

## 第 11 讲 怎样管理和利用数据 II-数据库语言与数据挖掘

### 1、快速浏览---本讲视频都讲了什么？

#### 【视频 11.1 由关系模型到结构化数据库语言 SQL】

操作数据库需要有数据库语言。目前商品化数据库管理系统都支持标准的数据库语言 SQL。本段视频概述了什么是 SQL 语言，并简要介绍了如何用 SQL 语言创建数据库和数据库数据的增删改操作。后续各段视频将着重介绍其中的一条语句-SQL-SELECT 语句的应用，为此介绍了后续各段视频中将要使用的数据库中的各个关系及其属性。

#### 【视频 11.2 SQL-SELECT 之简单使用】

本段视频以丰富的示例，介绍了 SQL-SELECT 的应用方法，即如何用 SQL 表达各种查询。本段视频主要讲了对单一表的各种查询。

SELECT...FROM...WHERE...ORDER BY ...。

#### 【视频 11.3 SQL-SELECT 之多表联合操作】

本段视频以丰富的示例，继续介绍了 SQL-SELECT 的应用方法，即如何用 SQL 表达各种查询。本段视频主要讲了对多个表的联合查询。也讲了一个查询如何嵌入在另一个查询中的方法。

SELECT...FROM...WHERE...ORDER BY ...。

SELECT...FROM...WHERE....(SELECT ... FROM ... WHERE ...) ...。

#### 【视频 11.4 SQL-SELECT 之分组聚集操作】

本段视频以丰富的示例，继续介绍了 SQL-SELECT 的应用方法，即如何用 SQL 表达各种查询。本段视频主要讲了对多个表的分组聚集操作，例如如何分组求平均、求和等。

SELECT...FROM...WHERE....GROUP BY...HAVING....

#### 【视频 11.5 由数据库到数据挖掘 I-数据挖掘示例之背景与概念】

本段视频和下一段视频将给大家介绍一个数据挖掘的例子---关联规则数据挖掘。本段视频首先介绍了示例的背景与相关概念，尤其是对即将介绍的 Apriori 算法的相关概念做了介绍，在看下一视频前，应将其理解透彻。

#### 【视频 11.6 由数据库到数据挖掘 II-数据挖掘示例之计算过程】

本段视频采用数据示例模拟的方法介绍了 Apriori 算法的思想，进一步介绍了关联规则的产生方法和语义。还简要介绍了这个算法还能挖掘哪些规则。

**【视频 11.7 数据抽象与设计-怎样抽象与抽象的层次】**

**本段视频被列为选学内容。**主要讲的是方法论，计算学科的基本研究方法是抽象-理论-设计，那什么是抽象呢，视频中讲抽象是“理解-区分-命名-表达”，通过对前面介绍的数据库相关内容的回顾，比如“数据表”抽象成关系，抽象成数据库语言的过程，再比如“多维数据表”抽象成方体模型，进一步抽象出数据仓库语言的过程的分析，揭示了抽象就是“理解-区分-命名-表达”。

进一步，揭示了不同抽象层次的含义，现实-模型-元模型，现实世界-信息世界-数据世界，方法论及方法论的应用等不同的抽象层次。如下图示例。如何理解不同抽象层次，如何确定在不同抽象层次进行问题的讨论，对计算学科的学生而言很重要。请看课程视频如何解释不同抽象层次。

