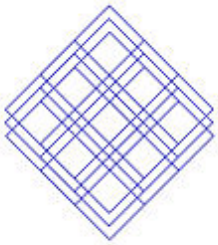

ENTERPRISE BIG DATA PROFESSIONAL

企业大数据专家级考试

多选项测试

90分钟考卷

说明

1. 所有60个考题都应该尝试回答。
2. 所有答案都需要填充在答题格上。
3. 请在所提供的答题纸上使用铅笔而不是墨水笔标记您的答案。
4. 每个问题只会有一个正确答案
5. 90分钟内完成试卷
6. 必须答对39个或以上的题才能通过考试

1 大数据四大特征之一是什么？

a) Validity - 合法性
b) Volume - 体量
c) Value - 价值
d) Variability - 可变性

2 以下哪一项是很多企业难以通过大数据实现竞争优势的最主要原因之一？

a) 他们将信息视为战略资产
b) 他们所有的大数据都在积极的被使用
c) 他们没有充分定义大数据战略
d) 他们在数据的严谨性上竞争

3 当数据集中的大多数值大于平均值时，指的是何种类型的偏度(skew)？

a) 大偏度
b) 小偏度
c) 正偏度
d) 负偏度

4 通常与监督机器学习相关的技术有哪些？

a) 分类和关联
b) 关联和回归
c) 分类和回归
d) 回归和聚类

5 以下哪些陈述适用于结构化数据？

1. 以预先定义的格式或表格进行组织
2. 通常存储在Excel文件或SQL数据库中
3. 通常写满了各类文字，可能包含日期，数字和其它内容
4. 显示字段之间建立关系的位置

a) 1, 2, 3
b) 1, 2, 4
c) 1, 3, 4
d) 2, 3, 4

6 完成了一些简单的大数据项目之后，一个组织现在希望确保他们的大数据团队保留他们的大数据知识并采用持续学习的文化。

实现哪种组织成功因素最能支持这个目标？

a) 该组织希望通过大数据实现目标的清晰视角
b) 有效的数据治理和数据管理流程
c) 一个集中并卓越的大数据中心
d) 正在进行的大数据专家的培训计划

7 分析深度学习技术主要用于哪种数据

a) 分类和随机
b) 字母和数字
c) 图像和音频
d) 机器和整数

8 大数据成熟度等级上的哪个级别是贯穿组织所有领域去管理数据和分析能力，但尚未完全优化其所有大数据活动的？

a) 2级 - 分析能力本地级
b) 3级 - 分析能力运营级
c) 4级 - 分析能力企业级
d) 5级 - 数据驱动企业级

9 认知分析与其他形式的分析的差异在哪里？

a) 根据所感知环境和个性化特征作出决策
b) 决策过程考虑了广义的意识和情商
c) 认知原理和一系列广义的条件被纳入决策过程
d) 决策是自动化的，与任何个人偏好无关

10 在制定大数据战略时，应按以下顺序进行以下操作？

1. 定义业务目标
2. 确定用例并划分优先顺序
3. 执行现状评估
4. 制定大数据路线图

a) 1, 3, 2, 4
b) 1, 3, 4, 2
c) 3, 1, 2, 4
d) 3, 2, 4, 1

11 大数据对用来做预测的数据样本来说怎样更有效？

a) 需要针对结果需要比较小的数据子集
b) 样本量需要更接近总数
c) 需要在所有数据样本之间创建关系
d) 样本需要独立于整体

12 以下哪项是参考架构最好的描述？

a) 一套项目经理或其他相关部门可参考的最佳实践文件
b) 一个使用计算机内的通用结构配置的信息源
c) 一批在其他人获得的通用知识基础上共享的建议
d) 一个基于其他人经验形成的并基于标准设计原则的建议

13 根据大数据框架的说法，大数据的定义是什么？

a) 一组无法手动处理或通过自动计算机软件处理的大量信息
b) 创建能够处理大量数据的计算机程序
c) 内部，外部，结构化，半结构化，非结构化和企业数据的整合
d) 一套去探索如何对通过海量数据去演绎出有价值的见解的技巧、技能与技术

