

第三届全国工业和信息化技术技能大赛
——工业大数据（数据库运行管理员）赛项

理论考试

命
题
方
案

2025年9月

为了参赛选手更有效的备赛，以及确保竞赛组织的严密性和有效性，根据第三届全国工业和信息化技术技能大赛——工业大数据赛项的技术方案，对理论考试的命题工作进行说明。理论考试样题见附件。

一、理论考试的范围及所占总分比例

理论考试的知识范围分 5 个模块，分别是：

数据库原理及应用，占 20%；数据收集与导入，占 20%；工业大数据平台架构，占 20%；人工智能理论基础，占 20%；智能制造技术基础，占 20%。

二、试题类型和分值

试题全部是客观题，分别是：（1）单项选择题；（2）多项选择题；（3）判断题。考试试题由 30 道单项选择题、20 道多项选择题、30 道判断题组成，共计 80 道题目，题目总分 100 分，其中单项选择题每题 1 分、多项选择题每题 2 分、判断题每题 1 分。

三、考试时间

理论考试时间为 1 小时。

四、考试方式

采用计算机考试。

五、命题和组卷方式

命题采用专家命题，建立具有一定规模的竞赛题库，采用按比例随机组卷的方式生成理论试题。

六、复习参考书

每项内容指定 1~2 本复习参考书，复习参考书的选取范围以职业院校现行的教材或最新的权威教材为主，在命题时避免题目的二义性，如果发生则以指定参考书的表述为准。

供参考教材清单如下：

[1] 数据结构(Python 语言描述)：张玉华、吕强、朱晓旭主编，清华大学出版社

[2]人工智能应用基础：史茨中、钱晓忠主编，电子工业出版社

[3] 计算思维与人工智能基础：周勇主编，人民邮电出版社

[4] 机器学习及应用：李克清、时允田主编，人民邮电出版社

[5] 工业大数据分析实践：田春华、李闯、刘家扬 等主编，电子工业出版社

[6] 机器视觉及其应用技术：刘韬、葛大伟主编，机械工业出版社

七、每套赛卷各模块的题目类型和数量分配

根据每套赛卷的总量和各模块占比，综合计算，按如下方式分配题目数量。

表 1 每套赛卷的题目类型数量和理论知识模块对照表

序号	题目类型	理论模块	数据库原理及应用	数据收集与导入	工业大数据平台架构	人工智能理论基础	智能制造技术基础	小计
1	单项选择题	8	8	8	8	8	8	40
2	多项选择题	4	4	4	4	4	4	20
3	判断题	8	8	8	8	8	8	40
合计		20	20	20	20	20	20	100

八、题库题目数量

每套赛卷是 100 道题目，按照不重复可组 10 套的总量来命题，计总题目有 1000 道。

附件：第三届全国工业和信息化技术技能大赛——工业大数据赛项理论考试样题

附件

第三届全国工业和信息化技术技能大赛

——工业大数据赛项理论考试样题

考试时间：60 分钟

一、单项选择题（共 30 题，每题 1 分，共 30 分）

1. 【单选题】 在数据库的分区表设计中，分区的目的是？
 - A. 提高数据冗余
 - B. 提高查询效率并减少锁冲突
 - C. 增加数据的存储空间
 - D. 降低表的维护成本

2. 【单选题】 在数据库的备份策略中，“增量备份”指的是？
 - A. 备份数据库中的所有数据
 - B. 备份自上次完整备份后发生变化的数据
 - C. 备份数据表的结构
 - D. 备份数据库的日志

