览记录数据库
- B. 某医院的患者电子病历系统存储的病例信息
- C. 某市气象局公开的历年气象监测数据
- D. 某物流公司车辆的实时 GPS 定位数据
48. 某电商数据运营团队对 2025 年 6 月首次下单的新用户建立同期群，并跟踪其后续每月复购率。这种同期群分析（Cohort Analysis）的主要目的是？
- A. 比较不同时段获取的新用户在后续行为上的差异
- B. 预测整个平台下一季度的总销售额
- C. 识别高价值商品的最佳定价区间
- D. 评估不同广告素材的点击率高低
49. 2, 6, 12, 20, 30, ?
- A. 42
- B. 44
- C. 46
- D. 48
50. 下列句子中，没有语病且句意明确的一项是：
- A. 集团要求各子公司必须建立健全数据安全管理制度，防止数据泄露事件不再发生。
- B. 通过数据治理，使数据质量得到了显著提升。
- C. 数据运营分公司致力于为客户提供高效、安全、可靠的数据服务。
- D. 能否提高数据处理能力，是衡量企业数字化水平的重要标准。
51. 一项关于企业数字化转型的研究发现，在引入大数据分析系统后，企业的平均库存周转率提升了 20%。因此研究者认为，大数据分析系统能够有效帮助企业降低库存成本。以下哪项如果为真，最能削弱上述结论？
- A. 引入大数据分析系统的企业同时也优化了供应链管理流程
- B. 未引入大数据分析系统的企业库存周转率也有所提升
- C. 有些企业引入大数据分析系统后库存成本反而上升
- D. 该研究选取的企业均为行业领军企业，不具有普遍性

52. 某数据公司 2025 年第一季度营收为 1200 万元，第二季度比第一季度增长 15%，第三季度比第二季度下降 10%，则第三季度营收为多少万元？
- A. 1242
B. 1260
C. 1280
D. 1300
53. 下列各句中，表达最得体的一项是：
- A. 对于您的批评，我们虚心接受，并立即整改。
B. 请您务必在三天内完成数据报告，否则后果自负。
C. 我们诚挚邀请您参加数据要素论坛，届时敬请光临。
D. 您提出的问题非常幼稚，我们不予回答。
54. 某数据运营团队就“数据平台能否按时上线”进行讨论，甲：数据平台能按时上线。乙：如果数据清洗工作按时完成，数据平台就能按时上线。丙：数据平台不能按时上线。丁：数据清洗工作没有按时完成，但数据平台能按时上线。已知四人中只有一人说假话。由此可以推出：
- A. 数据平台按时上线了
B. 数据清洗工作按时完成了
C. 甲和丙的预测都是真的
D. 乙和丁的预测均为假
55. 某数据运营团队计划对一批数据资产进行清洗，原计划每天处理 120 万条记录，实际每天比计划多处理 30 万条，结果提前 2 天完成任务。则这批数据资产共有多少万条记录？
- A. 960
B. 1080
C. 1200
D. 1440
56. 下列句子中，没有语病的一项是：
- A. 数据运营分公司通过大数据分析，精准识别了客户需求，提升了服务质量。
B. 由于数据管理不善，导致了数据质量下降，影响了业务决策。
C. 集团领导对数据运营工作给予了高度评价，并提出了宝贵意见。
D. 随着数据要素市场的不断完善，使数据交易更加活跃。
57. 某数据运营公司规定，凡是通过数据安全审计的系统，都必须实施访问控制。所有实施访问控制的系统都部署了日志审计模块。有些部署了日志审计模块的系统没有通过数据安全审计。根据以上陈述，以下哪项一定为真？
- A. 有些没有通过数据安全审计的系统没有部署日志审计模块

- B. 所有通过数据安全审计的系统都部署了日志审计模块
- C. 有些部署了日志审计模块的系统通过了数据安全审计
- D. 所有实施访问控制的系统都通过了数据安全审计

二、多选题（共 3 题） 本部分共 3 题。

1. 下列数据挖掘算法中，属于无监督学习的有？
 - A. K-means 聚类算法
 - B. 决策树算法（C4.5）
 - C. Apriori 关联规则算法
 - D. 线性回归算法
2. 我国《数据管理能力成熟度评估模型》（GB/T 36073-2018，简称 DCMM）定义了数据管理能力域。下列属于 DCMM 能力域的有？
 - A. 数据战略
 - B. 数据治理
 - C. 数据架构
 - D. 数据应用
3. 根据《中华人民共和国个人信息保护法》，处理个人信息应当遵循的原则包括？
 - A. 合法
 - B. 正当
 - C. 必要
 - D. 诚信

