

中国科学院研究生教学丛书

分布式数据库系统及其应用

(第二版)

邵佩英 编著

科学出版社

北京

内 容 简 介

本书在第一版的基础上做了较大的修改,增加了与分布式数据库技术密切相关的、新的网络与数据库的接口技术,基于组件的新的软件开发方法,WWW 数据库,移动数据库,数据仓库和数据挖掘等内容,反映了分布式数据库技术和应用开发的新的发展趋势。本书还增加了习题和部分习题的参考答案。本书对分布式数据库系统的基本内容提供了翔实的讲解,对一些有争议的问题进行了开放式讨论,但对深奥的理论没有采用繁琐的证明方法,而是采用大量的图示和实例做浅显的解释,从而增加了它的知识性和易理解性。

本书可作为计算机专业高年级本科生的选修课教材,特别适合作研究生的专业课教材,也可作为从事计算机信息处理研究或相关应用开发人员的参考书。

图书在版编目(CIP)数据

分布式数据库系统及其应用/邵佩英编著.-2版.北京:科学出版社,2005
(中国科学院研究生教学丛书)
ISBN 7-03-015113-5

I.分… II.邵… III.分布式数据库-数据库系统 IV.TP311.13

中国版本图书馆 CIP 数据核字(2005)第 016526 号

责任编辑:鞠丽娜/责任校对:刘彦妮

责任印制:吕春珉/封面设计:槐寿明

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

印刷

科学出版社发行 各地新华书店经销

*

2000 年 6 月第 一 版 开本:787×1092 1/16

2005 年 3 月第 二 版 印张:25 3/4

2005 年 3 月第五次印刷 字数:608 000

印数:8 001—12 000

定价:38.00 元

(如有印装质量问题,我社负责调换〈新欣〉)

销售部电话 010-62136131

编辑部电话 010-62138978-8002

《中国科学院研究生教学丛书》总编委会

主 任：白春礼

副 主 任：余翔林 师昌绪 杨 乐 汪尔康 沈允钢

黄荣辉 叶朝辉 李 佩

委 员：朱清时 匡廷云 叶大年 王 水 冯克勤

冯玉琳 刘政凯 龚 立 侯建勤

《中国科学院研究生教学丛书》

技术学科编委会

主 编：师昌绪

副 主 编：冯玉琳

编 委：刘政凯 徐至展 陈先霖 王占国 马颂德

吴承康 史忠植

《中国科学院研究生教学丛书》序

在 21 世纪曙光初露,中国科技、教育面临重大改革和蓬勃发展之际,《中国科学院研究生教学丛书》——这套凝聚了中国科学院新老科学家、研究生导师们多年心血的研究生教材面世了。相信这套丛书的出版,会在一定程度上缓解研究生教材不足的困难,对提高研究生教育质量起着积极的推动作用。

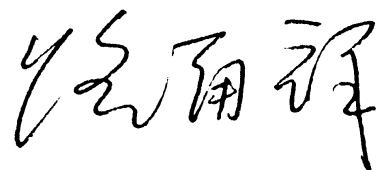
21 世纪将是科学技术日新月异,迅猛发展的新世纪,科学技术将成为经济发展的最重要的资源和不竭的动力,成为经济和社会发展的首要推动力量。世界各国之间综合国力的竞争,实质上是科技实力的竞争。而一个国家科技实力的决定因素是它所拥有的科技人才的数量和质量。我国要想在 21 世纪顺利地实施“科教兴国”和“可持续发展”战略,实现小平同志规划的第三步战略目标——把我国建设成中等发达国家,关键在于培养造就一支数量宏大、素质优良、结构合理,有能力参与国际竞争与合作的科技大军,这是摆在我国高等教育面前的一项十分繁重而光荣的战略任务。

中国科学院作为我国自然科学与高新技术的综合研究与发展中心,在建院之初就明确了出成果出人才并举的办院宗旨,长期坚持走科研与教育相结合的道路,发挥了高级科技专家多,科研条件好,科研水平高的优势,结合科研工作,积极培养研究生;在出成果的同时,为国家培养了数以万计的研究生。当前,中国科学院正在按照江泽民同志关于中国科学院要努力建设好“三个基地”的指示,在建设具有国际先进水平的科学研究基地和促进高新技术产业发展基地的同时,加强研究生教育,努力建设好高级人才培养基地,在肩负起发展我国科学技术及促进高新技术产业发展重任的同时,为国家源源不断地培养输送大批高级科技人才。

质量是研究生教育的生命,全面提高研究生培养质量是当前我国研究生教育的首要任务。研究生教材建设是提高研究生培养质量的一项重要基础性工作。由于各种原因,目前我国研究生教材的

建设滞后于研究生教育的发展。为了改变这种情况,中国科学院组织了一批在科学前沿工作,同时又具有相当教学经验的科学家撰写研究生教材,并以专项资金资助优秀的研究生教材的出版。希望通过数年努力,出版一套面向 21 世纪科技发展,体现中国科学院特色的高水平的研究生教学丛书。本丛书内容力求具有科学性、系统性和基础性,同时也兼顾前沿性,使阅读者不仅能获得相关学科的比较系统的科学基础知识,也能被引导进入当代科学研究的前沿。这套研究生教学丛书,不仅适合于在校研究生学习使用,也可以作为高校教师和专业研究人员工作和学习的参考书。

“桃李不言,下自成蹊。”我相信,通过中国科学院一批科学家的辛勤耕耘,《中国科学院研究生教学丛书》将成为我国研究生教育园地的一丛鲜花,也将似润物春雨,滋养莘莘学子的心田,把他们引向科学的殿堂,不仅为科学院,也为全国研究生教育的发展作出重要贡献。

A handwritten signature in black ink, reading 'Guo Junqiang' (郭俊强) in a cursive style.

前 言

本书作为中国科学院研究生教学丛书之一已经出版四年多了。据有关部门反映,这是一本备受欢迎的数据库系统教材。它全面、系统地阐述了分布式数据库系统的基本理论、概念原理、结构特点、存在问题、设计思想、技术方法;重点论述了分布式数据库的查询优化、事务管理、并发控制、可靠性、安全性和目录管理;讨论了客户机/服务器模式的概念、结构、工作方式以及与分布式数据库系统的关系;展望了分布式数据库系统的发展前景。

现代数据库管理技术在当前信息社会中的关键作用已经变得更加清晰,毕竟 Internet 仅仅是一个通信系统,它的真正价值是从数据库中读取或存入数据和信息。因此,随着 Internet 和无线技术的发展,特别是 WWW 建立以来,它们与分布式数据库有着极其紧密的关系。从特定意义上说,整个 Internet 就是一个巨大的“分布式数据库”。同时,随着时间的推移,本书中的内容,特别是后面部分的内容,已经有了相当大的变化。与分布式数据库技术密切相关的、新的数据库与网络的接口技术、新的软件开发方法、WWW 数据库、移动数据库的产生和应用,以及数据仓库、联机分析处理(OLAP)和数据挖掘已经成为了数据库技术研究、开发和应用最活跃的热点之一。

考虑到上述原因,作者根据本人的教学实践和有关反馈信息,对第一版中的内容做了如下几个方面的修改。

对第一至四章的修改主要是调整结构、补充和进一步完善有关内容。

对第五至七章,除修改、补充和进一步完善外,新增加的内容包括分布式事务管理的可串行化调度测试,两阶段封锁协议的实现方法,基于时间戳排序的多版本技术,采用验证锁的多版本两阶段封锁技术;关于分布式可靠性协议中的网络分割与提交协议;分布式数据库安全中的用户账户和数据库审计,统计数据库的安全性,数据加密技术,目录管理的重要性等内容。

对第八章的修改除调整结构、补充和进一步完善外,新增加了客户机/服务器计算模式的远景展望,创建基于客户机/服务器模式系统的自动化设计工具,创建基于客户机/服务器模式系统的技术和方法,基于组件的软件开发技术,异构数据源访问接口技术和方法等内容。

删除了第一版的第九章,本书新增加的第九章是:分布式数据库与 WWW 数据库和移动数据库。这两类数据库与分布式数据库有着极其紧密的关系,研究支持它们的基本技术,并考察它们将会如何影响分布式数据库系统。

对第十章的修改除调整结构、补充和进一步完善外,新增加了对象-关系数据库系统的产生和发展,重点介绍 SQL3 语言对对象-关系数据库系统的支持,讨论嵌套关系模型。

本书增加了第十一章:数据仓库与数据挖掘。本章概括地论述了数据库技术的最新应用领域:数据仓库、联机分析处理和数据挖掘的关键概念、数据模型、功能结构、技术方法,通过实例给出支持决策的方法,以及介绍了一些预期的研究方向。数据仓库可以看成是要

求事先进行多种活动的一种过程;数据挖掘可以认为是从现存的数据仓库中发现知识的一个活动。

本书增加的章后练习习题涉及了概念、理论、算法、技术和方法。给出章后练习习题的目的在于使学生们能对每一章学到的知识进行复习、回顾和系统化,有的习题还要求将所学知识应用到一个小的、实际的,但有约束的例子中,以进一步巩固所学的知识。

