第17章数据库新技术与新应用

北京理工大学 计算机学院 张文耀

zhwenyao@bit.edu.cn



- 信息管理内容的不断扩展和新技术的层出不穷,数据库技术面临着前所未有的挑战
 - 应用领域由企业内变成跨企业
 - 数据管理由简单数据发展到复杂数据
 - 新技术与数据库技术的结合
- 面对新的数据形式,人们提出了丰富多样的数据模型(层次模型、网状模型、关系模型、面向对象模型、半结构化模型等),同时也提出了众多新的数据库技术(XML 数据管理、数据流管理、Web 数据集成、数据挖掘等)。

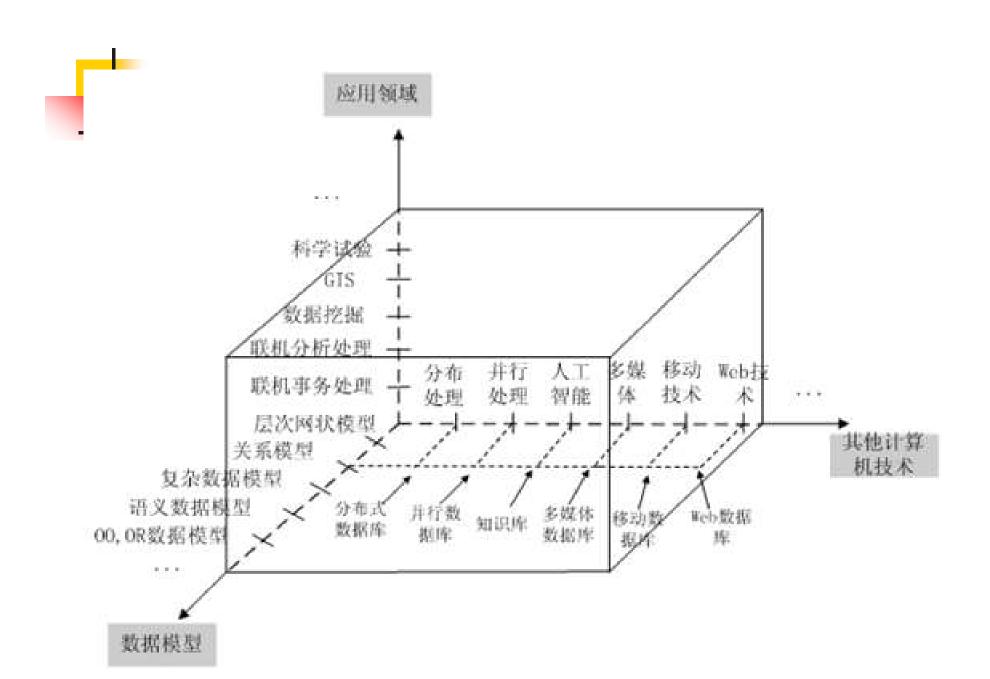
- 4
 - 建立和实现了一系列新型数据库系统,如分布式数据库系统、面向对象数据库系统、演绎数据库系统、统、知识库系统、多媒体数据库系统等,它们共同构成了数据库系统的大家族。
 - 传统的数据库系统仅是数据库大家族的一员,当然,它也是最成熟的和应用最广泛的一员。它的核心理论、应用经验、设计方法等仍然是整个数据库技术发展和应用开发的先导和基础。



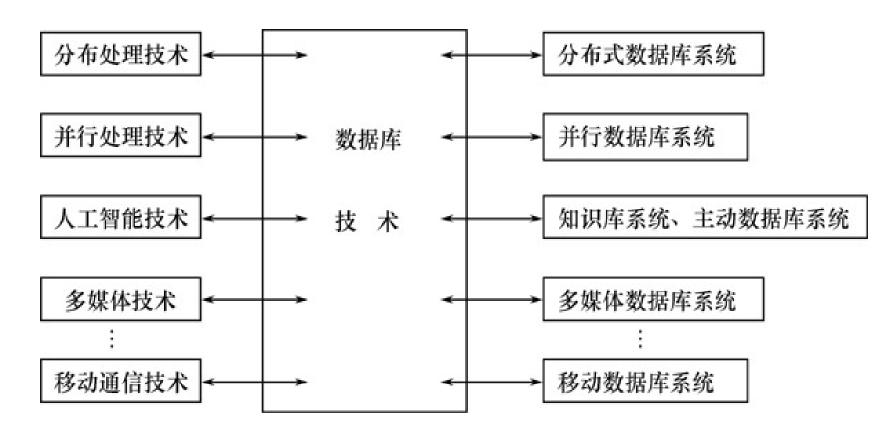
- 数据库新技术的发展可以划分为以下几类
 - 在整体上有较大的技术改进
 - 面向对象数据库技术
 - 实时数据库技术
 - 主动数据库技术
 - 时态数据库技术
 - 在体系结构方面集成新技术
 - 内存数据库技术
 - 并行数据库技术
 - 分布式数据库技术
 - 数据挖掘技术