三、判断题（共 5 题） 本部分共 5 题。

1. 数据生命周期管理通常包括数据采集、数据存储、数据使用、数据共享、数据归档和数据销毁等阶段。（ ）
2. 根据《中华人民共和国个人信息保护法》，处理个人信息应当遵循最小必要原则，即收集个人信息应当限于实现处理目的的最小范围，不得过度收集个人信息。（ ）
3. 数据中心运营中，定期进行数据备份是保障数据安全的重要措施，备份数据应当存储在异地或不同介质上，以防止本地灾难导致数据丢失。（ ）
4. 在数据资产管理中，数据资产确权仅涉及数据的使用权，不涉及数据的所有权和收益权。（ ）
5. 根据《关于构建数据基础制度更好发挥数据要素作用的意见》（数据二十条），我国数据产权制度实行“三权分置”，即数据资源持有权、数据加工使用权和数据产品经营权。（ ）

四、资料分析（共 2 题） 本部分共 2 题。

1. 国家能源局发布 2025 年全国电力工业统计数据显示，2023—2025 年，全国全社会用电量由 92241 亿千瓦时增加至 105000 亿千瓦时。全社会用电量统计范围包含农林牧渔业、工业、建筑业、交通运输业、服务业和城乡居民生活用电，采用用电终端采集系统汇总，剔除了线路损耗。为反映产业结构调整成效，本文按第一、二、三产业及城乡居民生活用电四类划分。在碳达峰碳中和目标引领下，我国深入实施节能降碳和电能替代战略，2024 年、2025 年全社会用电量同比分别增长 6.8% 和 6.6%。政策推动第三产业数字化、智能化加速，数据中心、5G 基站等新基建用电需求增长较快。总体看，用电量增速保持在 6.5%—7.0% 区间，高于 GDP 平均增速，反映电气化水平持续提升。第二产业用电量仍是“压舱石”，但占比由 2023 年的 65.9% 下降至 2025 年的 64.1%；第三产业占比由 18.1% 提升至 19.1%，居民生活用电占比稳定在 15% 左右。结构变化上，高技术制造业用电量增长较快，传统高耗能行业用电量增幅收窄；受城镇化率提高和居民收入增长影响，居民生活用电保持 8% 左右的增速。分城乡看，2025 年居民生活用电中城镇居民占 75%，城镇居民用电量同比增长 9.2%，农村居民用电量增速相对较低。表 1 展示了 2023—2025 年全社会及分产业用电量数据（具体数据见下表）。

年份	全社会用电量 (亿千瓦时)	第一产业 (亿千 瓦时)	第二产业 (亿千 瓦时)	第三产业 (亿千 瓦时)	居民生活用电 量 (亿千瓦时)
2023	92241	1278	60745	16694	13524
2024	98521	1357	63874	18348	14942
2025	105000	1450	67300	20100	16150

- (1) 根据材料，2025 年全社会用电量较 2023 年增长约多少？
- A. 12.8%
B. 14.5%
C. 13.8%
D. 15.2%
- (2) 2025 年第三产业用电量占全社会用电量的比重较 2024 年约提高了多少个百分点？
- A. 0.3
B. 0.5
C. 0.7
D. 0.9
- (3) 2025 年农村居民用电量同比增长约多少？
- A. 4.7%
B. 3.8%
C. 5.6%
D. 6.5%
- (4) 2023—2025 年第二产业用电量年均增长率约为多少？
- A. 5.2%
B. 4.8%
C. 5.8%
D. 6.3%
- (5) 2023—2025 年，第三产业用电量增长量对全社会用电量增长量的贡献率约为多少？