本书第二版继续保持了第一版的特点。所有概念都以一种技术上完备,而又易于理解的方式表达,并尽量减少各种符号的使用。它没有采用繁琐的证明方法,而是采用了大量的图示和实例描述概念和预期的结果,从而进一步增加了它的知识性和易理解性。本书内容涵盖了分布式数据库研究领域的各个关键问题,并以真实的实例系统引入分布式数据库的设计技术和方法。书中对分布式数据库技术的基本内容提供翔实的讲解,对深奥的理论做了浅显的解释,对一些有争议的问题进行了开放式的讨论,并充分反映了分布式数据库技术和应用开发的新的发展趋势。通过本书的学习,读者可以对分布式数据库系统的整体结构、设计技术和关键问题有深入的认识和理解,同时可以熟识分布式数据库系统的理论基础和应用方法。

作者认为应当用一年的时间来学习这门课程。但是如果仅仅有一个学期的时间,可以选择本书的第一至七章进行系统学习,而将八至十一章作为参考资料。

本书适合作为计算机专业高年级本科生的选修课教材,特别适合作为研究生的专业课教材,还可作为高校教师及从事计算机信息处理研究或相关应用开发人员工作和学习的参考资料。在学习本书之前,读者应具有数据库的基本概念和基础知识。

罗晓沛、孙淑玲、李伯民、陈伯飞仔细审阅了全书;宋纲、陈铭、王辉、祝孔强、阚劲松、颜涯等同学为本书提供有关资料。科学出版社的有关工作人员为本书的出版做了大量工作,在此一并表示衷心的感谢。

邵佩英

2004年12月

目 录

第一章 分布式数据库系统概述	1
1.1 分布式数据库系统的由来与发展	1
1.1.1 分布式数据库系统的由来	1
1.1.2 分布式数据库系统的发展	1
1.2 分布式数据库系统的定义与分类	4
1.2.1 分布式数据库系统的定义	4
1.2.2 分布式数据库系统的分类	6
1.3 分布式数据库系统的体系结构和组成成分	8
1.3.1 分布式数据库系统的体系结构	8
1.3.2 分布式数据库的组成成分	9
1.3.3 分布式数据库中数据的分片与分布	9
1.3.4 分布式数据库的模式结构.....	12
1.3.5 分布式数据库管理系统的功能结构	15
1.3.6 分布式数据库管理系统的一种参考模型	17
1.4 分布式数据库系统中数据的独立性与分布透明性.....	25
1.4.1 分布式数据库系统中数据的独立性与分布透明性概述	25
1.4.2 分布式数据库简单查询与分布透明性	26
1.4.3 分布式数据库复杂查询与分布透明性	29
1.4.4 分布式数据库更新应用与分布透明性	31
1.5 分布式数据库系统的优点和存在的技术问题.....	34
1.5.1 分布式数据库系统的优点.....	34
1.5.2 分布式数据库系统中存在的技术问题	35
1.6 本章小结.....	37
习题	38
第二章 分布式数据库系统的设计	39
2.1 分布式数据库系统设计概述.....	39
2.1.1 分布式数据库系统的创建方法	39
2.1.2 分布式数据库系统设计的内容	41
2.1.3 分布式数据库设计的目标.....	41
2.1.4 分布式数据库设计的方法.....	43
2.2 自顶向下设计分布式数据库.....	44
2.2.1 自顶向下设计分布式数据库的步骤和内容.....	44
2.2.2 数据的分片设计	44

2.2.3 数据库片段的位置分配设计	49
2.3 DATAID-D 方法	52
2.3.1 DATAID-D 方法概述	52
2.3.2 分布要求分析阶段	53
2.3.3 分布设计阶段	54
2.4 实例研究:飞机订票系统	55
2.4.1 实例研究概述	55
2.4.2 飞机订票系统中的分布要求分析	57
2.4.3 飞机订票系统中的分布设计	59
2.5 自底向上设计分布式数据库	61
2.5.1 自底向上设计分布式数据库要解决的问题	61
2.5.2 构造全局模式的设计问题和解决方法	62
2.5.3 自底向上综合的一个例子	65
2.6 本章小结	66
习题	67
第三章 分布式数据库中的查询处理和优化	69
3.1 分布式查询优化概述	69
3.1.1 分布式查询优化的目标	69
3.1.2 分布式查询优化的准则和代价估算	69
3.1.3 分布式查询策略的重要性	70
3.2 分布式查询优化中的基础知识	72
3.2.1 用关系代数表达式和 SQL 语句表示一个查询	72
3.2.2 查询树	73
3.2.3 等价变换规则的概念和术语	74
3.2.4 等价变换规则	75
3.3 分布式查询的分类与层次结构	76
3.3.1 分布式查询的分类	76
3.3.2 分布式查询处理的层次结构	78
3.4 基于关系代数等价变换的查询优化处理	79
3.4.1 基于关系代数等价变换查询优化处理的基本原理和实现方法	79
3.4.2 基于关系代数等价变换的查询优化处理举例	80
3.5 基于半连接算法的查询优化处理	83
3.5.1 采用半连接方法表示连接操作	83
3.5.2 采用半连接算法优化连接操作的基本原理和步骤	84
3.6 基于直接连接算法的查询优化处理	85
3.6.1 利用站点依赖信息的算法	85
3.6.2 分片和复制算法	87
3.6.3 站点依赖和数据复制结合	89

3.6.4 Hash 划分算法	90
3.6.5 不同方法的比较	92
3.7 直接连接操作的常用策略	92
3.7.1 直接连接操作的一般常用策略	92
3.7.2 利用并行性的直接连接操作策略	93
3.8 本章小结	94
习题	95
第四章 分布式数据库中的事务管理和恢复	97
4.1 分布式事务概述	97
4.1.1 分布式事务定义和特性	97
4.1.2 分布式事务的结构和事务状态	99
4.1.3 分布式事务管理的问题和目标	102
4.2 分布式事务的执行与恢复	104
4.2.1 分布式事务管理的抽象模型	104
4.2.2 分布式事务执行的控制模型	105
4.2.3 分布式数据库系统中的故障	107
4.2.4 事务故障恢复的基本概念	108
4.2.5 事务故障的恢复	111
4.2.6 分布式事务的执行与恢复举例	113
4.3 两阶段提交协议	114
4.3.1 两阶段提交协议的基本思想和内容	114
4.3.2 两阶段提交协议的通信结构	117
4.3.3 两阶段提交协议与故障恢复	122
4.4 分布式数据库中的数据更新	124
4.4.1 多站点的数据更新	124
4.4.2 主文本更新法	125
4.4.3 快照方法	126
4.5 分布式事务增强数据库一致性	127
4.5.1 业务规则的一致性	127
4.5.2 冗余数据的一致性	128
4.6 本章小结	130
习题	130
第五章 分布式数据库中的并发控制	131
5.1 并发控制的概念和理论	131
5.1.1 并发控制的概念	131
5.1.2 事务可串行化理论的基本概念	133
5.1.3 分布式事务的可串行化理论	134
5.1.4 分布式事务的可串行化调度测试	136

5.1.5	并发控制机制的常用方法及其分类	139
5.2	分布式数据库系统并发控制的封锁技术	141
5.2.1	基于封锁的并发控制方法概述	141
5.2.2	两阶段封锁协议	147
5.2.3	两阶段封锁协议的实现方法	150
5.2.3	多粒度封锁与意想锁	152
5.3	分布式数据库系统中的死锁处理	156
5.3.1	全局死锁与等待图	156
5.3.2	死锁的预防方法	158
5.3.3	死锁的检测和解决方法	159
5.4	分布式数据库系统并发控制的时标技术	162
5.4.1	基于时标的并发控制方法	162
5.4.2	基本时标法	164
5.4.3	保守时标法	164
5.5	分布式数据库系统并发控制的多版本技术	166
5.5.1	基于时间戳排序的多版本技术	166
5.5.2	采用验证锁的多版本两阶段封锁	167
5.6	分布式数据库系统并发控制的乐观方法	168
5.7	本章小结	170
	习题	171
第六章	分布式数据库中的可靠性	173
6.1	分布式数据库可靠性的概念及其度量	173
6.1.1	分布式数据库可靠性的概念	173
6.1.2	平均故障间隔时间和平均修复时间	174
6.2	分布式数据库系统的故障原因和容错技术	175
6.2.1	分布式数据库系统中系统失败的原因	175
6.2.2	基本的容错方法和技术	177
6.3	分布式数据库的可靠性协议	179
6.3.1	分布式数据库可靠性协议的组成	179
6.3.2	两阶段提交协议的演变	180
6.3.3	事务阻断与终结协议	181
6.3.4	两阶段提交协议的终结协议	182
6.3.5	两阶段提交协议的恢复协议	185
6.3.6	三阶段提交协议	187
6.4	网络分割与提交协议	190
6.4.1	网络分割概述	190
6.4.2	网络分割的提交协议	191
6.5	不一致性的检测和解决方法	197

