

国家“十三五”重点出版规划项目
上海高校优秀教材奖获得者主编

上海市高校精品课程
特色教材(立体化新形态)

数据库原理及应用

——基于SQL Server 2016

主编 贾铁军 徐方勤

副主编 邓红霞 戴春妮 王佃来 曹锐

参编 陶维天 张野 降爱莲

 机械工业出版社
CHINA MACHINE PRESS

第11章 数据库新技术



第11章 目录

- 11.1 云数据库及分布式数据库
- 11.2 大数据及主动数据库
- 11.3 数据仓库与数据挖掘
- 11.4 数据库其他新技术
- 11.5 数据库新技术应用发展趋势
- 11.6 本章小结



教学目标

- 了解数据库新技术的发展及趋势
- 了解各类型新技术产生的背景和基本概念
- 数据库新技术的特点、基本原则和目标
- 理解数据库新技术的发展趋势

友情
提示



同步实验(上机)



■ 11.1.1 云数据库的发展及趋势

■ 1. 云数据库

云数据库——指被优化或部署到一个虚拟计算环境中的数据库。可以实现按需付费、按需扩展、高可用性以及存储整合等优势。云数据库是指被优化或部署到一个虚拟计算环境中的数据库。

■ 11.1.1 云数据库的发展及趋势

■ 2. 云数据库的特点

第一，云数据库信息的“留存率”更高；

第二，扩展容易；

第三，数据迁移更为简便；

第四，相比本地数据库，维护成本也大大降低。

■ 11.1.1 云数据库的发展及趋势

■ 3. 云数据库供应商

①传统的数据库厂商：Teradata, Oracle, IBM DB2 和Microsoft SQL Server;

②涉足数据库市场的云供应商：Amazon, Google 和Yahoo;

③新兴小公司：Vertica, LongJump和EnterpriseDB。

■ 11.1.1 云数据库的发展及趋势

■ 3. 云数据库产品

- ① Amazon的云数据库产品;
- ② Google的云数据库产品;
- ③ Microsoft的云数据库产品;
- ④ 开源云数据库产品
- ⑤ 其他云数据库产品

■ 11.1.2 分布式数据库的发展及趋势

■ 1. 分布式数据库

是指数据分存在计算机网络中的各台计算机上的数据库。分布式数据库的基本思想是将原来集中式数据库中的数据分散存储到多个通过网络连接的数据存储节点上，以获取更大的存储容量和更高的并发访问量。

■ 11.1.2 分布式数据库的发展及趋势

■ 2. 分布式数据库的特点

高可扩展性:能够动态地增添存储节点以实现存储容量的线性扩展。

高并发性:分布式数据库必须及时响应大规模用户的读/写请求,能对海量数据进行随机读/写。

高可用性:分布式数据库必须提供容错机制,能够实现对数据的冗余备份,保证数据和服务的高度可靠性。

■ 11.1.2 分布式数据库的发展及趋势

■ 3. 分布式数据库相对传统集中式数据库的优点

更高的数据访问速度:采用备份的策略实现容错,客户端可以并发地从多个备份服务器同时读取,从而提高了数据访问速度。

更强的可扩展性:分布式数据库可以通过增添存储节点来实现存储容量的线性扩展,而集中式数据库的可扩展性十分有限。

更高的并发访问量:分布式数据库由于采用多台主机组成存储集群,所以相对集中式数据库,它可以提供更高的用户并发访问量。

■ 11.1.3 云数据库与传统的分布式数据库

云数据库和传统的分布式数据库具有相似之处，比如，都把数据存放到不同的节点上。

不同：

分布式数据库随着节点的增加会导致性能快速下降。云数据库具有很好的可扩展性。

使用方式上，云数据库也不同于传统的分布式数据库。云数据库通常采用多租户模式。

■ 11.2.1 大数据的介绍

■ 1. 大数据概念

大数据是一个体量特别大，数据类别特别大的数据集，并且无法在一定时间 范围内用常规软件工具进行捕捉、管理和处理的数据集合。大数据需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力以适应海量、高增长率和多样化的信息资产。