- A. 19.6%
- B. 26.7%
- C. 24.3%
- D. 31.2%

2. 国家能源局统计数据显示，本文数据来自国家能源局年度电力统计快报，统计口径为全国全口径发电量，涵盖火电、水电、风电、太阳能发电等主要电源类型，其中火电包括燃煤、燃气、生物质发电，水电不含抽水蓄能，风电与太阳能发电均按并网装机实际发电量统计。表格中分项之和与总发电量可能因四舍五入及存在其他电源而略有差异。在碳达峰碳中和目标引领下，能源结构转型步伐明显加快。‘十四五’以来，我国深入实施能源安全新战略，加快构建新型电力系统。面对全球气候变化挑战，我国作出碳达峰碳中和重大战略决策，能源绿色低碳转型是实现‘双碳’目标的关键路径。2023 年国家发展改革委、国家能源局联合印发《关于新形势下配电网高质量发展的指导意见》，2024 年国务院《政府工作报告》提出加强大型风电光伏基地和外送通道建设，2025 年国家能源局进一步推动绿色电力交易与碳市场衔接。在政策驱动下，全国发电量稳步增长，电源结构持续优化。从总体趋势看，2022—2025 年全国发电量由 87000 亿千瓦时增至 107000 亿千瓦时，累计增长 23.0%，年均增长约 7.1%。其中 2023 年发电量同比增长 6.9%，2024 年增长 7.0%，2025 年增长 7.5%，增速稳中有升，与国内生产总值增长及电气化水平提高趋势基本一致。结构变化方面，火电发电量由 59000 亿千瓦时增至 68000 亿千瓦时，但占比较 2022 年的 67.8% 下降至 2025 年的 63.6%；非化石能源发电量（含水、风、光）合计占比由 31.0% 提升至 36.4%，其中风电、太阳能发电量年均增速分别达 23.3% 与 10.1%，成为增长最快的电源类型。影响结构变化的主要因素包括新能源装机容量持续扩大、煤电灵活性改造推进，以及电力市场化改革促进绿色电力优先消纳。此外，随着电力现货市场逐步推广、绿电绿证交易机制不断完善，清洁能源发电的消纳条件进一步改善，为能源电力行业高质量发展注入了新动能。表 1 列出 2022—2025 年全国发电量及主要电源发电量数据。

年份	全国发电量 (亿千瓦时)	火力发电量 (亿千瓦时)	水力发电量 (亿千瓦时)	风力发电量 (亿千瓦时)	太阳能发电量 (亿千瓦时)
2022	87000	59000	13000	8000	6000
2023	93000	62000	14000	10000	6000
2024	99500	65000	15000	12000	7000
2025	107000	68000	16000	15000	8000

- (1) 2023 年全国发电量同比增长约（）。
- A. 7.5%
 - B. 6.9%
 - C. 6.2%
 - D. 8.1%
- (2) 2025 年非化石能源发电量（水电、风电、太阳能合计）占总发电量的比重较 2022 年提升约（）个百分点。
- A. 6.3 个百分点
 - B. 5.4 个百分点
 - C. 4.8 个百分点
 - D. 7.0 个百分点

- (3) 2022—2025 年火力发电量年均增长率约为 ()。
- A. 4.8%
 - B. 5.2%
 - C. 4.5%
 - D. 5.6%
- (4) 2024 年风力发电量增速比 2023 年风力发电量增速 ()。
- A. 高 5 个百分点
 - B. 低 2 个百分点
 - C. 低 5 个百分点
 - D. 高 2 个百分点
- (5) 2025 年全国发电量同比增长 7.5%，其中火力发电量同比增长约 4.6%，若其他电源（非火电）2024 年发电量为 34500 亿千瓦时，则 2025 年其他电源发电量同比增速约为 ()。
- A. 13.0%
 - B. 11.5%
 - C. 14.8%
 - D. 10.2%

参考答案与解析

一、单选题（共 57 题）

1. 答案：A。解析：设单程距离为 1，则去程时间 $1/80$ ，回程时间 $1/120$ ，总时间 $1/80 + 1/120 = (3 + 2)/240 = 5/240 = 1/48$ ，总路程 2，平均速度 $2 \div (1/48) = 96$ Mbps。

2. 答案：D。解析：数据要素具有非竞争性，即一个人使用数据并不减少他人使用的可能性，同时数据可以无限复制，不具有完全排他性。因此 D 选项错误。

3. 答案：B。解析：数据治理强调对数据资产的管理与控制。B 项中对交易数据进行脱敏并建立访问权限，属于数据安全治理，符合定义。A 项是数据应用，C 项是数据分析，D 项是编辑工作，均不属于数据治理。

4. 答案：B。解析：数据治理是一套持续管理数据资产的活动，其核心是提高数据质量与数据价值，确保数据准确、一致、可用、安全，从而支撑业务决策和运营。降低存储成本、扩充机柜、减少备份介质均非核心目标。

5. 答案：A。解析：提价后价格 $= 200 \times (1 + 20\%) = 240$ 元，打八折后 $= 240 \times 0.8 = 192$ 元。