6.5.1	决定网络的状态	197
6.5.2	不一致性的检测和解决方法	199
6.5.3	检查点和冷启动	201
6.6	本章小结	203
	习题	204
第七章	分布式数据库的安全性 与目录管理	206
7.1	数据库安全性概述	206
7.1.1	数据库安全性的概念	206
7.1.2	分布式数据库的不安全因素	207
7.1.3	分布式数据库安全需求和安全措施分析	208
7.1.4	分布式数据库的用户账户和数据库审计	210
7.2	安全数据模型与多级安全数据库	211
7.2.1	数据库安全术语与基本概念	211
7.2.2	基于授予/收回权限的自主访问控制	213
7.2.3	多级安全 BLP 模型	217
7.2.4	基于多级安全性分类级别标记的强制访问控制	219
7.3	统计数据库的安全性	222
7.4	数据加密	223
7.4.1	数据加密概述	223
7.4.2	公开密钥加密算法和数据签名	224
7.5	计算机系统与数据库管理系统的安全评估标准	226
7.5.1	计算机系统的安全评估标准	226
7.5.2	数据库管理系统的安全评估标准	227
7.5.3	当前流行的几种 RDBMS 安全机制	228
7.6	分布式数据库的目录结构和管理	231
7.6.1	分布式数据库目录的重要性	231
7.6.2	分布式数据库目录的内容及用途	232
7.6.3	分布式数据库目录系统的组织方式及逻辑结构	234
7.6.4	分布式数据库目录的分布方式	236
7.6.5	具有站点自治性的对象命名和目录管理	237
7.7	分布式数据库中权限保护和用户识别	238
7.7.1	分布式数据库中的权限和保护	238
7.7.2	分布式数据库中权限规则的分布	239
7.7.3	分布式数据库系统中的用户识别和分类	239
7.8	本章小结	240
	习题	241
第八章	分布式数据库与客户机/服务器模式	242
8.1	客户机/服务器计算模式	242

8.1.1	客户机/服务器计算模式概念	242
8.1.2	客户机/服务器环境下应用成分的分布和开放性	244
8.1.3	客户机/服务器模式的体系结构	246
8.2	客户机/服务器模式的定位	249
8.2.1	计算环境演变分析	249
8.2.2	客户机/服务器模式的定位	253
8.3	客户机/服务器模式的特性和优缺点	254
8.3.1	客户机/服务器模式的特性	254
8.3.2	客户机/服务器模式的优点与不足	255
8.4	客户机/服务器模式的远景展望	256
8.5	分布式数据与分布式访问	259
8.5.1	从集中式到分布式数据的转移	259
8.5.2	数据分布的基本形式	260
8.5.3	数据分布的技术	264
8.5.4	分布式数据的访问	266
8.6	创建基于客户机/服务器模式的数据库系统	269
8.6.1	创建客户机/服务器系统的自动化设计工具	269
8.6.2	创建客户机/服务器系统的技术和方法	274
8.6.3	基于组件的软件开发技术和方法	275
8.6.4	异构数据源访问接口技术和方法	278
8.7	本章小结	281
	习题	281
第九章	分布式数据库与 WWW 数据库和移动数据库	283
9.1	WWW 数据库系统的产生及其运行环境	283
9.1.1	WWW 数据库系统的产生	283
9.1.2	统一资源定位器和超文本传输协议	283
9.1.3	超文本标记语言和主页	285
9.2	WWW 数据库系统的体系结构	287
9.2.1	概述	287
9.2.2	WWW 浏览器	288
9.2.3	WWW 服务器	289
9.3	WWW 数据库	290
9.3.1	WWW 数据库的建立和访问	290
9.3.2	WWW 数据库接口	292
9.3.3	WWW 数据库面临的问题	295
9.4	移动数据库系统概述	296
9.4.1	移动数据库的产生背景	296
9.4.2	移动计算环境的体系结构	297

9.4.3 移动数据库系统的参考模型	299
9.5 移动数据库系统的数据管理问题	300
9.5.1 移动应用中的数据	300
9.5.2 移动计算环境下的数据管理问题	301
9.6 移动数据库系统实现技术和应用研究	302
9.6.1 移动数据库数据的同步复制	302
9.6.2 移动数据库数据的移动事务管理	304
9.6.3 移动数据库系统的微型化	305
9.7 间歇同步移动数据库	306
9.8 本章小结	306
习题	308
第十章 分布式数据库系统的发展趋势	309
10.1 并行数据服务器与分布式数据库系统	309
10.1.1 并行数据服务器体系结构	309
10.1.2 并行数据服务器数据定位	312
10.1.3 并行操作处理的算法	315
10.1.4 并行查询处理的算法	321
10.2 分布式知识库系统	322
10.2.1 知识库	322
10.2.2 逻辑查询处理	325
10.2.3 并行递归查询处理	327
10.3 分布式面向对象数据库	332
10.3.1 面向对象数据库	333
10.3.2 分布式对象管理	335
10.4 对象-关系数据库	337
10.4.1 对象-关系数据库系统产生背景	337
10.4.2 对象-关系数据库系统的特点	338
10.4.3 SQL3 标准对对象-关系系统的支持	339
10.4.4 扩展类型系统的实现以及相关问题	341
10.4.5 嵌套关系数据模型	342
10.5 本章小结	344
习题	345
第十一章 数据仓库和数据挖掘	347
11.1 数据仓库	347
11.1.1 数据仓库的定义、特征和结构	347
11.1.2 数据仓库的元数据	351
11.1.3 数据仓库的数据模型	353
11.1.4 数据仓库的设计	357

11.1.5 数据仓库实现中的困难和开放问题	362
11.2 联机分析处理	363
11.2.1 OLAP 的数据组织	363
11.2.2 OLAP 的衡量标准和多维数据分析功能	365
11.2.3 OLAP 的多层结构	368
11.3 知识发现和数据挖掘	370
11.3.1 知识发现和数据挖掘的概念	370
11.3.2 数据挖掘技术概述	371
11.3.3 关联规则	374
11.3.4 其他数据挖掘问题	377
11.3.5 数据挖掘的应用	379
11.3.6 商业化数据挖掘工具	380
11.4 本章小结	381
习题	382
习题参考答案	383
主要参考文献	395

第一章 分布式数据库系统概述

1.1 分布式数据库系统的由来与发展

1.1.1 分布式数据库系统的由来

分布式数据库系统(distributed data base system, DDBS)的研究始于 20 世纪 70 年代中期。这是由于数据库应用需求的拓展和计算机硬件环境的改变,特别是计算机网络与数字通信技术的飞速发展,卫星通信、蜂窝通信、计算机局域网、广域网和激增的 Intranet 及 Internet 得到了广泛应用,使分布式数据库系统应运而生,并成为计算机技术最活跃的研究领域之一。国际上每年都召开专门会议,研究探讨分布式数据库系统的各类问题及其解决方案。英国国家计算中心(national computing centre)专门对分布式数据库做了分析和预测,断言:“分布式系统,特别是以分布式数据库作为该系统的核心,将成为今后 10 年计算机科学发展的主要方向之一。”事实已经证明了这一点。

分布式数据系统符合当今信息系统应用的需求,符合当今企业组织的管理思想和管理方式。尤其是那些地域上分散而管理上又相对集中的大集团、大机关、大企业,如全球或全国性的公司、银行、连锁店、保险业、各类交通运输业,以及全国性人力、财力、资源、环境管理机构和军事国防单位等。在这些组织中,数据通常就已经是分布的了,因为企业中的每个部门都会很自然地维护与自己工作有关的数据。这样企业的整个信息资产就被分裂成通常所说的信息孤岛(island of information),而分布式数据库系统所起的作用就是为把这些小孤岛联系在一起提供桥梁。换句话说,分布式数据库的结构能够反映当今企业组织的信息数据结构:本地数据保存在本地维护,而同时又可以在需要时存取异地的数据。这就是说,往往既要有各部门的局部控制和分散管理,同时也要有整个组织的全局控制和在高层次的协同管理。这种协同管理要求各部门之间的信息既能灵活交流和共享,又能统一管理和使用,自然而然就提出了使用分布式数据库系统的要求。而且,随着应用需求的扩大和要求的提高,人们越来越认识到集中式数据库的局限性,迫切需要把这些子部门的信息通过网络连接起来,组成一个分布式数据库,或重新建立一个既有各部门独立处理又适合全局范围应用的分布式数据库系统。

世界上第一个分布式数据库系统 SDD-1 (system of distributed database)是由美国计算机公司(CCA)于 1976 年至 1978 年设计,并于 1979 年在 DEC-10 和 DEC-20 计算机上实现。

1.1.2 分布式数据库系统的发展

分布式数据库系统是数据库系统与计算机网络系统相结合的产物。计算机网络技术的飞速发展和广泛应用,使得对分布式数据库系统的研究和开发变得更加活跃。

分布式数据库系统产生于 20 世纪 70 年代末期,80 年代进入成长阶段。一方面因为计算机功能增强而成本下降,使得各行各业都购置了计算机,从而有利于数据的分散处理;另一方面也因为计算机网络技术的发展,降低了数据传输的费用。特别是微型机和超级微型机的出现和计算机局域网的广泛应用,则为分布式数据库系统的研制和实现提供了必要的条件。事实上,不论是军事上还是民用上,分布式数据库技术研究都有着深刻的应用背景。因此,各国在 DDBS 上都投入了大量的人力、财力和物力,美国、西欧、日本等相继提出规模宏大的 DDBS 研制计划。例如:

1) 德国斯图加特大学研制的 POREL 系统,历时 11 年,投资 450 万马克。

2) 美国 IBM 的 San Jose 研究室研制的 R * ($R * = R, RR, RRR, \dots$ 表示任意数目的 R) 和 System R。

3) 美国加州大学伯克利分校研制的分布式 INGRES 和荷兰阿姆斯特丹大学研制的扩展 INGRES,在 Unix / PDP 机上实现。

4) 法国 INRIA 研制的 SIRIUS-DELTA 系统和 IMAG 研究中心研制的 MICROBE 系统。

1987 年,关系数据库的最早设计者之一 C. J. Date (另一位是 E. F. Codd) 在《Distributed Database: A Closer Look》中提出了完全的、真正的分布式数据库管理系统应遵循的 12 条规则,这 12 条规则现已被广泛接受,并作为分布式数据库系统的理想目标或标准定义。它们是:

1) 本地自治性(local autonomy);

2) 不依赖于中心站点(no reliance on central site);

3) 可连续操作性(continuous operation);

4) 位置独立性(location transparency and location independence);

5) 数据分片独立性(fragmentation independence);

6) 数据复制独立性(replication independence);

7) 分布式查询处理(distributed query processing);

8) 分布式事务管理(distributed transaction management);

9) 硬件独立性(hardware independence);

10) 操作系统独立性(operating system independence);

11) 网络独立性(network independence);

12) 数据库管理系统独立性(DBMS independence)。

显然,这 12 条规则既不是相互独立的,也不是同等重要的,完全实现难度很大。但是,以这些规则为基础可以有助于理解分布式技术,还可以有助于规划一个特定的分布式系统的功能。这 12 条规则更可以有助于区分一个真正的、普遍意义上的分布式数据库系统与一个只能提供远程数据存取的系统。在一个远程数据存取的系统中,用户可以操作远程站点上的数据,甚至可以同时操作多个远程站点上的数据,但是,“远程与本地不是无缝连接的”,用户会或多或少地知道数据是在异地的,从而采取必要的相应操作。例如,一般意义下的客户机/服务器模式,或是浏览器/ Web 服务器/数据库服务器的三层模式数据库系统,都能够通过计算机网络访问和操作远程站点中的数据库,但是,用户清楚地知道数

据是存储在远程站点上的,需要采取相应的操作,也就是说,用户能感知远程与本地结合的接缝的存在。然而,在一个真正的、普遍意义上的分布式数据库系统中,远程与本地是无缝连接的。对用户来说,几乎完全感觉不到远程与本地结合的接缝的存在,即“一个分布式系统应该看起来完全像一个非分布式系统”。

在本书以后的阐述中,我们将使用“分布式数据库系统”这个术语来特指一个真正的、普遍意义上的分布式数据库系统。

20世纪90年代,分布式数据库系统已进入商品化应用阶段。一些数据库厂商也在不断推出和改进自己的分布式数据库产品,以适应应用的需要和扩大市场占有率。但是,实现和建立分布式数据库系统绝对不是数据库技术与网络技术的简单结合,而是在这两种技术相互渗透和有机融合后的技术升华,它又产生了很多新的技术。而且,分布式数据库系统虽然基于集中式数据库系统,但却有它自己的特色和理论基础。由于数据的分布环境造成了许多固有的技术难度,使得分布式数据库系统的应用被推迟。至今完全遵循分布式数据库系统12条规则,特别是实现完全分布透明性的商用系统仍很难见到。现在,一些商品化的数据库系统产品,如Oracle, Ingres, Sybase, Informix, IBM DB2等,大都提供了对分布式数据库的支持,尽管它们所提供的支持程度不一样。

我国对分布式数据库系统的研究大约从20世纪80年代初期开始,一些科研单位和高等学校先后建立和实现了几个各具特色的分布式数据库系统。如由中国科学院数学研究所设计,由该所与上海科技大学、华东师范大学合作实现的C-POREL,武汉大学研制的WDDBS和WOODDBS,东北大学研制的DMU/FO系统等。他们的工作对我国分布式数据库技术的理论研究和应用开发起到了积极的推动作用。

分布式数据库系统已有30多年的发展历史,经历了一个从产生到发展的过程,取得了长足的进步,许多技术问题被提出并得到了解决。当前,分布式数据库技术已经成熟并得到应用。一方面是实际应用的迫切需要;另一方面是分布式数据库系统的技术相当复杂,某些原理上理论性问题已经研究成熟,但实际应用时,特别是复杂情况下的效率、可用性、安全性、一致性问题并不容易解决。为了解决和减轻实现分布式数据库系统的技术难度,大部分主要的数据库厂商把他们的精力从开发一个“真正的”DDBMS产品,重新定位到开发基于客户机-服务器的系统,或异构型多数据库系统这样的只提供远程数据存取的数据系统。

目前,由于新应用领域的出现,如办公自动化系统(OA)、计算机辅助设计与制造系统(CAD/CAM)、计算机集成制造系统(CIMS)等,以及计算机相关学科与数据库技术的有机结合,如面向对象程序设计技术、多媒体技术、并行处理技术、人工智能技术等,将促进分布式数据库系统向面向对象分布式数据库系统、分布式智能库和知识库系统、数据仓库系统等广阔的领域发展。分布式全球网数据库在不久的将来将会成为现实。

值得我们注意的是以上所给出的系统,无论是原型系统还是商品化系统,它们都是关系数据库系统,至少它们都支持SQL。C. J. Date在《Distributed Database: A Closer Look》中指出:关系技术是(有效的)分布式技术的一个先决条件。事实上,有很多理由说明,如果一个分布式系统要成功,它一定是关系的。

1.2 分布式数据库系统的定义与分类

1.2.1 分布式数据库系统的定义

分布式数据库系统,通俗地说,是物理上分散而逻辑上集中的数据库系统。分布式数据库系统使用计算机网络将地理位置分散而管理和控制又需要不同程度集中的多个逻辑单位(通常是集中式数据库系统)连接起来,共同组成一个统一的数据库系统。因此,分布式数据库系统可以看成是计算机网络与数据库系统的有机结合。

在分布式数据库系统中,被计算机网络连接的每个逻辑单位是能够独立工作的计算机,这些计算机称为站点(site)或场地,也称为结点(node)。所谓地理位置上分散是指各站点分散在不同的地方,大可以到不同国家,小可以仅指同一建筑物中的不同位置。所谓逻辑上集中是指各站点之间不是互不相关的,它们是一个逻辑整体,并由一个统一的数据库管理系统进行管理,这个数据库管理系统称为分布式数据库管理系统(distributed database management system, DDBMS)。

在分布式数据库系统中,一个用户或一个应用如果只访问他注册的那个站点上的数据称为本地(或局部)用户或本地应用;如果访问涉及两个或两个以上的站点中的数据,称为全局用户或全局应用。

由此可见,一个分布式数据库系统应该具有如下的特点:

(1) 物理分布性

分布式数据库系统中的数据不是存储在一个站点上,而是分散存储在由计算机网络连接起来的多个站点上,而且这种分散存储对用户来说是感觉不到的。所以,分布式数据库系统的数据具有物理分布性,这是与集中式数据库系统的最大差别之一。

(2) 逻辑整体性

分布式数据库系统中的数据物理上是分散在各个站点中,但这些分散的数据逻辑上却构成一个整体,它们被分布式数据库系统的所有用户(全局用户)共享,并由一个分布式数据库管理系统统一管理,它使得“分布”对用户来说是透明的。这是分布式数据库的“逻辑整体性”特点,也是与分散式数据库的最大区别。区别一个数据库系统是分散式还是分布式,只需判断该数据库系统是否支持全局应用。

因此,分布式数据库系统中就有了全局数据库(GDB)和局部数据库(LDB)之分,所谓全局是从整个系统角度出发研究问题,全局数据库由全局数据库管理系统(GDBMS)进行管理。所谓局部是从各个站点自己的角度出发研究问题,局部数据库由局部数据库管理系统(LDBMS)进行管理。

(3) 站点自治性

站点自治性也称场地自治性,各站点上的数据由本地的DBMS管理,具有自治处理能力,完成本站点的应用(局部应用),这是分布式数据库系统与多处理机系统的区别。多处理机系统虽然把数据也分散存放于不同的数据库中,但从应用角度来看,这种数据分布与应用程序没有直接联系,所有的应用程序都由前端机处理,只不过对应用程序的执行是

由多个处理机进行,这样的系统仍然是集中式数据库系统。

由以上这三个分布式数据库系统的基本特点还可以导出它的其他特点,包括:

(1) 数据分布透明性

数据独立性是数据库方法追求的主要目标之一。在集中式数据库系统中,数据独立性包括两个方面:数据的逻辑独立性与数据的物理独立性,其含义是用户程序与数据的逻辑结构及数据的物理存储结构无关。

在分布式数据库系统中,数据独立性具有更多的内容。除了数据的逻辑独立性与数据的物理独立性外,还有数据分布独立性,亦称数据分布透明性(data distribution transparency)。所谓数据分布透明性是指用户不必关心数据是如何被逻辑分片的(数据分片透明性),不必关心数据及其片段是否被复制及复制副本的个数(数据复制透明性),也不必关心数据及其片段的物理位置分布的细节(数据位置透明性),同时也不必关心局部场地上数据库支持哪种数据模型。数据分布透明性也可以归入数据物理独立性的范围。