- 在应用功能方面进行增强
 - 工程数据库技术
 - 统计数据库技术
 - 空间数据库技术
 - 知识数据库技术
 - 多媒体数据库技术









- 发展对象数据库的原因
 - 关系数据库的数据模型和系统在设计和实现复杂的数据应用时往往不能满足应用要求。
 - 关系模型的语义表达能力有限。
- 面向对象数据库(OODBMS)是数据库技术与面向 对象技术结合的产物。
 - 既有持久性、存储管理、数据共享、数据可靠性、查询 处理、模式修改等DBMS的特性,
 - 又具有面向对象的封装性、继承性、重载、对象标识、 复合对象和可扩充等特性。



- ■面向对象数据库系统的研究方法
 - 开发全新的数据模型,建立完全的面向对象数据库系统。但纯粹面向对象数据库系统并不支持SQL语言,应用领域受到局限
 - 对传统的关系数据库加以扩展,增加面向对象特性。建立对象关系数据库管理系统。
 这种系统既支持已被广泛使用的SQL,具有通用性,又支持复杂对象和复杂对象行为



- 实时数据库是实时系统和数据库技术想结合的产物。
 - 利用数据库技术来解决实时系统中的数据管理问题;
 - 利用实时技术为实时数据库提供时间驱动调度和资源分配算法;
- 实时数据库并非实时系统和数据库的简单集成。
- 实时数据库是数据和事务都有显式定时限制的数据库,系统的正确性不仅依赖于事务的逻辑结果,而且依赖于逻辑结果所生产的时间。



- 实时数据库是数据库系统发展的一个分支,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。
- 实时数据库主要的理论研究有:实时数据建模、实时事务调度与资源分配策略、实时数据查询语言、实时数据通信等。



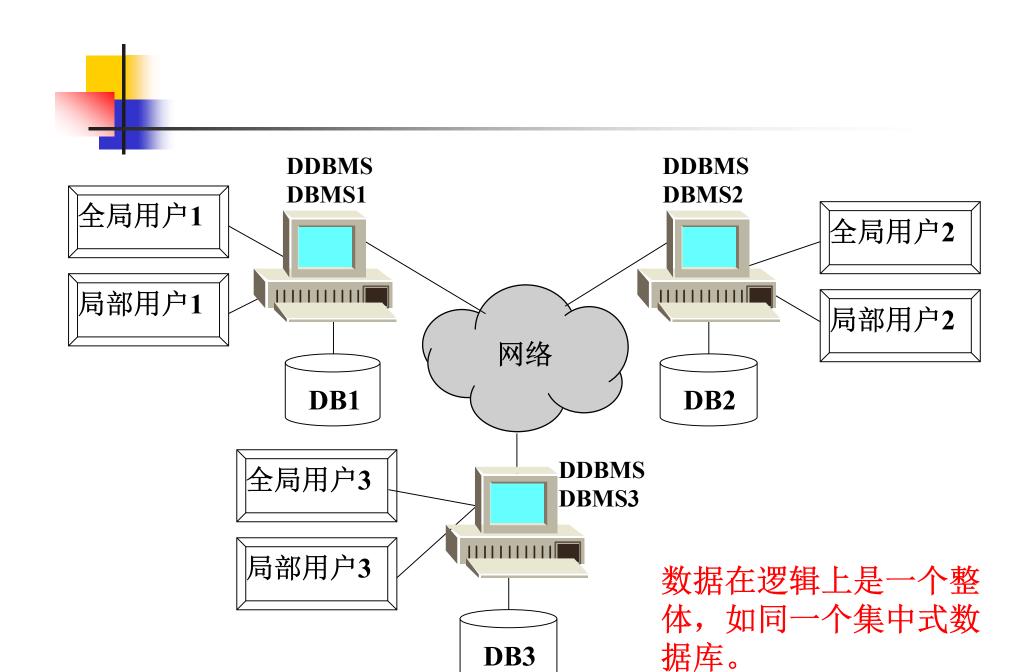
- 主动数据库是相对于传统数据库的被动性而言的。
- 在实际应用中,如计算机集成制造系统、管理信息系统、办公自动化系统中常常希望数据库系统在紧急情况下能根据数据库的当前状态,主动地做出反应,执行某些操作,向用户提供有关信息。
- 主动数据库是传统数据库与人工智能技术和面向对象技术结合起来的产物。