■ 11.2.1 大数据的介绍

■ 2. 大数据特点

业界通常用4个V(即Volume、Variety、Value、Velocity)来概括大数据的特征。

新说法(5V)：**Large Volume**(大体量)、**High Velocity**(高速度)、**Mary Varieties**(多类别)、**Difficult to Verify**(难验证)和**High Value**(高价值)。

■ 11.2.1 大数据的介绍

■ 3. 大数据技术

■ 数据采集：ETL工具负责将分布的、异构数据源中的数据如关系数据、平面数据文件等抽取到临时中间层后进行清洗、转换、集成，最后加载到数据仓库或数据集市，成为联机分析处理、数据挖掘的基础。

■ 数据存取：关系数据库、NOSQL、SQL等。

■ 基础架构：云存储、分布式文件存储等。

■ 11.2.1 大数据的介绍

■ 3. 大数据技术

■ 数据处理：自然语言处理(NLP, NaturalLanguageProcessing)是研究人与计算机交互的语言问题的一门学科。

■ 统计分析：假设检验、显著性检验、差异分析、相关分析、T检验、方差分析、卡方分析、偏相关分析、距离分析、回归分析、简单回归分析、多元回归分析、逐步回归、回归预测与残差分析、岭回归、logistic回归分析、曲线估计、因子分析、聚类分析、主成分分析等

■ 11.2.1 大数据的介绍

■ 3. 大数据技术

■ 数据挖掘：分类（Classification）、估计（Estimation）、预测（Prediction）、相关性分组或关联规则（Affinity grouping or association rules）、聚类（Clustering）、描述和可视化（Description and Visualization）、复杂数据类型挖掘等。

■ 模型预测：预测模型、机器学习、建模仿真。

■ 结果呈现：云计算、标签云、关系图等。

■ 11. 2. 2主动数据库的介绍

■ 1. 主动数据库

主动数据库 (Active Database) 是指在没有用户干预的情况下，能够主动地对系统内部或外部所产生的事件作出反应的数据库，是数据库技术和人工智能技术相结合的产物。

■ 11.2.2 主动数据库的介绍

■ 2. 主动数据库的功能模块

- 数据模型：在传统数据库中主要指层次模型、网状模型和关系模型等描述和处理实体间联系的方法，而在主动数据库中主要指知识模型。
- 执行模型：处理和执行主动规则的方式。
- 条件检测：如何检测规则的条件。
- 事务调度：如何控制事务执行的次序，使数据库状态满足完整性、一致性等要求。

■ 11.2.2主动数据库的介绍

■ 2.1知识模型

- ①指在主动数据库管理系统中描述、存储和管理ECA规则的方法。
- ②支持有关时间的约束条件。
- ③传统数据库为实现复杂的参照完整性和数据一致性引进了触发器，但只能描述“单个关系”的更新，且执行方式单一，条件的检查、动作的执行总是在触发后立即执行或事务提交前执行。

■ 11.2.2主动数据库的介绍

■ 2.2执行模型

- ①指ECA规则的处理、执行方式。
- ②提出了立即式、延迟式、隔离式等执行ECA规则的方式，克服了DBMS中触发器只能顺序执行规则的不足。
- ③主动规则执行分为五个阶段：