(2) 集中与自治相结合的控制机制

在分布式数据库系统中,数据的共享有两个层次:一是局部共享,即同一站点上的用户可共享本站点局部数据库中的数据,以完成局部应用。二是全局共享,即分布式数据库系统上的用户都可共享在分布式数据库系统的各个站点上存储的数据,以完成全局应用。

因此,分布式数据库系统常常采用集中和自治相结合的控制机制。自治控制机制是指各局部的 DBMS 可以独立地管理局部数据库,具有自治的功能。同时,系统又设有集中控制机制,协调各局部 DBMS 的工作,执行全局控制管理功能。

(3) 存在适当的数据冗余度

在集中式数据库系统中,尽量减少冗余度是系统目标之一。其原因是,冗余数据不仅浪费存储空间,而且容易造成各数据副本之间的不一致性,为了保证数据的一致性,系统要付出一定的维护代价。而在分布式数据库系统中却通过冗余数据提高了系统的可靠性、可用性和改善了系统的性能。在一个集中式系统中,一旦站点发生故障,则整个系统都无法使用。在分布式数据库中,单个站点(甚至是某些站点)的故障,只会使一部分数据可能不能获得,但是用户仍然可以继续操作,访问数据库。这是因为:当某一站点出现故障时,系统可以对另一站点上的相同副本进行操作,因而不会因一处故障而造成整个系统的瘫痪。

但是,数据冗余同样会带来冗余副本之间数据不一致的问题,这是分布式数据库系统必须着力解决的问题。一般来说,增加数据冗余度方便了检索,提高了系统的查询速度、可用性和可靠性,但不利于数据的更新,这将增加系统维护的成本。

(4) 事务管理的分布性

大型数据库分布在多个站点上,数据的分布性必然造成事务执行和管理的分布性,即一个全局事务的执行可分解为在若干个站点上子事务(局部事务)的执行。此外,每个站点执行的事务量比把所有事务提交到一个单一集中的数据库时的事务量要小,而且通过将多个查询放在不同站点上分别执行,或将一个查询分解成一组查询,以并行执行的方式来实现网间查询和网内查询并行,促使了执行性能的改进。同样,由于数据的分布性使得事务的原子性、一致性、可串行性、隔离性和永久性,以及事务的恢复也都应考虑分布性。

完全的、真正的分布式数据库系统应该满足 C. J. Date 提出的 12 条规则,但这 12 条规则不是同等重要的。通常认为,分布式数据库系统中的数据是物理分布在用计算机网络连接起来的各个站点上;每一个站点是一个集中式数据库系统,都有自治处理的能力,完成本站点的局部应用;而每个站点上的数据并不是互不相关的,它们构成一个逻辑整体,统一在分布式数据库管理系统管理下,共同参与并完成全局应用,并且分布式数据库系统中的这种“分布”对用户来说是透明的,也就是说,本地与远程结合的“接缝”是被隐蔽的,用户几乎完全感觉不到远程与本地结合的接缝的存在,即“一个分布式系统应该看起来完全像一个非分布式系统”。这是分布式数据库系统的定义,也是区分一个系统是否为真正意义上的分布式数据库系统的主要依据。那些只提供远程数据存取的数据库系统,即所谓网络数据库系统和多数据库系统,它们虽然也能通过计算机网络访问和操作远程站点上的数据,但是用户清楚地知道数据是存储在远程站点上的,需要采取相应的操作。

图 1.1 是分布式数据库系统的示意图。

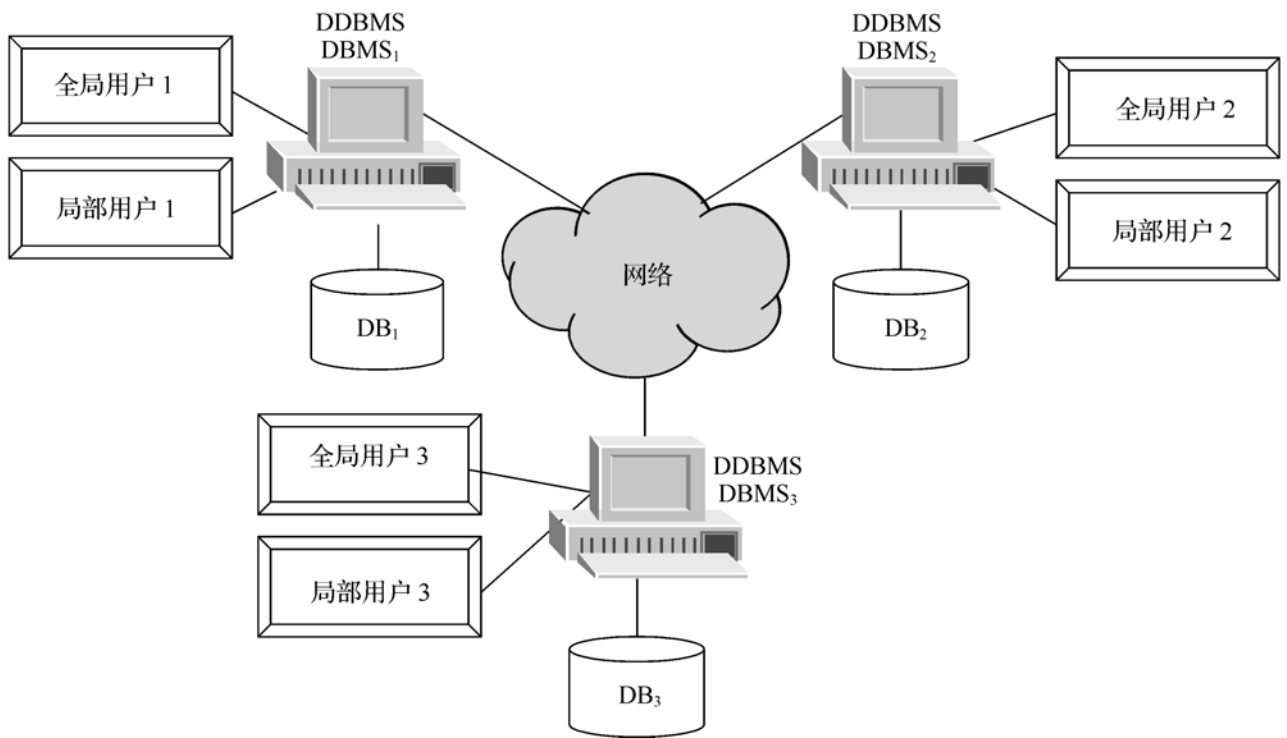


图 1.1 分布式数据库系统的示意图

由此可见,集中式数据库的许多概念和技术,包括数据独立性、数据共享和减少冗余度、查询优化、并发控制、事务管理、完整性、安全性和恢复等在分布式数据库系统中都有了不同的、更加丰富的内容。

1.2.2 分布式数据库系统的分类

分布式数据库系统的分类虽没有统一标准,但较为认同的分类方法有两种。一种是按构成分布式数据库系统的局部数据库管理系统的数据库模型来进行分类,另一种是按分布式数据库系统的全局控制系统类型来进行分类。

1. 按局部数据库管理系统的数据库模型分类

按分布式数据库系统中各站点中的局部数据库管理系统的数据库模型对 DDBS 进行分类是一种常见的分类方法。按照这种分类方法,分布式数据库系统可以分为如下两大类。

(1) 同构型(homogeneous)DDBS

有的翻译成同质型 DDBS,但在我国数据库界流行翻译成同构型 DDBS,对于“同质”另有一层含义。如果各个站点上的数据库的数据模型都是同一数据模型的(例如都是关系型),则称该数据库系统是同构型 DDBS。但是,具有相同类型的数据模型若为不同公司的产品,其性质,例如数据表示的格式,也不尽相同。因此,同构型 DDBS 又可分为同构同质型和同构异质型两种。

1) 同构同质型:如果各个站点上的数据库的数据模型都是同一类型的(例如都是关系型的),而且是同一种 DBMS(通常是同一个厂家的产品),则称该分布式数据库系统是同构同质型 DDBS。

2) 同构异质型:如果各个站点上的数据库的数据模型都是同一类型的,但不是同一种 DBMS(例如有 Sybase、Oracle 等),则称该分布式数据库系统是同构异质型 DDBS。

早期国际上著名的同构型分布式数据库系统主要有:

- ① SDD-1 美国 CCA 公司
- ② SYSTEM R * 美国 IBM 公司
- ③ POREL 德国斯图加特大学
- ④ DDM 美国 CCA 公司
- ⑤ SIRIUS-DELTA 法国

(2) 异构型(heterogeneous)DDBS

如果各站点上数据库的数据模型的类型是各不相同的,则称该分布式数据库系统是异构 DDBS。

早期国际上典型的异构型数据库系统主要有:

- 1) MULTIBASE 美国 CCA 公司 1981 年研制
- 2) IMDAS:H 美国佛罗里达大学 1984 年开发
- 3) DDTS 美国 HONEYWELL 公司 1980 年研制