6. 答案：C。解析：A 项缺少主语，应删去“通过”或“使”；B 项否定不当，“防止……不再发生”应改为“防止……再次发生”；D 项两面对一面，应删去“能否”。C 项表述正确。

7. 答案：C。解析：C 项指出统计口径发生变化，导致前后数据不可比，直接否定论据，削弱力度最强。A 项属于另有他因，也能削弱，但统计口径变化使得论据本身不成立，削弱力度更强。B 项覆盖少部分客户可能削弱适用范围，但不如 C 直接；D 项与论证无关。

8. 答案：B。解析：ETL 是 Extract（抽取）、Transform（转换）、Load（加载）的缩写，T 代表 Transform，即对抽取的数据进行清洗、规范化、去重、格式转换等处理，使其满足目标系统的要求。

9. 答案：B。解析：能被 3 整除的数有 33 个，能被 5 整除的有 20 个，能被 15 整除的有 6 个，所以能被 3 或 5 整除的有 $33 + 20 - 6 = 47$ 个，概率 $47/100 = 0.47$ 。

10. 答案：A。解析：（3）引出数据量增长的时代背景，（1）承接说明数据要素的重要性，（4）提出数据运营的问题，（2）给出解决方案，即成立数据运营分公司。因此顺序为（3）（1）（4）（2）。

11. 答案：C。解析：数据要素强调用于生产经营并带来经济效益。C 项用户购买记录可用于用户画像、精准营销等，能产生经济效益，属于数据要素。A 项考勤记录主要用于内部管理，B 项游客照片和 D 项学生档案不一定用于生产经营活动，均不符合。

12. 答案：B。解析：Hadoop 1.x 默认块大小为 64MB，Hadoop 2.x 和 3.x 默认块大小为 128MB（可在配置中调整，但默认值为 128MB）。因此选 B。

13. 答案：B。解析：数列规律为前一项 $\times 2 - 1$ ： $2 \times 2 - 1 = 3$ ， $3 \times 2 - 1 = 5$ ， $5 \times 2 - 1 = 9$ ， $9 \times 2 - 1 = 17$ ，所以 $17 \times 2 - 1 = 33$ 。

14. 答案：A。解析：A 项“未雨绸缪”比喻事先做好准备，符合语境。B 项“首鼠两端”意为犹豫不决，与“果断决策”矛盾。C 项“胸有成竹”指做事有把握，与“一败涂地”矛盾。D 项“各自为政”指不顾全局，各搞一套，与“数据孤岛”现象相符，但“各自为政”含贬义，此处使用恰当，但 A 项更符合语境，且 B、C 明显错误，D 项虽恰当但不如 A 项典型，故最佳答案为 A。

15. 答案：B。解析：题干翻译为：有效防范 $\rightarrow$ 建立制度。甲建立了制度属于肯定后件，不能推出必然能有效防范，因此不能推出 A、C、D。发生数据泄露的可能性依然存在，因此 B 正确。

16. 答案：B。解析：聚合条件应使用 HAVING 子句，且 GROUP BY 必须按非聚合列 user_id 分组。A 中 WHERE 子句不能直接使用聚合函数 SUM；C 错误地按 amount 分组；D 先过滤单笔金额大于 1000 的订单，但需求是用户总金额大于 1000，逻辑不符。正确语法为 SELECT ... FROM ... GROUP BY ... HAVING ...。

17. 答案：A。解析：2025 年： $500 \times 1.2 = 600$ ；2026 年： $600 \times 1.2 = 720$ ；2027 年： $720 \times 1.2 = 864$ 台。

18. 答案：C。解析：A 项“惠赐佳音”用于请对方给予回音，但“深感兴趣”主语是“我们”，搭配不当，应改为“我们对此项目深感兴趣，希望贵公司尽快赐复”。B 项“不负众望”指不辜负大家的期望，此处用于感谢款待，不恰当。C 项“宝贵意见”“认真考虑”“尽快改进”表达得体。D 项“绝不拖延”语气过于绝对，不够得体。因此 C 项最得体。

19. 答案：A。解析：甲说‘所有都通过’与丁说‘有的没通过’是矛盾关系，必有一真一假。题干只有一真，因此乙、丙均为假。乙假说明丙组没有通过；丙假说明甲组通过了。所以丁为真，甲为假。因此 A 项‘甲组通过了’一定为真。