14 以下哪个是固态硬盘（SSD），光盘驱动或移动硬盘连接到计算机的例子？

a) 存储区域网络（SAN）
b) 网络附加存储（NAS）
c) 直接附加存储（DAS）
d) 直接区域网络（DAN）

15 以下哪一项是NIST大数据指导架构中的五个主要角色之一？

a) 数据制造商
b) 数字分析师
c) 数据提供者
d) 对话提供者

16 可以使用什么类型的分析来探讨受试者群体服用的维生素量与其寿命之间是否存在关系？

a) 聚类
b) 采样
c) 关联
d) 回归

17 哪些特征适用于大数据中的数据集？

a) 包含一些无意义的细节
b) 来自单一数据源
c) 在通用的结构中呈现
d) 利用了传统存储能力

18 以下哪项是精心设计的大数据实验室的关键特征？

1. 开放的协同空间
2. 没有干扰的独立工作空间
3. 创新实验环境
4. 具有4GB RAM的联网电脑

a) 1, 2, 3
b) 1, 2, 4
c) 1, 3, 4
d) 2, 3, 4

19 一家保险公司希望使用其现有客户样本的用户画像来确定用于新保险产品的有效的广告投放。

什么类型的业务目标可以解决这种需求？

a) 描述性业务目标
b) 推论性业务目标
c) 预测性业务目标
d) 机理性业务目标

20 Hadoop是如何克服正在存储或处理的数据丢失风险的？

a) 每天多次将整台设备都备份到集中的数据库系统中
b) 使用并行编程模型来降低处理过程中丢失数据的风险
c) 使用存储区域网络（SAN）来增强存储设备
d) 复制数据包并将它们存储在多个不同的设备上

21 对20万名考生的测试结果的分析显示，99.9%的候选人得分在76%至89%之间。

有18名考生得分低于20%。 用什么统计术语来描述这些异常？

a) 极端值（Outliers）
b) 随机数据点
c) 偏差（Deviations）
d) 多变量观察

22 在下面的句子中填上适当的词。

大数据［ ？ ］需要帮助重点关注大数据分析和解析方面的组织投入。

a) 战略
b) 架构
c) 团队
d) 职能

23 哪种说法不适用于Hadoop？

a) 是用于处理大数据数据集的开源软件框架
b) 由假设硬件很少发生故障而设计的模块组成
c) 假定硬件故障可以被框架自动化处理
d) 已经成为广为人知的并能连接不同大数据解决方案的生态系统

24 4,000名学生在全国各地完成了一份考试卷。 这些结果需要传达给所有考试中心。

什么类型的图形表示会显示最高分和最低分、平均分数、上下四分位数以及不符合其他一般模式的任何单独分数？

a) 散点图
b) 双标图
c) 箱形图
d) Q-Q图

25 以下哪一项是使用“开放”大数据参考架构的好处？

a) 保证对原始数据进行快速准精确的结果分析
b) 支持使用不断变化的解决方案来解决类似的问题
c) 在应用术语方面的解释将不受限制并开放
d) 为利益相关者提供一种通用语言