另外,还有一些准分布式数据库系统,这类系统具有某些分布式系统的特征,但又未能实现或达到分布式数据库综合指标的系统,它们基本上是分布式数据库系统的雏形(有的现在已经发展成为分布式数据库系统),主要有:

- 1) TANDEM 公司的 ENCOMPASS 系统
- 2) IBM 公司的 CICS/ISC 系统
- 3) Oracle 的 SQL * STAR
- 4) 分布式 Ingres 产品为 Ingres/Star(伯克利)
- 5) Applied Data Research(ADR)的 D-NET
- 6) CULLINAE CORP. 的 IDMS DDS

- 7) International Computer Limited(ICL)的 IDMS-DB50
- 8) Siemens AG 的 VDS-D
- 9) Software AG 的 Net-work
- 10) Sybase 的 Replication Server

2. 按分布式数据库系统的全局控制系统类型分类

按分布式数据库控制系统的类型来进行分类,可以分为如下三类:

(1) 全局控制集中型 DDBS

如果 DDBS 中的全局控制机制和全局数据字典(global data dictionary, GDD)位于一个中心站点,由中心站点完成全局事务的协调和局部数据库转换等所有控制功能,则称该 DDBS 为全局控制集中型 DDBS。全局控制集中型 DDBS 的控制方式简单,有助于实现数据更新一致性。但由于全局控制机制和全局数据字典集中存放在一个中心站点,不但容易产生瓶颈问题,而且系统较脆弱,一旦该中心站点失效,整个系统就将崩溃。

(2) 全局控制分散型 DDBS

如果 DDBS 中的全局控制机制和全局数据字典分散在网络的各个站点上,而且每个站点都能完成全局事务的协调和局部数据库转换,每个站点既是全局事务的参与者又是协调者,则称该 DDBS 为全局控制分散型 DDBS。这种系统可用性好,站点独立自治性强,单个站点故障、进入或退出系统,都不会影响整个系统的运行;但是全局控制机制的协调和保持信息的一致性较困难,需要有复杂的设施。

(3) 全局控制可变型 DDBS

也称主从型 DDBS。在这种类型的 DDBS 中,根据应用的需要,将 DDBS 系统中的站点分成两组,其中一组的站点中都包含全局控制机制和全局数据字典(可能是一部分),称为主站点组,它的每一个站点都是主站点;另一组中的站点都不包含全局控制机制和全局数据字典,称为辅站点组,它的每一个站点都是辅站点或从站点(所以也称主从型)。全局控制可变型 DDBS 介于全局控制集中型 DDBS 和全局控制分散型 DDBS 之间,若主站点组的站点数目等于 1 时为集中型;若 DDBS 中全部站点都是主站点时为分散型。

1.3 分布式数据库系统的体系结构和组成成分

1.3.1 分布式数据库系统的体系结构

一个系统的体系结构也称总体结构。它给出该系统的总体构架,定义整个系统的各组成部分及它们的功能,定义系统各组成部分之间的相互关系。在集中式数据库系统中,除了计算机系统本身的硬件和软件(包括操作系统、语言及语言编译程序、其他应用程序)外,主要组成成分有:数据库(DB)、数据库管理系统(DBMS)和数据库管理员(DBA)。分布式数据库系统在此基础上做了扩充:数据库分为局部 DB 和全局 DB;数据库管理系统分为局部 DBMS 和全局 DBMS;数据库管理员也有局部 DBA 和全局 DBA 之分。图 1.2 是分布式数据库系统体系结构的示意图。

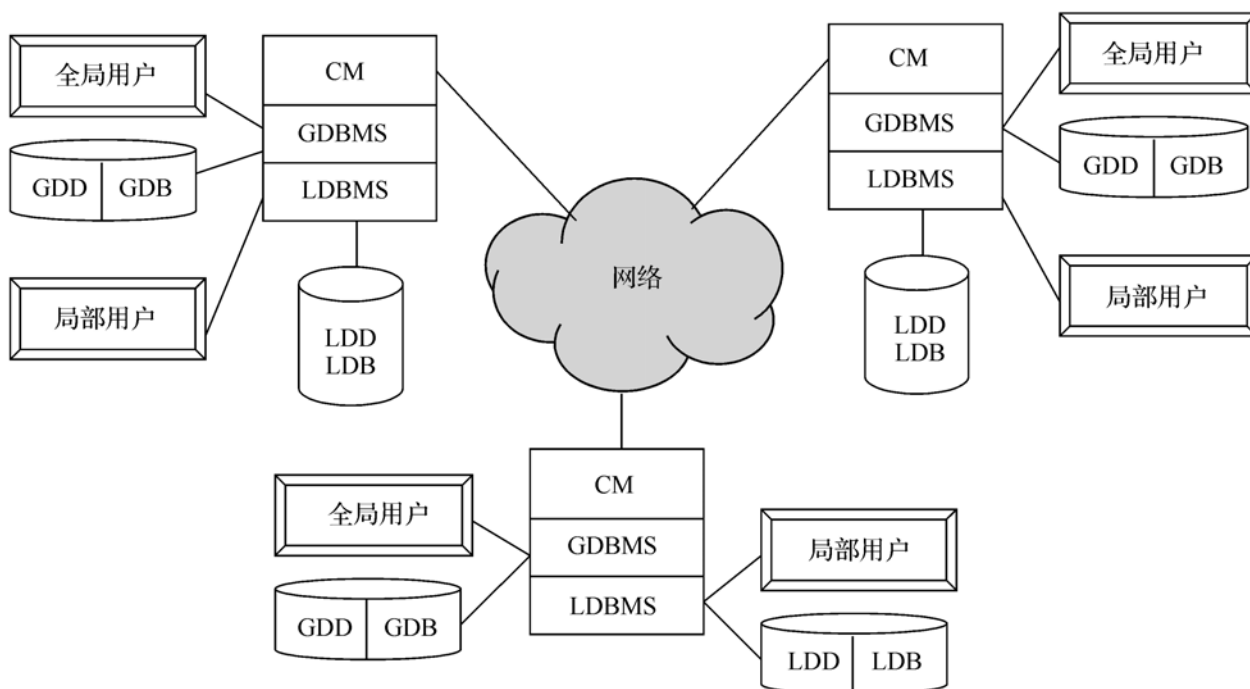


图 1.2 分布式数据库系统的体系结构

1.3.2 分布式数据库的组成成分

分布式数据库(DDB)是分布式数据库系统中各站点上数据库的逻辑集合。与集中式数据库一样,分布式数据库也是由两部分组成:一部分是关于应用所需要的数据的集合,称为应用数据库,它是分布式数据库的主体;另一部分是关于数据库中数据结构的定义,以及全局数据的分片、分布的描述,称为描述数据库,也称数据字典、数据目录或元数据。

在分布式数据库中,数据及描述它的数据目录,也有局部和全局之分,即局部数据和全局数据,局部数据目录和全局数据目录。所谓局部数据是指只提供本站点的局部应用所需要的数据;全局数据虽然也物理地存放在各个站点中,但它参与全局应用,是可被多个站点上应用访问的数据。所谓局部数据目录是指本站点中的局部数据字典(local data dictionary, LDD);而全局数据目录就是全局数据字典,又称网络数据字典,是提供全局数据的描述和管理的相关信息,如数据的结构定义,数据的分片、分布处理、授权、事务恢复等的必要信息。GDD由各站点上的全局数据库管理员(GDBA)建立和管理,而LDD则由局部数据库管理员(LDBA)和全局数据库管理员协调建立和管理。

由此可见,在分布式数据库系统中,分布式数据库的数据目录本身也构成了一个分布式数据库系统。有关分布式数据库目录系统的进一步阐述,将在第七章中给出。

1.3.3 分布式数据库中数据的分片与分布

在分布式数据库系统中,数据的分片和分布是两个重要的概念。事实上,分布式数据库大部分问题都是由数据的分片与分布引起的。它对整个系统的可用性、可靠性及效率有极大的影响,同时也与分布式数据库系统的其他方面的问题密切相关,尤其是分布式查询处理问题,往往同数据的分片与分布问题相互交织在一起。在本节中主要阐述数据分片和

分布的概念,至于如何进行数据的分片与分布,以及数据分布的优化策略等,将在分布式数据库系统设计中阐述。

1. 分布式数据库中数据的分片

数据分片(data fragmentation)也称数据分割,是分布式数据库的特征之一。在一个分布式数据库中,全局数据库是由各个局部数据库逻辑组合而成;反之,各个局部数据库是由全局数据库的某种逻辑分割而得。实际上这也是应用的需要,以关系型 DDBS 为例,一个关系描述了某些数据之间的逻辑相关性,但是,因为不同站点的用户需要该关系中的元组可能不同。例如,某个关系中的元组是与地区有关,上海站点的用户关心、需要的是有关“上海”的那些元组,而北京站点的用户关心、需要的是有关“北京”的那些元组等。这就需要对这个关系进行分割,并将分割后得到的各部分元组,称为该关系的逻辑片段(fragments),存放在相应的站点上。这样处理将各得其所,可以大大减少网络上的通信,从而提高系统的响应效率。