20. 答案：B。解析：Spark 的核心是弹性分布式数据集（RDD），优先将数据缓存到内存中，迭代计算时避免频繁读写磁盘，因此在机器学习、图计算等迭代任务中性能显著优于 MapReduce。A 错误，Spark 仍需编写程序；C 错误，Spark 同时支持批流处理；D 错误，Spark 本身不提供强制的图形界面。

21. 答案：A。解析：甲效率 $1/12$ ，乙效率 $1/18$ ，合作效率 $1/12 + 1/18 = 5/36$ ，合作 3 天完成 $3 \times 5/36 = 5/12$ ，剩余 $7/12$ ，乙单独需 $(7/12) \div (1/18) = 7/12 \times 18 = 10.5$ 天。

22. 答案：D。解析：A 项“以及”前不应有逗号，应删去“以及”前的逗号。B 项“数据安全法”应改为“《数据安全法》”。C 项“数据交易等”后应加句号，且“等”字后不应有标点，但此处句号缺失。D 项标点使用正确。

23. 答案：A。解析：A 项将身份证号替换为随机编号，既保留了数据特征（仍是身份证号格式），又消除了与真实身份的关联，符合定义。B 项是数据汇总，C 项是权限控制，D 项保留病历号仍可关联身份，均不属于数据脱敏。

24. 答案：D。解析：缺失值填充应尽量保留数据分布特征。均值填充、删除记录、模型预测都是常见方法。而添加极端异常数会严重扭曲数据分布，影响后续分析，因此最不合理。

25. 答案：A。解析：环比增长率 $= (240 - 200) / 200 = 40 / 200 = 20\%$ 。

26. 答案：C。解析：A 项“今年的目标”可指分公司今年的目标，也可指“我们”今年的目标，有歧义。B 项“借了”可理解为借入或借出，有歧义。C 项表述清晰，无歧义。D 项“新来的”可修饰“数据管理员”，也可修饰“经理”，有歧义。因此 C 项正确。

27. 答案：B。解析：论证从‘收集数据’推出‘提升效率’，必须假设数据能够用于优化效率，即 B 项。A 项‘真实个人信息’与题干‘购买记录’不完全等同，且不是必须；C 项涉及法律合规问题，但不是论证成立的必要假设；D 项与论证无关。

28. 答案：B。解析： $DAU/MAU = 6000 \text{ 万} / 1.2 \text{ 亿} = 0.5 = 50\%$ ，表示月活跃用户中平均每日活跃的占比，是衡量用户粘性的常用指标。

29. 答案：A。解析：总成本 $= 5000 + 8000 \times 0.5 = 5000 + 4000 = 9000$ 元。

30. 答案：C。解析：A 项“通过……使……”造成主语残缺，应删去“通过”或“使”。B 项两面对一面，“是否”与“关键在于”不搭配，应删去“是否”。C 项表述正确。D 项“数据安全法”应加书名号，但此处未加，属于标点错误，故不选。因此 C 项正确。

31. 答案：B。解析：B 项解释了运动如何促进学习，建立了运动与成绩之间的因果机制，对结论支持作用最强。A 项说明样本有代表性，有一定支持但力度弱；C 项属于因果倒置，削弱；D 项排除部分干扰，支持力度不如 B。

32. 答案：B。解析：星型模型中，事实表存放业务过程的数值度量及外键，维度表存放描述属性，事实表通过外键与维度表关联。雪花模型是对维度表的规范化拆分，不是星型模型的特点。C 描述的是雪花模型，A、D 明显错误。

33. 答案：A。解析：数列为平方数： $1^2, 2^2, 3^2, 4^2, 5^2$ ，所以下一个为 $6^2 = 36$ 。

34. 答案: A。解析: (1) 引出数据交易的话题, (2) 说明数据交易的作用, (3) 转折指出数据交易面临的挑战, (4) 提出对策。因此顺序为 (1) (2) (3) (4)。

35. 答案: B。解析: 数据运营的核心在于利用数据优化业务。B 项分析用户行为并优化推荐策略, 符合定义。A 项是数据安全维护, C 项是合规工作, D 项是数据系统建设, 均不属于数据运营。

