

学校编码: 10384
学 号: 0028002

分类号 _____ 密级 _____
UDC _____

学 位 论 文

基于 ORACLE 的分布式数据库的并发控制与数据分配的研究

王劲波

指导教师: 薛永生 教授
申请学位类别: 硕 士
专业名称: 计算机应用
论文提交日期: 2003年5月 日
论文答辩日期: 2003年 月 日
学位授予单位: 厦 门 大 学
学位授予日期: 2003年 月 日

答辩委员会主席: _____
评 阅 人: _____

二〇〇三年五月

厦门大学博硕士学位论文摘要库

摘 要

分布式数据库是数据库领域的一个重要分支，尤其是在当今网络技术飞速发展的时代，分布式数据库的应用是越来越广泛。国家政府部门、银行系统、证券交易系统、电力管理系统、交通运输部门等各种机构无不需分布式数据库的管理和支持。当今对分布式数据库的研究主要集中在并发控制、数据分配、事务管理、数据恢复等方面。本文主要以 ORACLE 的分布式数据库管理为基础来研究分布式数据库的并发控制和数据分配。

在并发控制方面，本文通过分析几种常用的并发控制模型以及它们在分布式环境下的拓展，提出了一种基于加锁粒度的分布式高优先级两段锁模型。在该模型中，引入了优先级分配机制和加锁检测机制。优先级分配机制主要是处理好事务优先级与执行时间及截止期的关系，同时分配好全局事务中子事务的优先级。而通过加锁检测机制可以确定哪些冲突事务可以通过降低加锁粒度来解决冲突。同时提出了该模型的算法。该算法在避免死锁、降低事务高重启率、提高事务的并发度等方面有较大的改善。

在数据分配方面，本文主要通过实例讨论数据分配的原则，分别对非冗余分配、冗余分配、静态启发式分配等数据分配方法采用量化的方法来进行研究，并对动态分配也做了一些比较分析，在研究过程中，把分布式检索和更新处理问题结合起来考虑，以获得减少通信代价、减少检索代价及更新代价的最优解决方案，从而明确得出全局关系划分后的逻辑片段应该置放的最佳场地。

关键词： 分布式数据库、ORACLE、事务管理、并发控制、数据分配、DHP-2PL、优先级分配机制、加锁粒度、逻辑片段、通信代价

Abstract

Distributed database is an important branch of database field. Especially for the era of network technology developing quickly, the application area of distributed database becomes wider and wider. The distributed database system can be used in the government institution, bank system, stock exchange system, electric power system, and transport system. And these systems all need the support and management of distributed system. Now the research on distributed system pays more attention on concurrency control, data allocation, transaction management, data recovery and so on. In this paper, the concurrency control and data allocation are researched which based on the distributed database management system in ORACLE.

About concurrency control, some lock models in concurrency control is introduced, and how to using them on the distributed environment is discussed. At the same time a locking granularity-based distributed high priority two phase locking model is presented. At this model, a priority allocation scheme and a locking checking scheme are introduced. The problem between consistence and deadlines can be deal with harmoniously by the priority allocation scheme and the priority of sub-transactions also is allocated by it. And some conflicts can be disappeared by decreasing locking granularity with locking check scheme. At last the algorithm of this model is presented. This algorithm has a lot of improvement on the way of avoiding deadlock, increasing transaction concurrency and decreasing the restart ratio of transaction.

About data allocation, the principles of data allocating are discussed with instances in this paper. We also research some data allocation way such as non-redundance allocating, redundance allocating, static and heuristic allocating by quantitative method. Also, there are some comparative analyses of dynamic allocating. During research process, considering distributed search and updating problem, we achieves optimum settlement by reducing cost of communication, search and updating. Thus, we define the best place of logic segments after divided from global relation.