3. 【单选题】 以下哪种技术适用于处理大规模并发访问的数据库系统？

- A. 数据库分片
- B. 数据库分区
- C. 索引优化
- D. 事务日志

4. 【单选题】在数据库系统中，数据的物理存储结构通常是由以下哪个部分管理的？

- A. 数据库管理员
- B. 操作系统
- C. 数据库管理系统
- D. 网络管理员

5. 【单选题】在关系型数据库中，主键的作用是？

- A. 唯一标识表中的每一行数据
- B. 防止数据重复
- C. 加快数据查询速度
- D. 控制数据的访问权限

6. 【单选题】以下哪种数据库模型使用表格（二维表）来存储数据？

- A. 层次模型
- B. 网络模型

- C. 关系模型
- D. 面向对象模型

7. 【单选题】在数据库的安全性控制中，授权的数据对象的（ ），授权子系统就越灵活。

- A. 范围越小
- B. 范围越大
- C. 约束越细致
- D. 范围越适中

8. 【单选题】数据库的事务 ACID 特性中的 D 代表什么？

- A. Durability（持久性）
- B. Data（数据）
- C. Database（数据库）
- D. Direction（方向）

9. 【单选题】在数据库中，外键的主要作用是什么？

- A. 唯一标识记录
- B. 保证数据的完整性
- C. 提高查询效率
- D. 限制表的大小

10.【单选题】数据库中的“冗余数据”指的是什么？

- A. 数据重复
- B. 数据丢失
- C. 数据更新
- D. 数据查询

11.【单选题】在工业大数据中，以下哪个不是数据采集的主要来源？

- A. 传感器数据
- B. 机器日志
- C. 社交媒体
- D. 操作员输入

12.【单选题】在工业大数据中，以下哪个指标不是衡量数据质量的？

- A. 准确性
- B. 完整性
- C. 可访问性
- D. 多样性

13.【单选题】在工业大数据中，以下哪个不是数据预处理的常见步骤？

- A. 数据清洗

- B.数据转换
- C.数据压缩
- D.数据增强

14.【单选题】在工业大数据中，以下哪个技术不是用于数据存储的？

- A.Hadoop
- B.Spark
- C.SQL 数据库
- D.机器学习算法

15.【单选题】大数据的利用过程是（ ）。

- A.采集——挖掘——清洗——统计
- B.采集——统计——清洗——挖掘
- C.采集——清洗——挖掘——统计
- D.采集——清洗——统计——挖掘

16.【单选题】支撑大数据业务的基础是（ ）。

- A.数据科学
- B.数据应用
- C.数据硬件
- D.数据人才

17. 【单选题】 以下哪些工具不属于数据清洗工具？（ ）

A.OpenRefine

B.DataCleaner

C.Weka

D.Visual.ly

18. 【单选题】 大数据治理范围共包括六个关键域，即战略、组织、大数据质量、大数据生命周期、（ ）和大数据架构。

A.大数据技术

B.大数据安全

C.大数据应用

D.大数据政策

19.【单选题】下列关于大数据的分析理念的说法中，错误的是（ ）。

A.在数据基础上倾向于全体数据而不是抽样数据

B.在分析方法上更注重相关分析而不是因果分析

C.在分析效果上更追究效率而不是绝对精确

D.在数据规模上强调相对数据而不是绝对数据

20. 【单选题】 下列关于可视化工具中高级分析工具的说法，错误的是？（ ）

A.R 是属于 GNU 系统的一个自由、免费、源代码开放的软件

B.Weka 主要用于社交图谱数据可视化分析，可以生成非常酷炫的可视化图形

C.Gephi 主要用于社交图谱数据可视化分析，可以生成非常酷炫的可视化图形

D.R 通常用于大数据集的统计与分析

21.【单选题】工业人工智能算法中，用于分类问题的常见算法是：

A.线性回归 B.支持向量机 C.聚类分析 D.主成分分析

22.【单选题】在工业人工智能算法中，以下哪个不是深度学习模型？

A.卷积神经网络（CNN）

B.循环神经网络（RNN）

C.决策树

D.生成对抗网络（GAN）

23.【单选题】在工业人工智能中，以下哪个算法最适合处理时间序列数据？

A.随机森林

B.支持向量机

C.长短期记忆网络（LSTM）

D.K-最近邻算法

24. 【单选题】在工业人工智能中，以下哪个算法不是用于回归问题的？

- A.线性回归