在分布式数据库中,数据存放的单位是数据的逻辑片段。对关系型数据库来说,一个数据的逻辑片段是关系的一部分。数据分片有三种基本方法,它们是通过关系代数的基本运算来实现的。

(1) 水平分片

按特定条件把全局关系的所有元组分划成若干个互不相交的子集,每一子集为全局关系的一个逻辑片段,简称片段。它们通过对全局关系施加选择运算得到,并可通过对这些片段执行合并操作来恢复该全局关系。

(2) 垂直分片

把全局关系的属性集分成若干子集。为得到这些子集,对全局关系作投影运算,要求全局关系的每一属性至少映射到一个垂直片段中,且每一个垂直片段都包含该全局关系的键。这样,可以通过对这些片段执行连接操作来恢复该全局关系。

(3) 混合分片

是以上两种方法的混合。可以先水平分片再垂直分片,或先垂直分片再水平分片,但他们的结果是不相同的。

例 1.1 关系 S (S# , SNAME , AGE , SEX)

1) 定义关系 S 的两个水平分片:

```
DEFINE FRAGMENT SHF1
  AS SELECT * FROM S WHERE SEX = 'M'
DEFINE FRAGMENT SHF2
  AS SELECT * FROM S WHERE SEX = 'F'
```

2) 定义关系 S 的两个垂直分片:

```
DEFINE FRAGMENT SVF1
  AS SELECT S# ,AGE,SEX FROM S
DEFINE FRAGMENT SVF2
  AS SELECT S# ,SNAME FROM S
```

3) 定义关系 S 的两个混合分片:

```
DEFINE FRAGMENT SF1
  AS SELECT S# ,SNAME FROM SHF1
DEFINE FRAGMENT SF2
  AS SELECT * FROM SVF1 WHERE SEX = 'M'
```

在定义各类片段时要遵守如下规则:

1) 完备性条件:必须把全局关系的所有数据映射到各个片段中,绝不允许有属于全局关系的数据却不属于它的任何一个片段。

2) 可重构条件:必须保证能够由同一个全局关系的各个片段来重建该全局关系。对于水平分片可用并操作重构全局关系,对于垂直分片可用连接操作重构全局关系。

3) 不相交条件:要求一个全局关系被分割后所得的各数据片段互不重叠(对水平分片)或只包含主键重叠(对垂直分片)。

2. 分布式数据库中数据的分布

数据分布(data distribution)是分布式数据库的又一特征。所谓数据分布是指分布式数据库中的数据不是存储在一个站点的计算机存储设备上,而是根据需要将数据划分成逻辑片段,按某种策略将这些片段分散地存储在各个站点上。数据分布的策略有:

(1) 集中式

所有数据片段都安排在同一站点上。这种分布策略因系统的数据都存放在同一站点上,对数据的控制和管理都比较容易,数据的一致性和完整性能够得到保证。但由于对数据的检索和修改都必须通过这个站点,使这个站点负担过重,系统对这个站点的依赖性过多。一则容易出现瓶颈,二则一旦这个站点出现故障,将会使整个系统崩溃,系统的可靠性较差。为了提高系统的可靠性,该站点的设施就要提高。

(2) 分割式

所有数据只有一份,它被分割成若干个逻辑片段,每个逻辑片段被指派在某个特定的站点上。这种分布策略可充分利用各个站点上的存储设备,数据的存储量大。在存放数据的各个站点上可自治的检索和修改数据,发挥系统的并发操作能力。同时,由于数据是分布在多个站点上,当部分站点出现故障时,系统仍能运行,提高了系统的可靠性。对于全局查询和修改,所需的时间会比集中式长些,因为数据不在同一场地上,需要进行通信。

(3) 复制式

全局数据有多个副本,每个站点上都有一个完整的数据副本。采用这种策略的系统可靠性高,响应速度快,数据库的恢复也较容易,可从任一场地得到数据副本。但是要保持各个站点上数据的同步修改,将要付出高昂的代价。另外,整个系统的数据冗余很大,系统的数据容量也只是一个站点上数据库的容量。

(4) 混合式

全部数据被分为若干个子集,每个子集安置在不同的站点上,但任一站点都没有保存全部的数据,并且根据数据的重要性决定各个子集的副本的多少。这种分布策略,兼顾了分割式和复制式的做法,也获得了两者的优点,它灵活性好,能提高系统的效率,但同时也

包括了两者的复杂性。

1.3.4 分布式数据库的模式结构

回顾集中式数据库的模式结构,它是三级模式结构:内模式(internal schema)、概念模式(conceptual schema)(又称模式)、外模式(external schema)。这种三级模式的体系结构体现了数据库数据的三个抽象级别。它把数据的具体组织留给数据库管理系统去管理,使用户能抽象地处理数据,而不必关心数据在计算机中的表示和存储方式。这三个级别的模式之间存在很大差别,为实现模式之间的转换,数据库管理系统在这三个模式之间提供了两次映像(mappings):外模式/模式映像和模式/内模式映像。

分布式数据库是基于计算机网络连接的集中式数据库的逻辑集合。因此,分布式数据库模式结构既保留了集中式数据库模式结构的特色,又比集中式数据库模式结构复杂。图 1.3 是分布式数据库模式结构的示意图。

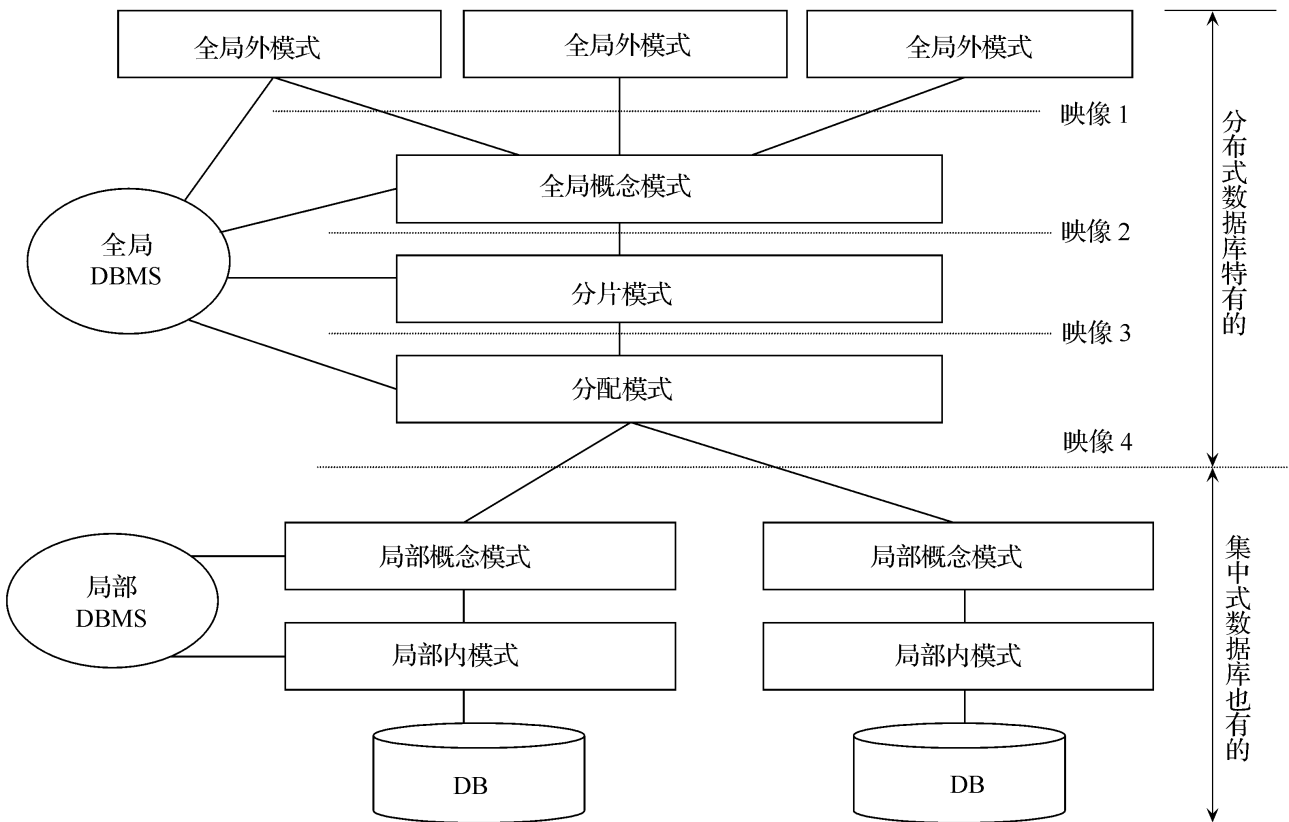


图 1.3 分布式数据库的模式结构

分布式数据库是多层(级)模式结构,层次的划分尚无统一的标准,国内数据库界一般划分为四层。①全局外层(global external level):全局外模式。②全局概念层(global conceptual level):全局概念模式、分片模式、分配模式。③局部概念层(local conceptual level):局部概念模式。④局部内层(local internal level):局部内模式。