Keywords: distributed database, ORACLE, transaction management, concurrency control, data allocation, DHP-2PL, priority allocation scheme, locking granularity, logic segments, communication cost

目 录

| | |
|--------------------------------------|----|
| 第一章 分布式数据库综述..... | 1 |
| 1.1 概述..... | 1 |
| 1.2 分布式数据库系统的组成与特点..... | 1 |
| 1.3 分布式数据库管理系统的研究概况及趋势..... | 3 |
| 1.4 当前分布式数据库系统的需要改善的问题..... | 6 |
| 1.5 本文的主要工作..... | 7 |
| 第二章 ORACLE 的分布式数据库管理..... | 8 |
| 2.1 ORACLE 分布式数据库的体系结构..... | 8 |
| 2.2 ORACLE 分布式数据库中的数据库链接..... | 9 |
| 2.3 分布式数据库系统的事务管理..... | 10 |
| 2.3.1 分布式事务的特性..... | 10 |
| 2.3.2 分布式事务管理目标..... | 11 |
| 2.3.2 分布式事务的恢复机制..... | 12 |
| 2.4 ORACLE 中的分布式事务管理..... | 12 |
| 2.4.1 设置分布式事务的初始化参数..... | 13 |
| 2.4.2 查看分布式事务处理的信息..... | 13 |
| 2.4.3 处理悬而未决的事务..... | 15 |
| 2.5 ORACLE 中的并发控制..... | 16 |
| 第三章 分布式数据库的并发控制研究..... | 18 |
| 3.1 可串行化调度..... | 18 |
| 3.1.1 串行调度与可串行调度..... | 18 |
| 3.1.2 可串行化调度的判定..... | 19 |
| 3.2 分布式并发控制简介..... | 20 |
| 3.2.1 锁模型..... | 20 |
| 3.2.2 时间印模型..... | 21 |
| 3.2.3 乐观的并发控制方法..... | 21 |
| 3.3 两段锁协议(2PL)及其它几种锁模型..... | 22 |
| 3.3.1 两段锁协议(2PL)及分布式两段锁协议(D2PL)..... | 22 |

| | | |
|---------|---------------------------|----|
| 3.3.2 | 利它锁方法(ALTruistic locking) | 23 |
| 3.3.3 | 积极加锁法与消极加锁法 | 24 |
| 3.3.4 | 高优先级两段锁协议(HP-2PL 协议) | 24 |
| 3.3.5 | 混合两段锁协议(Hb2PL 协议) | 25 |
| 3.3.6 | 实时静态两段锁协议(RT-S2PL) | 25 |
| 3.4 | 基于加锁粒度的分布式高优先级两段锁协议的研究 | 26 |
| 3.4.1 | 分布式高优先级两段锁协议(DHP-2PL) | 26 |
| 3.4.2 | 基于加锁粒度的分布式高优先级两段锁协议 | 27 |
| 3.4.2.1 | LGDHP-2PL 的优先级分配策略 | 28 |
| 3.4.2.2 | LGDHP-2PL 的冲突算法 | 31 |
| 3.4.3 | LGDHP-2PL 算法的分析 | 33 |
| 第四章 | 分布式数据库的数据分配操作 | 34 |
| 4.1 | 数据分配定义 | 34 |
| 4.2 | 数据分配方法 | 35 |
| 4.2.1 | 参数约定 | 35 |
| 4.2.2 | 非冗余分配最佳法 | 36 |
| 4.2.3 | 选择所有收益场地法 | 36 |
| 4.2.4 | 添加副本法 | 37 |
| 4.3 | 基于实例的数据分配方法分析与研究 | 37 |
| 4.3.1 | 运用“非冗余分配最佳法”进行数据分配 | 37 |
| 4.3.1.1 | 第一种途径：根据本地访问次数为标准 | 38 |
| 4.3.1.2 | 第二种途径：把片段分配到各个场地 | 38 |
| 4.3.1.3 | 对非冗余分配最佳法的分析 | 39 |
| 4.3.2 | 选择所有收益场地法 | 39 |
| 4.3.3 | 用添加副本法进行冗余分配及遗传算法比较 | 40 |
| 4.4 | 各种数据分配方法的总结 | 41 |
| 第五章 | 结束语 | 43 |
| | 参考文献 | 44 |
| | 致 谢 | 46 |