■ 11.2.2 主动数据库的介绍

■ 2.2 执行模型

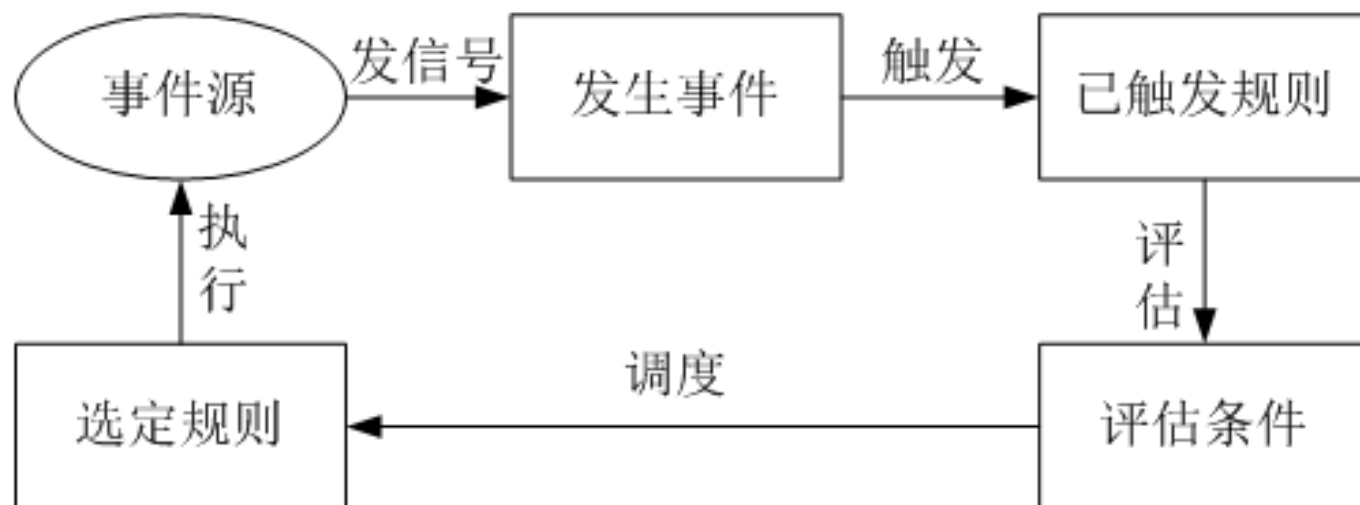


图 11-1 主动数据库执行模型

■ 11.2.2 主动数据库的介绍

■ 2.3 事务调度

- ①指如何控制事务的执行次序，使得事务满足一定的约束条件。
- ②传统DBMS中并发事务的调度应满足可串行化以保证数据库的一致性。
- ③主动数据库中还需满足事务时间方面的要求。

■ 11.2.2 主动数据库的介绍

■ 3. 主动数据库一般模型

在功能结构上，一个主动数据库系统(ADBS)由一个传统数据库系统(DBS)和一个事件驱动的知识库(EB)和相应的事件监视器(EM)组成。即： $ADBMS = DBS + EB + EM$

■ 11.2.2 主动数据库的介绍

■ 3. 主动数据库一般模型

$$ADBMS = DBS + EB + EM$$

DBS: 存储数据和对数据进行维护、管理和运用。

EB: 事件驱动知识库，是一组由事件驱动的知识的集合。

EM: 是一个随时监视EB中事件是否发生的监视模块。

■ 11.2.2 主动数据库的介绍

■ 4. 主动数据库系统体系结构

- ①事件监测器：确定规则所关心的事件是否发生。原始事件由数据库或外部源提供，复合通报是指原始事件加上从历史记录中获得的已发生事件的信息。
- ②条件评估器：评估与被监测事件相关联的规则的条件部分。
- ③调度器：比较当前被触发规则与先前被触发规则，修改冲突集，触发被调度为立即处理的规则。
- ④查询执行器：执行数据库查询或动作。