36. 答案: C。解析: 数据质量常用维度包括准确性、完整性、一致性、及时性、唯一性、有效性等, 可扩展性属于系统架构的非功能属性, 不属于数据质量维度。

37. 答案: A。解析: 至少会一种的人数 $= 6 + 5 - 3 = 8$, 两者都不会 $= 10 - 8 = 2$ 人。

38. 答案: D。解析: A 项“由于……导致”搭配不当, 应删去“由于”或“导致”。B 项“修订和完善”语义重复, 应删去“和完善”。C 项“关于”多余, 应删去。D 项表达准确、简洁。

39. 答案: C。解析: C 项说明拥堵指数下降是因为行驶速度提高, 但寻找停车位的时间增加, 导致总通勤时间增加, 合理解释了表面矛盾。A 项车流量增加会导致拥堵指数上升, 无法解释指数下降; B 项通勤距离长是固有特征, 无法解释变化; D 项统计范围问题与题干矛盾无关。

40. 答案: C。解析: 数据生命周期一般包括采集、存储、使用、共享、归档、销毁等阶段。企业应根据业务和法律要求确定保留期限, 并非所有数据都无限期保存。“无限期永久保存”不符合数据生命周期管理原则。

41. 答案: B。解析: 设原成本为 100, 则原售价 $= 100 \times (1 + 25\%) = 125$ 。新成本 $= 100 \times (1 - 20\%) = 80$, 新利润率 $= (125 - 80) / 80 = 45 / 80 = 56.25\%$ 。

42. 答案: A。解析: A 项表述正确。B 项两面对一面, 应删去“能否”。C 项“解决存在的问题”搭配不当, 应改为“解决存在的问题”或“解决问题”。D 项“随着……使”造成主语残缺, 应删去“使”。因此 A 项正确。

43. 答案: A。解析: 论证前提是‘增加服务器数量就能提升数据处理能力’, 这实际上将数据处理能力简单等同于服务器数量, 忽略了软件优化、架构设计等因素。A 项准确指出了这一缺陷。B 项也是忽略其他因素, 但 A 项更根本地指出了逻辑错误——将不同概念等同。C、D 项与论证不直接相关。

44. 答案: B。解析: 根据规定, 企业内部使用的数据资源, 符合无形资产定义和确认条件的, 应当确认为无形资产。确认无形资产要求使用寿命能够可靠估计, 并且该数据资源预期能够带来经济利益等。A、D 明显不符, C 为费用化处理。

45. 答案: A。解析: 成功概率 $= 0.9 \times 0.8 \times 0.7 = 0.504$ 。

46. 答案: B。解析: A 项“等”字前不应有逗号, 应删去“应用”前的逗号。B 项标点使用正确。C 项“2024 年”后应加逗号, 但此处没有, 属于标点错误。D 项“数据安全法”应加书名号, 但此处未加, 属于标点错误。因此 B 项正确。

47. 答案: C。解析: 数据资产需同时满足“企业拥有或控制”和“能够带来未来经济利益”两个条件。A 项用户浏览记录数据库由电商平台拥有, 可用于精准营销, 属于数据资产; B 项患者电子病历由医院控制, 可用于医疗研究和管理, 属于数据资产; D 项 GPS 定位数据由物流公司拥有, 用于调度优化, 降低油耗, 属于数据资产。C 项公开气象数据是免费向社会公开的, 不属于特定企业拥有或控制, 也不能直接为企业带来经济利益, 故不属于数据资产。

48. 答案: A。解析: 同期群分析将用户按首次行为时间分组, 跟踪各群组在后续周期(如月度)的留存、复购等指标, 用于评估用户质量与生命周期价值。B 为时间序列预测, C 为价格弹性分析, D 为广告效果分析。

49. 答案: A。解析: 数列规律为 $n(n+1)$: $1 \times 2 = 2$, $2 \times 3 = 6$, $3 \times 4 = 12$, $4 \times 5 = 20$, $5 \times 6 = 30$, 所以下一个为 $6 \times 7 = 42$ 。

50. 答案: C。解析: A 项否定不当, “防止……不再发生”应改为“防止……再次发生”。B 项主语残缺, 应删去“通过”或“使”。C 项表述正确。D 项两面对一面, 应删去“能否”。因此 C 项正确。

51. 答案：A。解析：题干由引入大数据分析系统后库存周转率提升，得出大数据分析系统能降低库存成本。A 项说明这些企业同时优化了供应链管理流程，存在其他可能原因，属于另有他因，直接削弱了题干中大数据分析系统的作用，削弱力度最强。B 项只是说明未引入系统的企业也有提升，但无法否定引入系统企业自身的因果关系，削弱力度较弱；C 项“有些”属于量小，削弱力度有限；D 项说明样本代表性不足，属于质疑样本，但不足以彻底否定结论，力度不如 A 项。故答案为 A。

