**判断题**

1. 元数据根据其用途可以分为技术元数据和业务元数据

T

1. 技术元数据是用于描述其使用的数据挖掘算法的数据

F

1. 业务元数据是从单位业务的角度描述数据仓库的元数据

T

1. 数据备份的历史记录属于技术元数据

T

1. 联机事务处理比联机分析处理重要 一样重要

F

1. 联机分析处理的多维分析有切片、切块、钻取和数据混合 没有数据混合

F

1. 在数据分析中，数据定义为“名义”、“连续”等是数据度量的方式

T

1. 联机分析处理中，ROLAP是指基于多维数据结构组织的OLAP实现 是基于关系数据库的OLAP实现，多维数据结构组织是MLAP

F

1. 粒度是对数据仓库中数据的综合程度高低的一个衡量

T

1. 数据越详细，粒度越小，综合程度越高

F

1. 粒度的不同选择会导致其逻辑模型具有差异

T

1. 粒度的不同选择会导致数据存储容量的差异

T

1. 数据集市是为了特定的应用目的，从数据仓库中独立出来的一部分数据

T

1. 数据集市是数据买卖交易的专用场所 不是场所

F

1. 数据集市可以汇集为数据仓库，数据仓库也可以按某一种方式分解为数据集市

T

1. 数据集市的数据也称为部门数据或主题数据

T

1. OLAP和OLTP均为面向主题的数据处理OLAP是面向主题，OLTP是面向应用的

F

1. 联系分析处理的多维分析有切片、切块、钻取和旋转

T

1. 在数据分析中，数据标记为“名义”或“连续”等是数据度量的方式

T

1. OLAP根据其存储数据的方式分为ROLAP、MOLAP和HOLAP三类

T

1. 数据仓库系统由数据源、数据的存储和管理、OLAP服务器和前端工具与应用组成

T

1. 神经网络是数据存储和管理中最主要的算法 并不是数据管理和存储

F

1. 数据仓库系统的数据源可以是静态的，也可以是动态的

T

1. 服务器是数据仓库系统的应用基础

F

1. 数据仓库系统由数据源、数据的存储和管理、OLAP服务器和前端工具与应用组成

T

1. 数据仓库系统的核心是数据源 核心是数据的存储和管理

F

1. 数据仓库系统的基础是数据存储和管理 基础是数据源

F

1. 前端、后端及中间件三层架构是常用的数据仓库系统的体系结构 是两层架构

F

1. 贝叶斯网络是白盒分析

T

1. 贝叶斯网络是一种对数据的黑盒分析方法

F

1. 人工神经网络是白盒分析 黑盒分析

F

1. 贝叶斯网络的状态节点只有两种状态，即0或1 节点值可以是离散的，也可以是连续的

F

1. 人工神经网络不能预测类别属性以外的数据

T

1. 神经网络的隐含层可以有多层

T

1. 神经网络隐含层每个层的节点个数是相同的

F

1. Hopfield模型是一种反馈型的神经网络

T

1. 神经网络不能没有隐含层

F

1. BP网络是一种前馈神经网络

T

1. 神经网络在结构化数据上表现较好，不能处理图像等其他的数据

F

1. 贝叶斯网络可以做预测

T

1. 贝叶斯网络可以做诊断

T

1. 贝叶斯网络训练时，是对其网络节点的概率和节点间的条件概率进行训练

F

1. 神经网络的隐含层的每个层的节点个数是一致的

F

1. 单层神经网络有一个隐含层

F

1. 单层神经网络没有隐含层

T

1. 不存在没有隐含层的网络

F

**填空题**

1. 数据仓库环境最重要的3个环节包括抽取、转换和加载，简称为（　ETL　）
2. 在数据分析的过程中，将丢失或不全的数据进行补充，该过程属于（　数据清洗　）。
3. 维度表由主键、分类层次和描述属性组成，对主键的选择可以采用（　自然键　）
4. 在数据分析的过程中，在最初填充数据仓库时常用的ETL方法是（　静态抽取　）。
5. 数据仓库模型中，不仅提供了事实表，还提供了维表和明细分类表的是（　雪花型模式　）
6. 数据仓库模型中，仅提供事实表和维表的是（　星型模式　）。
7. 信息包图主要用于数据仓库设计（　　）
8. 数据仓库产品有（　　）、（　　）
9. 信息包图拥有（　指标　）（　维度　）（　类别　）三个重要对象？
10. 数据仓库管理包括（　安全与权限的管理，数据更新的追踪，数据质量的检查，元数据的管理与更新，数据仓库使用状态的检测与审计，数据复制与删除，数据分割与分发，数据备份与恢复，数据存储管理。　）
11. 在层次聚类分析中常用的相似性度量有（　最小距离　）（　最大距离　）（　均值距离　）（平均距离）
12. KDD是（　基于数据库的支持发现　）