第一章 分布式数据库综述

1.1 概述

在当今的信息社会里，人们的全部社会活动都离不开信息，信息总是以数据为载荷符号来反映和记录的。从 50 年代引入的文件管理系统，就是一种专门从事数据管理的系统。但文件系统缺乏数据共享性，难以实现对数据的统一管理，故在 60 年代后期，一种新的数据管理方法——数据库技术出现了。由于数据库对数据的管理具有持久性、有效性、共享性和冗余的可控性等特点，以及在消除数据的不一致性，提高存储和查询效率等方面具有的优势，数据库系统很快就占领了整个数据管理市场。但是，在数据库应用的开始阶段，并没有重视地域分散的特定需求，在强调集中时又产生了不协调，同时随着计算机技术的发展，尤其是网络技术的发展，集中式数据库系统对于处理地域分散、数据共享的应用显得力所不及。这样，分布式数据库系统就被提出来。人们设想以“数据库+计算机网络”来实现分布式数据库系统，这样既达到对数据的集中管理与共享，又能使地域的分散性被系统隐蔽起来。分布式数据库适合于一些地域上分散的大团体、大机构的组织结构与管理。因为在这些单位中，往往既要有各子单位的分散管理，也要在高一级层次上实现协同管理，既要有局部控制，同时也需要高层次的协作。故从 70 年代末开始，分布式数据库进入一个飞速发展的时期。

1.2 分布式数据库系统的组成与特点

分布式数据库系统(DDBS)的定义是：由分布式数据库(DDB)和分布式数据库管理系统(DDBMS)所构成，分布式数据库是一组数据集，逻辑上它们属于同一系统，而物理上它们分散在用计算机网络连接的多个场地上，并统一由一个

分布式数据库管理系统管理。分布式数据库管理系统分布式数据库管理系统包括三个主要成分：全局数据库管理系统(GDBMS)、局部数据库管理系统(LDBMS)、通讯管理程序(CM)。GDBMS 负责管理分布式数据库中的全局数据。LDBMS 是 DDBS 中和场地的数据库管理系统，而 CM 是保证 DDBS 中场地间信息传送的部分。整个 DDBS 的系统结构如图 1-1：