■ 11.3.1 数据仓库

■ 1. 数据仓库的概念

数据仓库（DataWarehouse），是企业所有级别的决策制定过程，提供所有类型数据支持的战略集合。它是单个数据存储，出于分析性报告和决策支持目的而创建。

■ 2. 主要特点

①面向主题的。

②集成的。

③时变的。

④非易失的。

■ 11.3.1 数据仓库

■ 3. 数据仓库的结构

数据仓库包括三部分内容：

数据层实现对企业操作数据的抽取、转换、清洗和汇总，形成信息数据，并存储在企业级的中心信息数据库中；

应用层通过联机分析处理，甚至是数据挖掘等应用处理，实现对信息数据的分析；

表现层通过对前台分析工具，将查询报表、统计分析、多维联机分析和数据挖掘的结论展现在用户面前。

■ 11.3.1 数据仓库

■ 4. 数据仓库的类型

数据仓库可分为企业数据仓库（EDW）、操作型数据库（ODS）和数据集市（datamart）三种类型。

■ 11.3.2 数据挖掘

■ 1. 数据挖掘概念

数据挖掘(Data Mining)也称为数据库中的**知识发现**。从技术角度考虑, 是从大型数据集中发现可行信息的过程, 是通过分析每个数据, 从大量数据中寻找其规律的技术, 主要有**数据准备**、**规律寻找**和**规律表示**3个步骤。

■ 11.3.2数据挖掘

■ 2. 数据挖掘的分类

按照数据库类型分类：根据数据模型不同可分为关系型、面向对象型和数据仓库型数据挖掘技术；根据应用类型不同分为空间、时间序列、文本、多媒体数据挖掘技术。

按照知识类型分类：根据数据挖掘功能分为特征化、区分、关联、聚类、孤立点分析、演变分析等数据挖掘技术。

■ 11.3.2 数据挖掘

■ 2. 数据挖掘的分类

按照所有的技术分类：根据用户交互程序的不同，例如自动系统、交互探索系统、查询驱动系统等，所用的数据分析方法分为面向对象数据库技术、数据仓库技术、统计学方法和神经网络方法等来进行描述。

按照应用分类：通常会根据应用系统的需求与特点确定数据挖掘的类型

。

■ 11.3.2 数据挖掘

■ 3. 数据挖掘的步骤

数据挖掘完整的步骤如下：

- ① 理解数据和数据的来源（understanding）。
- ② 获取相关知识与技术（acquisition）。
- ③ 整合与检查数据（integration and checking）。
- ④ 去除错误或不一致的数据（data cleaning）。
- ⑤ 建立模型和假设（model and hypothesis development）。
- ⑥ 实际数据挖掘工作（data mining）。
- ⑦ 测试和验证挖掘结果（testing and verification）。
- ⑧ 解释和应用（interpretation and use）。

■ 11.3.2 数据挖掘

■ 4. 数据挖掘的方法

数据挖掘完整的步骤如下：

- ① 分类
- ② 回归分析
- ③ 聚类
- ④ 关联规则
- ⑤ 神经网络方法
- ⑥ web数据挖掘

■ 11.3.3数据仓库与数据挖掘的区别

数据挖掘就是从大量数据中提取数据的过程。

数据仓库是汇集所有相关数据的一个过程。

数据挖掘和数据仓库都是商业智能工具集合。

数据挖掘是特定的数据收集。

数据仓库是一个工具来节省时间和提高效率，将数据从不同的位置不同区域组织在一起。

■ 11.4.1概述

各种学科与数据库技术的有机结合，使数据库领域中新内容、新应用、新技术层出不穷，形成了各种各样的数据库系统：**面向对象数据库系统、多媒体数据库系统、并行数据库系统、移动数据库、分布式数据库系统。**

数据库系统被应用到特定的领域后，又出现了空间数据库、工程数据库、时态数据库、科学数据库、文献数据库等；

■ 11.4.2空间数据库

■ 1. 空间数据结构

空间数据（Spatial Database）是用于表示空间物体的位置、大小、形状、和分布特征等诸方面信息的数据，适用于描述所有二维、三维和多维分布的关于区域的数据。它的特点既包括物体本身的空间位置及状态信息，又包括表示物体的空间关系（即拓扑关系）的信息。