- 主动数据库通常采用在传统数据库系统中嵌入事件—条件—动作规则,相当于系统提供了一个"自动监测"机构,主动地不时地检查这些规则中包含的事件是否已发生,一旦事件被发现,就主动触发执行相应的动作。
- 主动数据库主要解决的问题有知识模型、执行模型、条件检测、事务调度、体系结构、系统效率。



- 分布式数据库是数据库技术与计算机网络技术相结合的产物。
- 分布式数据库系统是将分散在各处的数据库系统通过计算机网络连接起来。这种系统由多个数据库服务器组成,各服务器之间由通信网络相互联系,从系统角度来看,这些数据在逻辑上是一个整体,如同一个集中式数据库。
- 关于分布式数据库系统可参见第14章。





- 数据仓库技术是从数据库技术发展而来的,是面向 主题的、稳定的、综合的、随时间变化的数据集 合。
- 数据挖掘是从超大型数据库或数据仓库中发现并提取隐藏在内部的信息的一种新技术,其目的是帮助决策者寻找数据间潜在的关联,发现被经营者忽略的要素,从而作出正确的决策。
- 大数据分析与处理

多媒体数据库

- 多媒体数据库是数据库技术与多媒体技术相结合的 产物。
- 多媒体数据库应具备的功能有:能表示和理解多媒体数据,能刻画、管理和表现各种媒体数据的特性和相互关系;具备物理数据独立性和媒体数据独立性,媒体类型可扩展;提供更为灵活的模式定义和修改功能,支持模式进化与演变,具备某些长事务处理的能力;提供多媒体访问的各种手段,近似性查询,混合方式访问等。
- 多媒体数据库是数据库主要发展方向之一。



- 并行数据库将传统的数据库管理技术与并行处理技术相结合,以提供高性能、高可用性与高可扩展性的一种数据库。
- 并行数据库能充分利用多处理器的能力,通过多种 并行性,在联机事务处理与联机分析处理应用中提 供优化的响应时间与事务吞吐量。
- 目前,并行数据库已经得到了较为广泛的应用,许 多主商用DBMS都支持并行数据库。



- 并行数据库系统的体系结构有以下三种
 - 共享内存(Shared-Memory)结构
 - 共享磁盘(Shared-Disk)结构
 - 无共享资源(Shard-Nothing)结构
- 一个理想的并行数据库系统应能充分利用硬件平台的并行性,采用多进程或多线程结构,提供四种不同粒度的并行性:
 - 不同用户事务间的并行性
 - 同一事务内不同查询间的并行性
 - 同一查询内不同操作间的并行性
 - ■同一操作内的并行性

空间数据库

- 空间数据库(Spatial Database)是一种为存储、管理和查询空间相关数据(如点、线、多边形等)而优化的一种数据库。
- 空间数据则是各种空间对象特征和关系的符号化表示,包括空间位置、属性特征(简称属性)及时域特征三部分。
- 空间数据呈现不定长的非结构化特征,不满足关系数据库理论中1NF要求。
- 空间数据库是作为一种应用技术而诞生和发展起来的,其目的是为了使用户能够方便灵活地编辑和查询所需的空间数据。

- 4
 - 地理信息系统(Geographic Information System,GIS)是对地理分布数据进行采集、储存、管理、运算、分析、显示和描述的计算机软件系统;处理、管理的对象是多种地理空间实体数据及其关系,包括空间定位数据、图形数据、遥感图像数据、属性数据等。
 - GIS可以说是建立在空间数据库管理系统上的应用信息系统。但传统意义上的GIS与空间数据库管理系统很难划清明显的界限,所以空间数据管理系统和GIS的概念经常可以互相交换。

XML数据库

- XML数据库一种可以对XML文档进行存取管理和 数据查询的数据库。
- XML文档——半结构化的数据

```
<Customer>
     <Name>ABC Industries</Name>
     <Address>
          <Street>123 Main St. 
          <City>Fooville</City>
          <State>CA</State>
          <Country>USA</Country>
          <PostCode>95041</PostCode>
     </Address>
 Customer>
```



- XML文档存储的方式有4种:
 - 关系数据库方式
 - ■面向对象数据库方式
 - 文本方式
 - XML数据的查询和更新都很困难
 - XML数据库方式

- 4
 - XML数据库为XML文档定义了一个逻辑模型并根据这个模型存取数据。这个模型包括元素、属性、PCDATA和文档顺序。
 - XML数据库以XML文档作为基本的逻辑存储单位,它与关系数据库中的记录类似。
 - XML数据库对底层的物理存储结构没有固定的要求,它可以建立在关系数据库、面向对象数据库的基础上,也可以采用特定的存储格式。

■ 主流的商用DBMS(如Oracle、SQL Server、DB2)中都增加了支持XML数据处理的功能。这些数据库被称为支持XML的数据库(XML-enables DBMS),与此对应的是一种专为处理XML而开发的数据库,这种数据库被称为原生XML数据库(Native XML DBMS)。