26 在数据管理流程中，以下关于数据管理过程的数据改进和验证活动的陈述哪些是正确的？

1. 目标是减少数据集的错误。
2. 如果检测到任何损坏的数据，它会触发警报。

a) 只有1是对的
b) 只有2是对的
c) 1和2都是对的
d) 1和2都不对

27 什么类型的统计能定量描述或总结一组信息的特征？

a) 描述性
b) 总结性
c) 表达性
d) 解释性

28 什么样的数据治理活动能确定组织外部供应商应该如何管理数据访问、检索、存储、销毁和备份的策略，以确保数据的管理和保护得以维护？

a) 制定数据质量战略
b) 监管和隐私审查要求
c) 制定数据治理政策
d) 分配角色和责任

29 下列哪一项不是Hadoop架构中的核心组件？

a) 名字节点（NameCode）
b) 映射归约（MapReduce）
c) 从属节点（Slave Node）
d) 任务发现（Job Seeker）

30 在分布图统计中，对于一组数据值来说，高标准差反映了什么？

a) 数据输入都是高质量
b) 数值接近预期值
c) 数据分布广泛
d) 数据分布在平均值附近

31 数值聚类到平均值附近数据分析过程中的哪一步将使用查找表来交叉参考和更正以错误格式输入的区域代码？

a) 数据识别
b) 数据收集和采购
c) 数据评审
d) 数据清洗

32 为什么偏度（skewness）在数据研究中很重要？

a) 更强调预测潜在的错误
b) 为均值偏离设定限制
c) 允许组合多个数据集
d) 从几个不同的角度呈现结果

33 制造企业应该使用什么类型的解析来理解为什么某些产品在东南亚市场表现良好？

a) 描述性解析(Descriptive analytics)
b) 诊断性解析(Diagnostic analytics)
c) 预测性解析(Predictive analytics)
d) 规范性解析(Prescriptive analytics)

34 数据识别图是如何在数据分析流程中使用的？

a) 从各种来源收集数据并确定数据的价值
b) 确定数据集中是否存在任何问题或结果
c) 确定数据集是否包含缺失值
d) 确定可能获得原始数据的位置

35 随着时间的推移，大数据的以下能力应该遵照什么样的次序发展？

1. 位置感知和以人为本的分析
2. 数据挖掘和统计分析
3. 网站分析和智能化

a) 1, 2, 3
b) 2, 1, 3
c) 2, 3, 1
d) 3, 2, 1

36 关于大数据分析架构的以下哪些陈述是正确的？

1. 离线分析比实时分析更昂贵。
2. 离线分析可以同时处理多批数据.

a) 只有1是对的
b) 只有2是对的
c) 1和2都是对的
d) 1和2都不对

37 在制定大数据战略时，根据哪个措施来确定用例的优先级？

a) 业务相关程度和利益相关人的认可
b) 针对成功的预定义措施的当前性能
c) 受影响的用户组数量，以及所需的数据源
d) 用户的预期收益和实施能力

38 下表列出了6月份伦敦一周内录得的气温 (0° C)。 在此期间的平均温度为19° C。

	周日	周一	周二	周三	周四	周五	周六	总计
温度 ° c	17	17	18	18	21	20	22	133
- 19° c	-2	-2	-1	-1	2	1	3	
(温度 -19° c) 2	4	4	1	1	4	1	9	24

结果: 24° c / 7 天 = 3.4° c

这个 " 3.4 " 代表什么度量？

a) 区间距
b) 四分位区间距
c) 方差
d) 标准偏差

39 Hadoop框架内的工作跟踪器（Job Tracker）的目的是什么？

a) 遵循处理作业的指示
b) 启动和协调处理工作
c) 跟踪数据的位置
d) 加载并分析最终结果

40 一家银行决定按照大数据框架中概述的步骤制定大数据战略。 他们已经确定了几个潜在的用例，其中大数据可能为组织提供长期的战略价值。 已经确定了27个用例，这比组织在预算和资源方面可以容纳的要多。

下一步将采取什么样的最佳措施？

a) 根据业务影响，预算和资源需求，审查每个用例的优先顺序
b) 制定大数据路线图，以确定哪些项目将首先执行
c) 执行现状评估以确定使用的用例是否和当前业务目标保持一致
d) 查找多个数据源之间的相关性，以确定哪些用例最可行

41 什么样的分布形态可以反映出客户在电子邮件营销后购买新产品的可能性？

a) 频率分布（Frequency）
b) 概率分布（Probability）
c) 抽样分布（Sampling）
d) 正态分布（Normal）

42 患者骨折的高分辨率X线的信息同时也包括了患者姓名，出生日期和患者编号，出生日期是属于什么例子？

a) 数据偏差
b) 数据可视化
c) 元数据
d) 非结构化数据

43 在下面的句子中填上遗漏的词。

标准化的统计流程对大数据来说很重要，因为它使数据集里面的［ ？］成为可能。

a) 消除偏差
b) 识别关系
c) 多个变量比较
d) 确定极端值

44 以下哪些描述是正确的？

1. 关联性评估是基于自变量(independent variables)对因变量(dependent variable)的影响来进行预测的。
2. 回归探索(regression explores)可以了解一个变量导致另一个变量发生变化的原因。