空间数据库常用的空间数据结构有**矢量数据结构**和**栅格数据结构**两种。

■ 11.4.2空间数据库

■ 2. 空间数据查询语言

空间数据查询包括位置查询、空间关系查询和属性查询等。

前两种查询是空间数据库特有的，基本方式有：面-面查询、线-线查询、点-点查询、线-面查询，点-线查询、点-面查询等。

■ 11.4.2空间数据库

■ 3. 空间数据库管理系统功能

- ①提供对空间数据和空间关系的定义和描述;
- ②提供空间数据查询语言, 实现对空间数据的高效查询和操作;
- ③提供对空间数据的存储和组织;
- ④提供对空间数据的直观显示等。

■ 11.4.3多媒体数据库

■ 1. 多媒体数据库概念和特征

多媒体数据库（Multimedia Database）是数据库技术与多媒体技术结合的产物。实现对格式化和非格式化的多媒体数据的存储、管理和查询，其主要特征有：

- ①多媒体数据库应能够表示多种媒体的数据。
- ②多媒体数据库应能够支持大对象。
- ③多媒体数据库应能够协调处理各种媒体数据，正确识别各种媒体数据之间在空间或时间上的关联。
- ④多媒体数据库应提供比传统数据管理系统更强的、适合非格式化数据查询的搜索功能。

■ 11.4.3多媒体数据库

■ 1. 多媒体数据模型

多媒体数据模型有三种：

- ①扩充关系模型
- ②语义数据模型
- ③面向对象模型

■ 11.4.4面向对象数据库

面向对象数据库是面向对象的程序设计技术与数据库技术相结合的产物，是为了满足新的数据库应用需要而产生的新一代数据库系统。

■ 1. 面向对象数据库的特点

- (1) 易维护。
- (2) 质量高。
- (3) 效率高。
- (4) 易扩展。

■ 11.4.4面向对象数据库

■ 2面向对象数据库与传统数据库的区别

- ①面向对象模型是一种层次式的结构模型；
- ②面向对象数据模型是将数据与操作封装于一体的结构方式；
- ③面向对象数据模型具有构造多种复杂抽象数据类型的能力；
- ④面向对象数据模型具有不断更新结构的模式演化能力。

■ 11.4.5移动数据库

移动数据库（Mobile Database）是指在移动计算环境中的分布式数据库，其数据在物理上分散而在逻辑上集中。它涉及数据库技术、分布式计算技术、移动通信技术等多个学科领域。

■ 11.4.6并行数据库

并行数据库（Parallel Database）是利用并行计算机技术使数个、数十、甚至成千上百个廉价的计算机协同工作，实现并行数据管理和并行查询功能，提供一个高性能、高可靠性、高扩展性的数据库管理系统，能够快速查询大数据量并处理大量数的事务。

■ 11.4.7工程数据库

工程数据库（Engineering Database）是一种能管理和
管理各种工程图形，并能为工程设计，是将工程设计方法、
人工智能技术与数据库技术相结合发展起来的智能化集成
系统，适合于CAD/CAM、计算机集成制造（CIM）等工程应
用领域。

■ 11.4.8其他数据库

实时数据库（Real-Time DataBase）是数据库技术结合实时处理技术产生的。它适用于处理不断更新的快速变化的数据及具有时间限制的事务处理。实时数据库的一个重要特性就是实时性，包括数据实时性和事务实时性。针对不同行业不同类型的企业，实时数据库的数据来源方式也各不相同。总的来说数据的主要来源有DCS控制系统、由组态软件与PLC建立的控制系统、数据采集系统（SCADA）、关系数据库系统、直接连接硬件设备和通过人机界面人工录入的数据。

■ 数据库新技术发展的主流趋势

- 非结构化数据库
- 数据库技术与多学科技术的有机结合
- 数据库技术及市场发展的有机结合
- 面向专门应用领域的数据库技术

■ 数据库新技术发展的特征

- 支持XML数据格式
- 传感器数据库技术
- 微小型数据库技术
- 信息集成
- 网格数据管理
- 移动数据管理

国家“十三五”重点出版规划项目
上海高校优秀教材奖获得者主编

上海市高校精品课程
特色教材

诚挚谢意！



数据库原理及应用

——基于SQL Server 2016