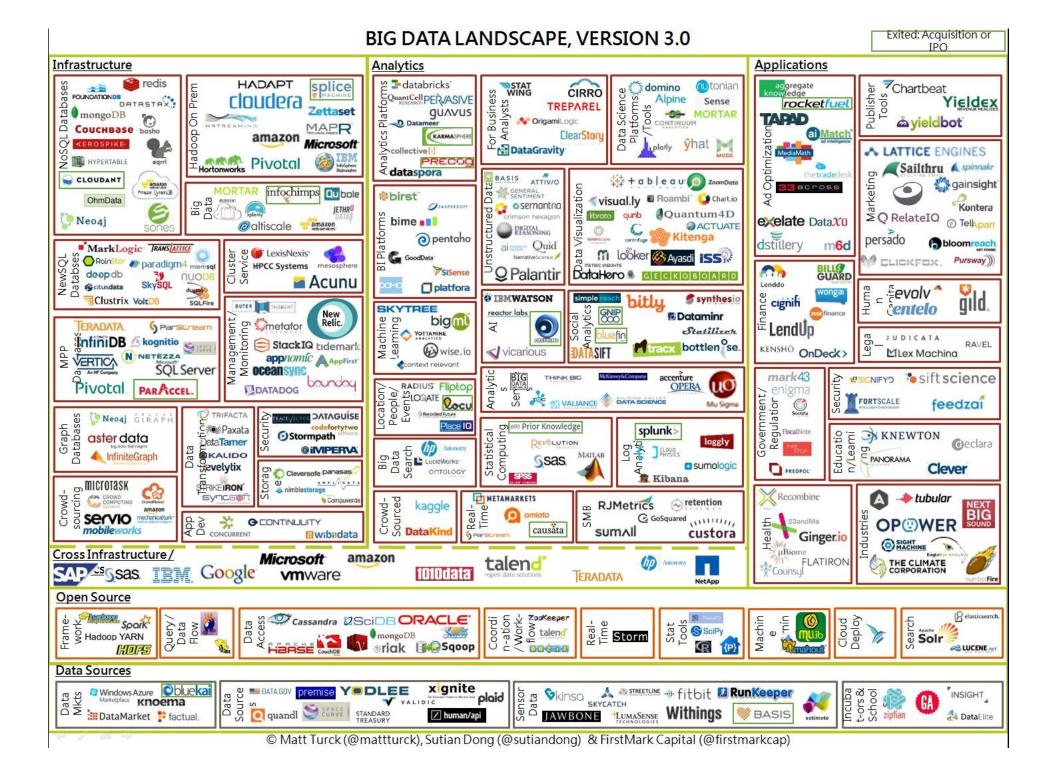


- 原生XML数据库(简称XML数据库)
 - 它为XML文档(而不是文档中的数据)定义了一个(逻辑)模型,并根据该模型存取文件。这个模型至少应包括元素、属性、PCDATA和文件顺序。
 - 它以XML文件作为其基本(逻辑)存储单位,正如关系 数据库以表中的行作为基本逻辑存储单位。
 - 它对底层的物理存储模型模型没有特殊要求。例如,它可以建在关系型、层次型或面向对象的数据库之上,或者使用专用的存储格式,比如索引或压缩文件。

大数据时代的数据库技术

从大的角度讲,可以简单的将数据库分为两类:

- 传统SMP架构的数据库,主要指代的是传统的关系 型数据库
 - DB2、Postgre、MySQL等。
- 新型数据库,主要指代的是支持大规模数量集、高 并发要求、高可扩展性等孕育而生的新型数据库。
 - 包括目前大数据生态当中主流MPP、NoSQL、NewSQL数据 库等。





根据对SQL的支持情况,可以将数据库分为:

- SQL Database
 - 01d SQL传统SQL数据库
 - New SQL新型SQL数据库
 - MPP
 - SQL on Hadoop
- NoSQL Database
 - Key-Value Database
 - Document Database
 - Column family Stores
 - Graph Database

内容	01dSQL	NoSQL	NewSQL
SQL支持	Good at	Give up SQL	Preserve SQL
ACID	follow ACID	Give up ACID	Preserve ACID
New OLTP support	Too slow, Does not scale	Locks consistency guarantees, Lowlevelinterface	Fast, scalable andconsistent, Su ports SQL
Scalability	hardly	easy scale out	Good scalability through modern innovative softwa e architecture
Performance	bottleneck depond on single node	Performance improved through distribute architecture	Performance improved through distribute archit cture

小结

从新型数据库类型的角度对数据库新技术进行了简要的介绍。

让数据说话,数据依旧,魅力倍增