填空题

第一章：

1.数据仓库就是一个面向主题的、集成的、相对稳定的、反映历史变化的数据集合。

2.元数据是描述数据仓库内数据的结构和建立方法的数据，它为访问数据仓库提供了一个信息目录，根据元数据用途的不同可将数据仓库的元数据分为技术元数据和业务元数据。

3.数据处理通常分为两大类，联机事务处理（OLTP）和联机分析处理(OLAP)。

4.多维分析是指以“维”形式组织起来的数据（多维数据集），采取钻取、切片、切块和旋转等各种分析动作，以求剖析数据，使用户能从不同角度、不同侧面观察数据仓库中的数据，从而深入理解多维数据集中的信息。

5.ROLAP是基于关系数据库的OLAP实现，而MOLAP是基于多维数据结构组织的OLAP实现。

6.数据仓库按照其开发过程，其关键环节包括数据抽取、数据存储与管理和数据表现等。

7.数据仓库系统的体系结构根据应用需求的不同，可以分为以下4种类型，两层架构、独立型数据集市、依赖型数据集市和操作型数据存储、逻辑型数据集市和实时数据仓库。

8.操作型数据存储实际上是一个集成的、面向主题的、可更新的、当前值的（但是可“挥发”的）、企业级的、详细的数据库，也叫运营数据存储。

9.“实时数据仓库”意味着源数据系统、决策支持服务和数据仓库之间以一个接近实时的速度交换数据和业务规则。

10.从应用的角度看，数据仓库的发展演变可以归纳为5个阶段：以报表为主、以分析为主、以预测模型为主、以运营导向为主和以实时数据仓库和自动决策为主。

第二章：

1.调和数据是存储在企业级数据仓库和操作型数据存储中的数据。

2.抽取、转换、加载过程的目的是为决策支持应用提供一个单一、权威数据源。因此，我们要求ETL过程产生的数据（即调和数据层）是详细的、历史的、规范化的、可理解的、即时的和质量可控制的。

3.数据抽取的两个常见类型是静态抽取和增量抽取。静态抽取用于最初填充数据仓库，增量抽取用于进行数据仓库的维护。

4.粒度是对数据仓库中数据的综合程度高低的一个衡量。粒度越小，细节程度越高，综合程度越低，回答查询的种类越多。

5.使用星型模式可以从一定程度上提高查询效率。因为星型模式中数据的组织已经经过预处理，主要数据都在庞大的事实表中。

6.维度表一般由主键、分类层次和描述属性组成。对于主键可以选择两种方式：一种是采用自然键，另一种是采用代理键。

7.雪花型模式是对星型模式维表的进一步层次化和规范化来消除冗余的数据。

8.数据仓库中存在不同综合级别的数据。一般把数据分成4个级别：早期细节级、当前细节级、轻度综合级和高度综合级。

第三章：

1.SQL Sever SSAS 提供了所有业务数据的统一整合视图，可以作为传统报表、在线分析处理、关键性能指示器记分卡和数据挖掘的基础。

2.数据仓库的概念模型通常采用信息包图法来进行设计，要求将其5个组成部分（包括名称、维度、类别、层次和度量）全面地描述出来。

3.数据仓库的逻辑模型通常采用星型图法来进行设计，要求将星型图的各类逻辑实体完整地描述出来。

4.按照事实表中度量的可加性情况，可以把事实表对应地事实分为4种类型：事务事实、快照事实、线性项目事实和事件事实。

5.确定了数据仓库的粒度模型以后，为提高数据仓库地使用性能，还需要根据用户需求设计聚合。

6.在项目实施时，根据事实表的特点和用户的查询需求，可以选用时间、业务类型、区域和下属组织等多种数据分割类型。

7.当维表中的主键在事实表中没有与外键关联时，这样的维称为退化维，它与事实表并无关系，但有时在查询限制条件（如订单号码、出货编号等）中需要用到。

8.维度可以根据其变化快慢分为无变化维度、缓慢变化维度和剧烈变化维度三类。

9.数据仓库的数据量通常较大，且数据一般很少更新，可以通过设计和优化索引结构来提高数据存取性能。

10.数据仓库数据库常见的存储优化方法包括表的归并与簇文件、反向规范化，引入冗余、表的物理分割（分区）。

第五章：

1.分类的过程包含获取数据、预处理、分类器设计和分类器测试。

2.分类器设计阶段包含三个过程：划分数据集、分类器构造和分类器设计。

3.分类问题中常用的评价准则有精确度、查全率和查准率、F-measure、几何均值。

4.支持向量机中常用的核心函数有多项式核函数、径向基核函数、S型核函数。

第六章：

1.聚类分析包括连续型属性、二值离散型属性、多值离散型属性以及混合类属性四种类型描述属性的相似度计算方法。

2.连续型属性的数据样本之间的距离有欧式距离、曼哈顿距离和明考斯基距离。

3.略 书上P138

4.层次聚类方法包括凝聚型层次聚类和分解型层次聚类。