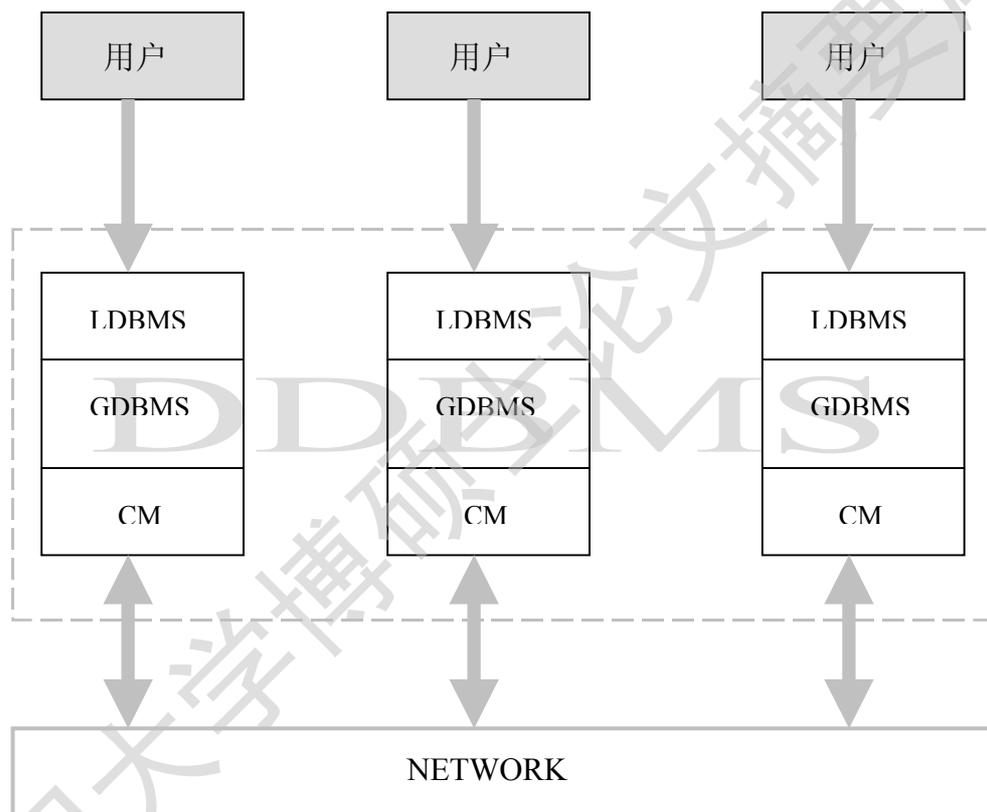


图 1-1：分布式数据库系统体系结构

因为分布式数据库系统是集中式数据库和计算机网络结合，是分散与集中的统一，故兼有二者的共同特性。其主要特性如下：

- I. 可靠性：分布式数据库可通过不同场地存放的冗余数据来恢复某个场地的失效的数据。
- II. 数据冗余的可控性：对于传输代价高于在存储代价时，可将数据冗余存储；当存储代价高于传输代价时，可控制冗余。

III. 事务管理的分布性：在分布式数据库中，一个事务(全局事务)的执行将划分成在许多场地上执行的子事务(局部事务)，子事务的执行结果合并而成全局事务的结果。

IV. 查询的全局优化与局部优化：全局优化主要决定在多副本中选取适当的场地副本，使得场地间的数据传输量以及次数最少；而局部优化跟集中式数据库的功能一样。

V. 系统透明性与分布透明性：用户在对分布式数据库操作不必知道数据的存储位置，用户可像使用集中式数据库一样来使用分布式数据库。

VI. 并行性：由于数据的冗余存储，分布式数据库可以通过并发控制来实现事务的并行处理。

1.3 分布式数据库管理系统的研究概况及趋势

对分布式数据库管理系统很早就展开了研究，但因为一些多方面的原因，如超复杂性，完全透明性，及其花费问题，故分布式数据库并不像集中式数据库那样很快就达到商品化、实用化。但单纯就技术问题而言，大部分的问题已经得到解决，分布式数据库的理论已较为完善。但实用化是需要大量投入和广泛测试才能完成，而在这方面是需要大量的资金和时间，同时具有大量资金的集团并没到不使用分布式数据库系统就不能生存的地步，故正是这一原因使得分布式数据库的应用被推迟。下面将列举一些在实例国内外分布式数据库发展史有名的系统。

早在 70 年代中期，美国计算机公司研制出第一个分布式数据库原型系统 SDD-1，该系统是在 1976~1978 年进行设计，1979 年在 DEC-10 和 DEC-20 计算机上实现的。它以 ARPANET 网络连接，以现有的 Datacomputer 数据库管理系统为基础。SDD-1 支持数据分片，包括水平分片和垂直分片；并支持数据分片的重复存储，即支持可控制冗余；支持分片和位置透明性。在查询处理优化方式方面采用了半联接技术，并且采取全局优化方法处理查询。在事务管理中采用时间印方法和由 RELNET 提供虚拟机的假脱机方式保证可靠性。

SDD-1 作为分布式数据库系统中第一个开发的原型，整体上来说它是成功

的，正是因为它概括了分布式数据库的几乎全部理论和实现技术，至今在有关分布式数据库的讨论中人们仍把它看作是一个典范。当然，由于当时条件和实现技术的限制，它也有着一些不足之处，但这些都是分布式数据库中次要的或不是十分重要的。总之，它给以后研究带来了许多有益的经验。

R*系统是 IBM 公司 San Jose 研究所研制的一个分布式数据库系统原型。它的目标是讨论在分布式数据库中场地自治性的问题。它是以关系数据库 System R 为基础的分布式版本。R*系统扩展了 SQL 语言定义和操作分布式数据库，它支持位置透明，但它不支持数据分片和数据重复存储；R*系统的优化策略和 System R 一样是代价计算，分布执行计划是由查询发生场地产生的；在事务管理中采用了成熟的两段锁和两段提交方式，并且设置了死锁检测机制。R*系统还试图使用快照和视图技术，以使分布式数据库系统 R*系统像 System R 一样在查询效率方面和用户使用上都有较好的功能。

在处理同构分布式数据库系统的研究中，比较成功的有：加州伯克利分校开发的分布式 D-INGRES(Distributed - INGRES)、德国斯图加特大学研制的 POREL、法国 INRIA 研制的 SIRIVS-DELTA 等。