**【视频 11.8 数据抽象与设计-怎样进行设计与怎样研究理论】**

**本段视频被列为选学内容。**主要讲的是方法论，计算学科的基本研究方法是抽象-理论-设计，那什么是理论呢，什么是设计呢。课程视频中讲理论研究就是“定义-公理-定理-证明”，设计就是“形式-构造-自动化”，课程视频是如何讲授的呢……

## 2、学习要点指南

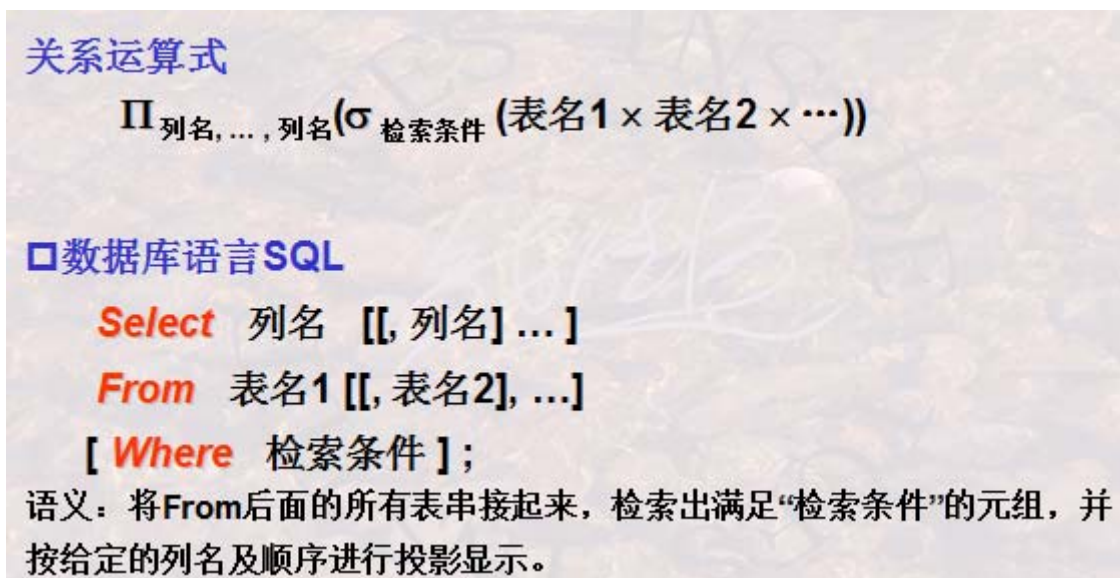
### 2.1 要点一：理解数据库语言，熟练掌握数据库的操纵技能

理解数据库语言，尤其是 SQL 语言，SELECT-FROM-WHERE-GROUP BY-HAVING-ORDER BY。一方面掌握数据库的操纵技能，另一方面体验一条数据库语句的千变万化的能力。

学习此部分内容，一定要结合具体的数据库查询需求来理解 SQL 语言。比如结合具体的专业，模拟可能建立的数据库，模拟可能的查询需求，将其用 SQL 语句表达出来，将会更深入的理解数据库。

另一方面，理解数据库语言可以结合着基于集合的关系运算的理解来理解，如下图的数据库

语言的标准语句，转换成关系运算则如图上半部所示意。即 FROM 后面相当于多个表的乘积运算，而 WHERE 语句相当于做选择运算，SELECT 语句则相当于做投影运算。因此任何形如所给出的 SELECT-FROM-WHERE 语句，相当于  $\pi(\sigma(R_1 \times R_2 \times \dots \times R_n))$ 。其执行过程逻辑上也等价于这个表达式所示意的操作次序。理解这一点对理解 SQL 语言的结果很有帮助，对将来进一步学习数据库系统实现也有帮助。



**关系运算式**

$$\Pi_{\text{列名}, \dots, \text{列名}} (\sigma_{\text{检索条件}} (\text{表名1} \times \text{表名2} \times \dots))$$

**数据库语言SQL**

**Select** 列名 [[, 列名] ...]

**From** 表名1 [[, 表名2], ...]

**[ Where 检索条件 ] ;**

语义：将From后面的所有表串接起来，检索出满足“检索条件”的元组，并按给定的列名及顺序进行投影显示。

## 2.2 要点二：理解数据挖掘，更重要的是理解能够挖掘什么？

视频以一个示例，展现了数据挖掘的基本思想。这个挖掘算法的理解本身是一方面，更重要的是它告诉我们，如果数据聚集到一定的程度，如何通过挖掘来利用这些数据；如果数据演变成大数据则它将改变人们的观念：“只注重关系，不关注因果”，“从抽样数据分析，到全数据集分析”等，什么是“一切以数据说话”。

## 2.3 要点三：理解抽象-理论-设计

理解什么是抽象“理解-区分-命名-表达”，理解什么是理论“定义-公理-定理-证明”，理解什么是设计“形式-构造-自动化”，更重要的是理解抽象-理论-设计三者之间的关系。如下图示的理解非常重要。

对方法论的理解将有助于我们形成科学的研究习惯和研究素养，有助于对问题的科学的严谨的把握。