52. 答案：A。解析：第二季度 = $1200 \times 1.15 = 1380$ 万元，第三季度 = $1380 \times 0.9 = 1242$ 万元。

53. 答案：C。解析：A 项“虚心接受”与“立即整改”搭配不当，应改为“虚心接受，并立即整改”。B 项“后果自负”语气强硬，不够得体。C 项表达得体。D 项“非常幼稚”不尊重对方，不得体。因此 C 项正确。

54. 答案：A。解析：设 P 为“数据平台能按时上线”，Q 为“数据清洗工作按时完成”。甲的预测为 P，丙的预测为 $\neg P$ ，二者构成矛盾关系，必有一真一假。由于只有一人说假话，则乙和丁的预测均为真。丁的预测为“ $\neg Q$ 且 P”，联言命题为真，则 $\neg Q$ 为真且 P 为真，即数据清洗工作未按时完成，且数据平台按时上线。因此 P 为真，故甲说的真话，丙说的假话。乙的预测为“ $Q \rightarrow P$ ”，由于 Q 为假，整个蕴含命题必然为真，与丁为真一致。综上可得数据平台按时上线了，A 项正确。B 项应为未按时完成，C 项甲真丙假，D 项乙丁均真，均错误。故答案为 A。

55. 答案：C。解析：设计划需要 x 天完成，则总记录数为 $120x$ 万条。实际每天处理 $120 + 30 = 150$ 万条，实际用时 $x-2$ 天，所以总记录数也为 $150(x-2)$ 万条。因此 $120x = 150(x-2)$ ，解得 $120x = 150x - 300$ ， $30x = 300$ ， $x = 10$ 。总记录数 = $120 \times 10 = 1200$ 万条。

56. 答案：A。解析：A 项表述正确。B 项“由于……导致”搭配不当，应删去“由于”或“导致”。C 项“提出了宝贵意见”搭配不当，应改为“提出了宝贵建议”。D 项“随着……使”造成主语残缺，应删去“使”。因此 A 项正确。

57. 答案：B。解析：设“通过数据安全审计”为 A，“实施访问控制”为 B，“部署日志审计模块”为 C。由题干“凡是通过数据安全审计的系统，都必须实施访问控制”可得 $A \rightarrow B$ ；“所有实施访问控制的系统都部署了日志审计模块”可得 $B \rightarrow C$ ；因此根据连锁推理可得 $A \rightarrow C$ ，即所有通过数据安全审计的系统都部署了日志审计模块，B 项正确。A 项：题干只说明有些部署了日志审计模块的系统没有通过审计，即有些 C 不是 A，不能推出有些非 A 不是 C。C 项：“有些 C 不是 A”不排除“有些 C 是 A”为真，但也不必然为真。D 项：题干给出 $A \rightarrow B$ ，不能推出 $B \rightarrow A$ 。因此只有 B 项一定为真。

二、多选题（共 3 题）

1. 答案：AC。解析：无监督学习指训练数据无标签，常见任务有聚类、关联规则。K-means 是聚类算法，Apriori 是关联规则算法，均属无监督。决策树和线性回归属于有监督学习。

2. 答案：ABCD。解析：DCMM 定义了 8 个能力域：数据战略、数据治理、数据架构、数据应用、数据安全、数据质量、数据标准、数据生存周期。题目中四项均属于。

3. 答案：ABCD。解析：《个人信息保护法》第五条规定：处理个人信息应当遵循合法、正当、必要和诚信原则，不得通过误导、欺诈、胁迫等方式处理个人信息。因此四项均正确。

三、判断题（共 5 题）

1. 答案：正确。解析：数据生命周期管理是指从数据产生到销毁的全过程管理，一般涵盖采集、存储、使用、共享、归档和销毁等关键阶段，各阶段均有相应的管理规范和控制措施，故题干表述正确。

2. 答案：正确。解析：《个人信息保护法》第六条规定，处理个人信息应当具有明确、合理的目的，并应当与处理目的直接相关，采取对个人权益影响最小的方式；收集个人信息，应当限于实现处理目的的最小范围。故题干表述正确。

3. 答案：正确。解析：数据备份是数据中心容灾的基础手段，定期备份可降低数据丢失风险；异地存储或采用不同介质可实现数据冗余，提高容灾能力，是数据中心运营规范的常见要求。故题干表述正