a) 只有1是对的
b) 只有2是对的
c) 1和2都是对的
d) 1和2都不对

45 根据观察到的症状，可以使用什么样的算法对患者进行诊断？

a) 计算
b) 关联
c) 聚类
d) 分类

46 以下关于本地式和分布式存储解决方案之间差异的说法是正确的？

1. 商业智能分析工具最适合分布式存储解决方案。
2. 对于本地数据存储解决方案中保存的数据，使用分析和可视化工具更为合适。

a) 只有1是对的
b) 只有2是对的
c) 1和2都是对的
d) 1和2都不对

47 分类算法是做什么的？

a) 指出变量之间是否存在关系
b) 根据特征的相似性对一组对象进行分组
c) 确定新的观测数据属于哪个已知类别
d) 将大数据集合汇总成汇总图

48 聚类算法有什么作用？

a) 指出变量之间是否存在关系
b) 根据特征的相似性对一组对象进行分组
c) 确定新的观测数据属于哪个已知类别
d) 将大数据集合汇总成汇总图

49 在分布式存储系统中，为加速数千台服务器上存储的数据访问而开发了什么？

a) 文件系统
b) NoSQL数据库
c) 并行编程模型
d) 直接附加存储（DAS）

50 以下哪些陈述是正确的？

1. 数据分析(data analysis)的主要目的是评审现有数据以支持决策。
2. 解析(analytics)的主要目的是分析数据集以优化未来.

a) 只有1是对的
b) 只有2是对的
c) 1和2都是对的
d) 1和2都不对

51 以下哪项不是大数据框架中六大要素之一？

a) 大数据架构(Big Data Architecture)
b) 大数据算法(Big Data Algorithms)
c) 大数据方法(Big Data Methodologies)
d) 大数据战略(Big Data Strategy)

52 希望在线客户在网站付款屏幕被建议购买其它产品，推荐使用什么类别的机器学习用于网站引擎？ 这些建议是基于其他进行类似购买行为的客户的画像(Profiles)。

a) 结构化
b) 监督式
c) 非监督式
d) 非结构化

53 NIST大数据参考架构中的系统调配器的功能是什么？

a) 将新的数据或信息注入大数据系统
b) 确保大数据环境的组件协同工作
c) 使用业务逻辑和功能将数据转换为期望的结果
d) 根据针对大数据优化的设计来存储和处理数据

54 哪一个大数据流程涵盖了如何分配角色和职责，以便在整个业务中保持一致和适当的数据处理？

a) 数据分析
b) 数据管理
c) 数据治理
d) 数据组织

55 哪个大数据团队角色需要能够提出创造性想法来帮助他们设计和开发算法？

a) 大数据分析师 (Big Data Analyst)
b) 大数据科学家 (Big Data Scientist)
c) 大数据工程师 (Big Data Engineer)
d) 大数据架构师 (Big Data Architect)

56 在人工智能中，哪个必要的能力是来自于人工处理问题后建模推理(model reasoning)的研究成果？

a) 自动推理 (Automated reasoning)
b) 知识表示 (Knowledge representation)
c) 机器学习 (Machine learning)
d) 自然语言处理 (Natural language processing)

57 什么叫关联（correlation）？

a) 两个数据集之间的因果关系
b) 将数据点转换为标准值
c) 两个随机变量之间的统计关系
d) 数据集中最大值和最小值之间的差异

58 在认知分析中，根据感知环境和特定用户的模式，做出决策和采取特定行动的叫什么？

a) 智能算法（Intelligent algorithm）
b) 节点中介（Active mediator）
c) 操作员（Human operator）
d) 理性智能体（Rational agent）

59 以下那个活动是根据图灵测试（Turing test）用来评估设备的智能水平，以确定其是等同于人类还是与人类没有区别的？

a) 数学方程（Mathematical equations）
b) 科学算法（Scientific algorithms）
c) 书面问题（Written questions）
d) 口头交流（Verbal communication）

60 一家快速增长的电信公司正打算建立其大数据能力， 已经实现这个目标已经完成了一些研究工作，公司现在可以采取什么最好的方法？

a) 聘请其他组织的有经验的数据研究员
b) 开始组建卓越的大数据中心
c) 设计大数据分析、大数据管理和大数据治理的流程
d) 为组织确定最适合的大数据工具