- B.决策树
- C.K-最近邻算法
- D.逻辑回归

25. 【单选题】在工业人工智能中，以下哪个不是神经网络的激活函数？

- A.Sigmoid
- B.ReLU
- C.Softmax
- D.K-最近邻

26. 【单选题】在工业人工智能中，以下哪个算法不是用于聚类分析的？

- A.K-means
- B.层次聚类
- C.逻辑回归
- D.DBSCAN

27. 【单选题】在工业人工智能中，以下哪个不是深度学习模型的部分？

- A.输入层
- B.隐藏层

- C.输出层
- D.决策树

28. 【单选题】 在机器学习中，以下哪种方法属于监督学习算法？

- A. K-means 聚类
- B. 支持向量机 (SVM)
- C. 主成分分析 (PCA)
- D. 自编码器

29. 【单选题】 有些国家的安全部门为了提高安全性，通过眼角膜识别系统控制安全门的打开和关闭。该过程使用了哪种计算机技术？

- A.语音识别技术
- B.虚拟现实技术
- C.模式识别技术
- D.智能代理技术

30. 【单选题】 智能扫地机器人能对周围环境进行探测辨别，自动躲避障碍物，这主要应用了信息技术的（ ）。

- A.微电子技术
- B.计算机技术
- C.通信技术
- D.传感技术

二、多项选择题（共 20 题，每题 1 分，共 20 分）

1.【多选题】可以从哪些维度评价数据价值：（ ）

- A.数据样本量
- B.数据品种
- C.数据完整性
- D.数据实时性

2.【多选题】数据分析主要实现哪三大作用：（ ）

- A.误差
- B.现状分析
- C.原因分析
- D.预测分析

3.【多选题】数据挖掘主要侧重解决哪几类问题：（ ）

- A.分类
- B.聚类
- C.关联
- D.预测

4.【多选题】下面关于数据分析与数据处理的描述，正确的是：（ ）

- A.数据分析过程通常会伴随着发生数据处理（或者说伴随着大量数据计算）
- B.数据分析和数据处理不存在紧密的关联关系
- C.二者是融合在一起的，很难割裂开来
- D.当用户在进行数据分析的时候，底层的计算机系统会根据数据分析任务的要求，使用程序进行大量的数据处理

5.【多选题】下面关于大数据处理与分析的描述，正确的是：（ ）

- A.在理论层面，数据分析需要统计学、机器学习和数据挖掘等知识
- B.在技术层面，包括单机分析工具（比如 SPSS、SAS 等）或单机编程语言（比如 Python、R），以及大数据处理与分析技术（比如 MapReduce、Spark、Hive 等）
- C.在大数据时代到来之前，数据分析主要以小规模的数据为主，一般使用单机分析工具（比如 SPSS 和 SAS）或者单机编程（比如 Python、R）的方式来实现分析程序
- D.到了大数据时代，数据量爆炸式地增长，数据分析就需要采用分布式实现技术，比如使用 MapReduce、Spark 或 Flink 编写分布式分析程序，借助于集群的多台机器进行并行数据处理分析

6.【多选题】大数据处理分析技术主要包括哪几种类型：（ ）

- A.批处理计算
- B.流计算

C.图计算

D.查询分析计算

7.【多选题】关于数据仓库 Impala 的描述正确的是：（ ）

A.Impala 是由 Oracle 公司开发的查询系统

B.与 Hive 类似，Impala 也可以直接与 HDFS 和 HBase 进行交互

C.Impala 采用了与商用 MPP 并行关系数据库类似的分布式查询引擎，可以直接从 HDFS 或者 HBase 中用 SQL 语句查询数据，而不需要把 SQL 语句转化成 MapReduce 任务来 执行

D.Impala 和 Hive 采用了不同的 SQL 语法、ODBC 驱动程序和用户接口

8.【多选题】Spark 的特点主要包括：（ ）

A.运行速度快

B.容易使用

C.通用性

D.运行模式单一

9.【多选题】以下哪些是人工智能在生活中的应用？（ ）

A.人脸识别

B.无人驾驶

C.机器翻译

D.智能机器人

10.【多选题】振动传感器中符合机电变换原理的有（ ）。

A.电动式

B.压电式

C.光电式

D.电感式

11.【多选题】温湿度传感器是指将温度量和湿度量转换成容易被测量处理的电信号的设备或装置，多以温湿度一体式的探头作为测温元件，将温度和湿度信号采集出来，转换成与温度和湿度呈线性关系的（ ）输出。