确。

4. 答案：错误。解析：数据资产确权是对数据资源相关权利进行界定和确认的过程，通常涵盖数据资源持有权、数据加工使用权、数据产品经营权等权利，并衍生出收益权等权属。因此，数据资产确权不仅涉及使用权，还包括所有权（或持有权）和收益权等，题干表述错误。

5. 答案：正确。解析：“数据二十条”明确提出建立保障权益、合规使用的数据产权制度，探索数据资源持有权、数据加工使用权和数据产品经营权三权分置的产权运行机制，故题干表述正确。

四、资料分析（共 2 题）

1. 答案：C；B；A；A；B

(1) C。由材料可知 2024 年、2025 年全社会用电量同比增速分别为 6.8% 和 6.6%。根据间隔增长率公式， $r = 6.8\% + 6.6\% + 6.8\% \times 6.6\% = 13.4\% + 0.4488\% \approx 13.85\%$ ，最接近 13.8%，故选 C。

(2) B。2025 年第三产业占比 = $20100 \div 105000 \approx 19.14\%$ ，2024 年占比 = $18348 \div 98521 \approx 18.62\%$ ，比重差 = $19.14\% - 18.62\% = 0.52$ 个百分点，约 0.5 个百分点。

(3) A。第一步，求 2025 年居民生活用电量增速： $(16150 - 14942) \div 14942 = 1208 \div 14942 \approx 8.09\%$ 。第二步，由城镇居民占 75%、增速 9.2%，设农村居民增速为 x ，利用十字交叉： $(9.2\% - 8.09\%) / (8.09\% - x) = 25\% / 75\% = 1/3$ ，即 $8.09\% - x = 3 \times (1.11\%) = 3.33\%$ ，解得 $x \approx 4.76\%$ ，约 4.7%。

(4) A。年均增长率 = $(67300 \div 60745)^{\frac{1}{2}} - 1 \approx (1.1079)^{\frac{1}{2}} - 1 \approx 1.0525 - 1 = 5.25\%$ ，约 5.2%。

(5) B。第三产业增长量 = $20100 - 16694 = 3406$ 亿千瓦时，全社会增长量 = $105000 - 92241 = 12759$ 亿千瓦时，贡献率 = $3406 \div 12759 \approx 26.7\%$ 。

2. 答案：B；B；A；C；A

(1) B。2023 年全国发电量为 93000 亿千瓦时，2022 年为 87000 亿千瓦时，同比增长量为 $93000 - 87000 = 6000$ 亿千瓦时，增速 = $6000 \div 87000 \approx 6.90\%$ ，因此 D 选项正确。

(2) B。2022 年非化石能源发电量 = $13000 + 8000 + 6000 = 27000$ 亿千瓦时，占比 = $27000 \div 87000 \approx 31.03\%$ ；2025 年非化石能源发电量 = $16000 + 15000 + 8000 = 39000$ 亿千瓦时，占比 = $39000 \div 107000 \approx 36.45\%$ ；提升约 $36.45 - 31.03 = 5.42$ 个百分点，即约 5.4 个百分点，因此 D 选项正确。

(3) A。设年均增长率为 r ，则 $59000 \times (1 + r)^3 = 68000$ ， $(1 + r)^3 = 68000 \div 59000 \approx 1.15254$ ，解得 $1 + r \approx 1.0488$ ， $r \approx 4.88\%$ ，即约 4.8%，因此 C 选项正确。

(4) C。2023 年风力发电量增速 = $(10000 - 8000) \div 8000 = 2000 \div 8000 = 25\%$ ；2024 年风力发电量增速 = $(12000 - 10000) \div 10000 = 2000 \div 10000 = 20\%$ ； $25\% - 20\% = 5$ 个百分点，即 2024 年增速比 2023 年低 5 个百分点，因此 D 选项正确。

(5) A。2025 年总发电量同比增量 = $107000 - 99500 = 7500$ 亿千瓦时；2025 年火力发电量同比增量 = $68000 - 65000 = 3000$ 亿千瓦时；则 2025 年其他电源发电量同比增量 = $7500 - 3000 = 4500$ 亿千瓦时；2024 年其他电源发电量为 34500 亿千瓦时，所以 2025 年其他电源发电量同比增速 = $4500 \div 34500 \approx 13.04\%$ ，即约 13.0%，因此 C 选项正确。