类同于集中式数据库,在这四层中的相应模式之间的转换,由全局数据库管理系统和局部数据库管理系统提供的多次映像来实现。

1. 全局外模式

全局外模式(global external schema)是全局应用的用户视图,所以也称全局视图(global view)。分布式数据库的全局视图与集中式数据库视图有同样的概念,不同的只是分布式数据库的全局视图的数据不是从某一个具体站点的局部数据库中抽取,而是从一个由各局部数据库组成的逻辑集合中抽取,即全局外模式是全局概念式的子集。然而,对全局用户而言,都可以认为在整个分布式数据库系统的各个站点上的所有数据库都如同在本站点上一样,只关心他们自己所使用的那部分数据。

2. 全局概念模式

全局概念模式(global conceptual schema)描述分布式数据库中全局数据的逻辑结构和数据特性,与集中式数据库中的概念模式是集中式数据库的概念视图一样,全局概念模式是分布式数据库的全局概念视图。从用户或用户程序角度来看,分布式数据库与集中式数据库应该没有什么区别,即他们可以把数据看作完全没有分布那样来定义分布式数据库中包含的全部数据。因此,可以用集中式数据库中定义概念模式的方法来定义全局概念模式。但是,定义全局模式中所用的数据模型应该便于定义该分布式数据库其他各层的映像,为此,一般采用关系数据模型。采用关系模型的全局概念模式由一组全局关系的定义(如关系名、关系中的属性、每一属性的数据类型和长度等)和完整性定义(关系的主键、外键及完整性其他约束条件等)组成。

3. 分片模式

分片模式(fragmentation schema)描述全局数据的逻辑划分。如前所述,在分布式数据库中,数据存储的单位是数据的逻辑片段。在关系型分布式数据库中,每个全局关系可以通过选择和投影这样的关系操作被逻辑划分为若干个片段,这就是数据分片。分片模式就是描述数据分片或定义片段,以及全局关系与片段之间的映像。这种映像是一对多的,即一个全局关系可对应多个片段,而一个片段只能来自一个全局关系。

4. 分配模式

由数据分片得到的片段仍然是分布式数据库的全局数据,每一片段在物理上可定位(分配)于计算机网络的一个或多个站点中。分配模式(allocation schema)就是根据选定的数据分布策略,定义各片段的物理存放站点,即定义片段映像的类型,确定分布式数据库是冗余的还是非冗余的,以及冗余的程度。如果一个片段分配在多个站点上,则片段的映像是一对多的,分布式数据库是冗余的,否则是不冗余的。冗余的程度描述片段映像到站点的个数,一般根据需要而定,每个片段映像的冗余程度不必相同。

5. 局部概念模式

一般情况下,局部概念模式(local conceptual schema)是全局概念模式的子集。全局概念模式经逻辑划分成一个或多个逻辑片段,每个逻辑片段被分配在一个或多个站点上,

称为该逻辑片段在某站点上的物理映像(physical image)或称物理片段。采用关系模型时,分配在同一站点上的同一个全局关系模式的一个或多个片段(物理片段),构成了该全局关系模式在该站点上的一个物理映像。对每个站点来说,在该站点上全部物理映像的集合称为该站点上的局部概念模式。或者说,一个站点上的局部概念模式是该站点上所有全局关系模式在该站点上物理映像的集合。

由此可见,全局概念模式与站点独立,而局部概念模式与站点相关,即局部概念模式“局部于”某个站点。

例 1.2 图 1.4 是全局关系 R 的分片和分配情况示意图,图中全局关系 R 被划分为 4 个逻辑片段:R₁,R₂,R₃,R₄ 并以冗余方式将这些片段分配到网络上的三个站点中。R₁ 在站点 1 和站点 2 上重复存储,得到 R₁ 的两个物理映像:R₁₁和 R₁₂;R₂ 在站点 1,2 和 3 上重复存储,得到 R₂ 的 3 个物理映像:R₂₁,R₂₂和 R₂₃;R₃ 和 R₄ 只存储在站点 3 上,各得一个物理映像,分别为 R₃₃和 R₄₃。由此得到全局关系 R 在站点 1,2 和 3 上的物理映像,分别为: P₁,P₂ 和 P₃。

同一个逻辑片段在不同站点上的物理映像相同,且其中一个为本,其他为它的副本。例如,R₂ 有 3 个相同的物理映像 R₂₁,R₂₂和 R₂₃,R₂₁,R₂₂是 R₂₃的副本,或 R₂₂,R₂₃是 R₂₁的副本,或 R₂₃,R₂₁是 R₂₂的副本。

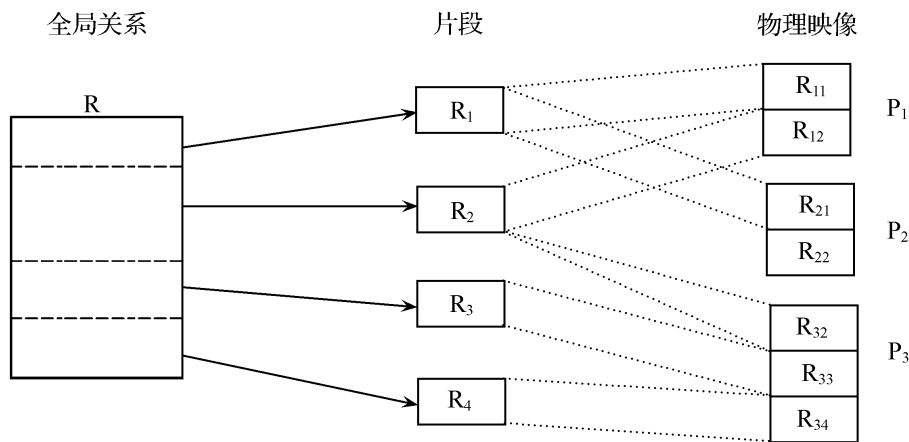


图 1.4 全局关系 R 的逻辑片段与物理映像

6. 局部内模式

局部内模式(local internal schema)是分布式数据库中关于物理数据库的描述,类同集中式数据库中的内模式,但其描述的内容不仅包含只局部于本站点的数据的存储描述,还包括全局数据在本站点的存储描述。

在图 1.3 的分布式数据库模式结构图中,全局概念模式、分片模式、分配模式与各站点特征无关,它们不依赖于各站点上的局部 DBMS 的数据模型。当全局数据库的数据模型与局部数据库的数据模型不同时,则物理映像与各局部数据库的数据模型之间还必须进行数据模型转换。即使数据模型相同,可能因产品的厂家不同,它们的数据类型和格式也不同,同样需要进行相应转换。这就是说,需要把物理映像转换为本地 DBMS 支持的数据模型和可操作的对象,这种转换(映射)称为本地化映射,由局部映射模式完成。具体的

映射关系,由各局部 DBMS 的类型决定。在异构分布式数据系统中,由于各站点上数据库的数据模型不同,各站点可拥有不同类型的局部映射模式。

这种分层的体系结构为理解分布式数据库提供了一种极通用的概念结构,它有三个显著特征:

1) 数据分片和数据分布概念的分离,形成了“数据分布独立性”的概念。

2) 数据冗余的显式控制:数据在各个站点上的分布情况在分布模式中一目了然,便于系统管理。

3) 局部 DBMS 的独立性:这个特征允许在不考虑局部 DBMS 的数据模型的情况下来研究分布式数据库管理的有关问题。

通常分布式数据库还支持局部用户和局部应用,而且局部用户定义的数据局部于本站点使用,不参与全局应用。因此,通常局部概念模式实际上还应该包含这些局部数据的局部概念模式和它的子集局部外模式,它们不是来自全局概念模式且由局部 DBA 描述和管理。

这时值得注意的是,全局数据虽然物理的存储在各站点上,但它们由全局 DBA 管理,而局部数据则由局部 DBA 管理。各站点上的用户是否有权访问全局数据,由全局 DBA 授权,局部 DBA 无权向局部用户授予全局数据的访问权,即使这些全局数据存放在本站点中。

1.3.5 分布式数据库管理系统的功能结构

1. 分布式数据库管理系统的功能

分布式数据库管理系统是分布式数据库系统的核心,负责实现 DDB 的建立、查询、更新、复制、维护等功能。包括提供分布透明性,查询优化,协调全局事务的执行,协调各局部 DBMS 共同完成全局应用,保证数据库的全局一致性,执行并发控制,实现更新同步和全局恢复等。

分布式数据库系统是在集中式数据库系统技术的基础上发展起来的,但不是简单地把集中式数据库分散地实现,它是具有自己的性质和特征的系统。为了实现前面提到的分布式数据库的特点,DDBMS 软件除了提供集中式 DBMS 所提供的功能外,还必须能够提供以下的集中式 DBMS 所不能提供的附加功能。

1) 数据跟踪:具有能够通过扩展 DDBMS 日志来记录数据分布、分片和复制的能力。

2) 分布式查询处理:具有能够通过通信网络存取远程站点的数据,以及在不同站点间传输请求和数据的能力。

3) 分布式事务管理:具有能够为需要从多个站点存取数据的查询和事务设计执行策略的能力,以及将分布数据的存取和整个数据库完备性的维持保持同步的能力。

4) 复制数据的管理:具有能够把数据库从单个站点故障和新类型故障(例如通信链路的故障)中恢复的能力。

5) 安全性:分布式事务的执行必须具有适当的数据安全管理,以及用户授权/存取权限的安全管理。