A.电压信号

B.电流信号

C.脉冲信号

D.震动信号

12.【多选题】实际生产生活中，可以采用“（ ）”三步法来提升数据的准确性。

A.收集

B.清洗

C.规整

D.分析

13.【多选题】数据库备份的方式包括（ ）

A. 完全备份

B. 增量备份

C. 差异备份

D. 连续备份

14.【多选题】关系型数据库的特征包括（ ）

A. 数据存储存储在表格中

B. 支持 SQL 查询语言

C. 数据表无固定结构

D. 数据间支持外键关系

15.【多选题】以下哪种操作会影响数据库的性能？（ ）

A. 索引创建

B. 查询优化

C. 数据库分区

D. 数据库加密

16. 【多选题】关于数据库事务的特性，下列说法正确的有（ ）

- A. 原子性
- B. 一致性
- C. 隔离性
- D. 持久性

17. 【多选题】NoSQL 数据库的特点包括（ ）

- A. 高可扩展性
- B. 灵活的模式
- C. 适合复杂查询
- D. 支持大规模分布式数据存储

18. 【多选题】以下哪些是 MySQL 的存储引擎（ ）

- A. InnoDB
- B. MyISAM
- C. Berkeley DB
- D. MongoDB

19. 【多选题】数据库的 ACID 特性包括（ ）

- A. 原子性
- B. 一致性

- C. 并发性
- D. 持久性

20. 【多选题】 以下哪些情况可能导致数据库出现死锁？（ ）

- A. 两个事务相互持有对方需要的锁
- B. 两个事务均尝试锁定同一资源
- C. 系统资源耗尽
- D. 事务之间未能释放锁

三、判断题（共 40 题，每题 1 分，共 40 分）

1. 【判断题】 备份操作通常不需要考虑数据库的数据一致性问题。
2. 【判断题】 在数据库的管理中，表的设计越复杂，查询效率通常会更高。
3. 【判断题】 数据库的正常运行需要定期进行碎片整理。
4. 【判断题】 数据库的日志文件记录了所有对数据库的修改操作。
5. 【判断题】 数据迁移时不需要考虑数据库的版本兼容性。
6. 【判断题】 数据库管理员需要定期监控数据库的磁盘空间。

7. 【判断题】关系型数据库通常不支持事务管理。
8. 【判断题】SQL 注入攻击是通过恶意构造 SQL 语句来破坏数据库的安全性。
9. 【判断题】数据库管理员负责数据库的高可用性设计和灾难恢复计划。
10. 【判断题】在数据库管理中，合理的用户权限设置可以有效防止未经授权的访问。
11. 【判断题】温湿度一体化模拟量型传感器可以同时把温度及湿度值变化转换成电流/电压值的变化，直接与各种标准的模拟量输入的二次仪表连接。
12. 【判断题】在工业生产自动化流程中，温度测量点要占全部测量点的四分之三左右。
13. 【判断题】常见的标定温度的方法主要有摄氏温标（ $^{\circ}\text{C}$ ）、华氏温标（ $^{\circ}\text{F}$ ）等。

14. 【判断题】空气的干湿程度叫“湿度”，常用绝对湿度、相对湿度、比较湿度、混合比、饱和差，以及露点等物理量来表示。

15. 【判断题】RS485 通信接口是对通信接口的硬件描述，它只需要两根通信线，即可以在两个或两个以上的设备之间进行数据传输。

16. 【判断题】RS485 通信数据传输的连接是半双工通信方式，即在某一个时刻，一个设备只能发送数据或者接收数据。

17. 【判断题】Modbus 数据通信采用 Master/Slave(主/从)方式。

18. 【判断题】准确的数据测量是数据采集的基础。在数据采集过程中，数据是否做过清洗，清洗的逻辑是否合理(例如，是否需要去重、是否做了去重)，数据是否有缺失，上游系统的数据回传是否正常，是否做了预警监控等。

19. 【判断题】脉冲量是在瞬间电压或电流由某一值跃变到另一值得信号量，它可以连续规律变化。

20. 【判断题】收集故障信息的具体内容为：故障时间信息、故障现象信息、故障部位信息、故障原因信息、故障性质信息、故障处理信息。

21. 【判断题】 维修是指设备维护和修理。

答案： 对

22. 【判断题】 数据库管理员的主要职责包括数据库的安装、配置、维护和备份恢复。

23. 【判断题】 数据库表的主键可以包含多个字段。

24. 【判断题】 数据库的索引是为了提高数据检索的效率而创建的。

25. 【判断题】 数据库管理员不需要负责数据库的安全性和权限管理。

26. 【判断题】 数据库的存储过程是由用户直接执行的程序。

27. 【判断题】 数据库管理员需要定期检查数据库的性能并进行优化。

28. 【判断题】 数据库中字段的索引不影响插入和更新操作的效率。

29. 【判断题】 数据库的正常备份应包括数据库文件和事务日志文件。

30. 【判断题】卷积神经网络中，对不同位置的特征进行聚合统计,称为池化 (pooling)。

池化不会丢失图像的信息，也不会降低其空间分辨率。