

- | | |
|---|------------------------------|
| 1 | 作业代号: catfish |
| 2 | 作业布置时间: 2024.9.25 |
| 3 | 作业截止提交时间: 2024.9.29 23:59:59 |
| 4 | 作业提交文件名: catfish.txt |

下列选择题为单选题，请选择正确的答案

一、下列描述中，哪一个对实体库集成系统和虚拟集成系统的描述是正确的：

- A. 实体库集成系统的数据集成框架与虚拟集成系统的框架所针对的数据源不同
- B. 实体库集成系统的数据集成框架是针对数据库的，而虚拟集成系统的框架是针对文件的
- C. 实体库集成系统的数据集成框架是先进的，而虚拟集成系统的框架是落后的
- D. 实体库集成系统的数据集成框架与虚拟集成系统的框架所基于的访问技术都是ODBC

二、下列对若干IT标准的描述中，哪句话是不准确的：

- A. SGML不是为互联网而设计的，但是，XML的诞生与WWW有密切的关系
- B. HTML的最早标准先于XML的标准，但是，HTML标准依然在进化之中
- C. XHTML的推出是W3C为了强化HTML规范性，但是违背了WWW的发展规律，从而导致失败
- D. W3C依然掌控着HTML的主导权，而WHATWG则掌控着XML的主导权

三、下列对于GML的描述，哪句话是准确的：

- A. GML是W3C推出的标准，主要应用于遥感产品的表达，更重要的是可以用来在Internet上进行数据交换
- B. GML是OGC推出的标准，主要应用于地理信息要素的表达，更重要的是可以用来在Internet上进行数据交换
- C. GML是W3C推出的标准，主要应用于地理信息要素的表达，更重要的是可以用来在数据库上进行数据交换
- D. GML是OGC推出的标准，主要应用于图形化的表达，更重要的是可以用来在Internet上进行数据交换

四、针对XML相关的几个概念，下面哪句话的描述是准确的？

- A. Xpath只能用于XML的解析，不能用于HTML的解析
- B. XQuery 可用来查询任何可作为 XML 形态呈现的数据，包括数据库
- C. XLink定义了 XML 文档中创建超级链接的标准方法，而XPointer 则定义了XML之中的指针标准
- D. XSLT 用来定义 XML 文件的标准化语言

五、对于XML、HTML，下面哪种描述是错误的：

- A. XML不是HTML的替代，但是HTML的补充
- B. XML允许自定义标签与结构，但是HTML不允许
- C. XML被设计用来传输和存储数据，其焦点是数据的内容；而HTML 被设计用来显示数据，其焦点是数据的外观
- D. XML是WHATWG组织的推荐的，而HTML则是W3C推荐的

六、使用 chrome 浏览器访问我们的服务器：<http://119.36.242.188:9990/xml/message.html>，观察响应的页面是正常的。但是，如果将该页面源码以及相应的数据文件 message.xml（页面html 源码以及数据文件均可在我们的教辅材料中下载）放置在本地同一个目录下，通过 chrome 在本地加载message.html，却无法在页面中显示 XML 中的内容。下面给出的原因中，哪一个是错误的？

- A. 由于 XMLHttpRequest 是服务器发起的，本地文件系统无web服务能力
- B. 由于CORS 策略的限制，http的请求与XML文件加载不同源
- C. 在加载XML文件时，文件夹路径错误
- D. XMLHttpRequest对象的实例，即xmlDoc，被初始化为`null`

七、对比OLTP、OLAP，下面哪句话的描述有问题：

- A. OLTP面向的用户是操作人员，底层管理人员，而OLAP面向的是决策人员，高级管理人员
- B. OLTP在数据层面上，主要是针对常规操作处理，而OLAP则是面向的决策分析
- C. OLTP处理的数据规模远大于OLAP处理的数据规模
- D. 在数据处理方式上，OLTP的特征是原始的、当前的、细节性的、二维的、独立的、可更新的，而OLAP则是提炼性的、历史的、综合性的、多维的、统一的、周期性增加的

八、对于ETL技术，下面哪句话的描述是准确的？

- A. ETL只能用于数据仓库技术，不能用于数据挖掘技术
- B. E=Extract，T=Transform，L=Load，分别表示抽取、转换、加载

- C. ETL技术是在常规数据挖掘中作用巨大，但是无法用于空间数据挖掘
- D. ETL技术中，不包含数据清洗

九、对于数据库技术、数据仓库技术两者而言，下面哪种表述存在问题：

- A. 数据库与数据仓库都需要应用程序接口完成对它们的访问
- B. 数据库与数据仓库技术本质上是一样的，只是数据仓库所表达的内容更加复杂，体量更大
- C. 数据仓库是一个面向主题的、集成的、相对稳定的、反映历史变化的数据集合，用于支持管理决策
- D. 数据仓库的更新往往比数据库更新的频率低

十、对于当前人工智能领域的重要分支--大语言模型，下面的叙述涉及到不同的侧面。请指出描述不准确的一句：

- A. ChatGPT是一种生成式推理模型。其本质上就是由数据学习联合概率分布 $P(X, Y)$ ，然后求出条件概率分布 $P(Y|X)$ ，作为预测的模型。
- B. 大语言模型本质上是统计学习方法在人工智能领域的成功应用，而统计学习方法的要素是：模型、策略、算法、以及算力。
- C. 之所以叫做「大语言模型」，是因为它是基于人类的自然语言为学习与推理载体的，因此对于不同的语种，需要不同的语料作为学习的依据。
- D. 统计学的发展，推动了人工智能的应用；而人工智能近年来的爆炸式发展，与基于GPU的算力大幅提升有密切的关系。

十一、数据集成的意义可总结为：系统轻载、应用闭环、数据价值。下面对此3个方面的描述中，哪一句是不准确的：

A. 由于历史数据中存在大量的「数据化石」，严重影响了系统的效率、稳定性，极大地增加了维护成本和系统开销。解决之道在于，改造底层的架构，轻载在线的热数据，而将以前离线的化石数据变成近线的“温”数据，并对这些近线数据提供有效、高速、便捷的访问技术。

B. 对生产系统中的数据实现应用闭环，可以改进现有的系统，如：对于一个推荐系统而言，可以大大提高推荐的准确度；对于数据项目而言，实现应用闭环则可以加大数据之间的关联度，从而提升数据的信息含量。

C. 数据的价值体现在「经济价值」和「社会价值」两方面。一般而言，经济价值越小的数据，社会价值反而越大，反之亦然。

D. 从系统轻载，到应用闭环，再到数据价值，可以看成是数据集成的3个阶段，伴随该3阶段，数据集成所发挥的作用也越来越重要。