D-INGRES 是关系数据库系统 INGRES 的分布式版本。是用于局域网，为用户提供全局模式；支持水平分片但不支持垂直分片。支持片段的重复存储；在查询优化处理中，支持更强的并行处理特性，使得在局域网中的传输代价几乎和 I/O 代价相当。D-INGRES 在事务管理中，亦使用两段锁机制，并设死锁检测装置以解决死锁问题，提交则相似于两段提交方式。

POREL 是为实行关系数据库机计划而研制的分布式数据库系统，目的之一是研究各种计算机的可移植性。

SIRIVS-DELTA 是从 1979 年开始，目标是研究同构异质中的问题。

在处理异构分布式数据库系统研究中，比较有代表的有美国计算机公司研制的软件系统 MULTIBASE 和 Honeywell 公司计算机科学中心研制的 DDTs (Distributed Database Tested System)。

MULTIBASE 是对异构分布式数据库中实现数据检索，并提供了只读的应用高级接口，所以在对数据位置和各场地数据库管理系统都是透明的。

DDTs 在设计中，充分的注意到模块化和灵活性，它由若干子各级系统组

成，为了支持某个实验可以很容易替换相应的功能模块，也可以把不同的数据库管理系统组合在一起。这对现在已经存在的集中式数据库系统组成统一的分布式数据库系统很有现实意义。

我国对分布式数据库系统的研究是 80 年代初开始的，虽然起步较晚，但起点相对都比较高，研制出来的分布式数据库系统都具有相当高的水平，在功能和性能方面都达到了当今世界先进水平。以下是几个国内具有代表性的系统原型。

WDDBS-32 是武汉大学数据库级研制的，其是以 UNIX 操作系统支持，用 C 语言开发的通用、完全透明、完全分布的分布式数据库系统。它允许场地动态增加或减少；支持水平、垂直分片，诱导水平分片和混合分片；支持重复副本，支持分片、重复存储和位置的透明性，采用目录辞典全复制方式管理，具有视图机制，授权机制；支持主键、异键，以及可用分片语义及其它语义规则实现完整性控制；事务管理采用多种锁机制的分布两段锁协议，用前后台控制的无阻塞可终止的改进两段提交协议。WDDBS-32 是我国自行设计和实现的面向实用的原型系统。

东南大学开发的 SUNDDDB 系统是在 SUNIII 工作站上开发的，追求实用性，在设计时支持重复副本和水平分片，提供快照技术。事务管理采用 2PL 协议及 2PC 协议。

中国科学院数学所、上海科技大学和华东师大联合研制的 C-POREL。是在 POREL 基础上经过改进设计后实施的。

中国人民大学数据与知识工程研究所研制的 DOS/SELS。是一个连接远程网上 IBM43 系列计算机上的 DL/I 数据库的分布数据查询系统。

在异构的分布式数据库系统比较有代表的是以东北大学为主研制的 DMU/FO，是支持网络模型的分布式数据库系统。南京大学研制的 LSZ，它由异种网络、微型机、操作系统构成的分布式系统上运行的异构型分布式数据库，提供全局数据语言，全局查询优化，有很好的并行性能。

经过二十几年的发展，分布式数据库系统理论已经较为成熟，其技术问题也基本解决。现在很多数据库商用产品已经具有分布式数据库的特色，如：SQL Server, Oracle, DB2 等。同时分布式数据库系统具有潜在的大市场。如银行

系统、电力管理系统、各种集团公司的管理等。故对分布式数据库系统的研究至今仍未中断过。

1.4 当前分布式数据库系统的需要改善的问题

以下就是目前分布式数据库产品在技术上的一些共同问题：

I. 分布和复制数据的透明管理

分布式数据库技术将数据分布和复制到一个网络上的其他多个计算机上，实现网络分布透明、复制透明和功分透明，访问数据库在理论上就是和集中式数据库管理系统的接口毫无区别。但是，当前大多数分布式数据库产品并没有达到足够的透明，主要表现在许多系统不允许跨多个数据库复制数据，而且受物理场地要被指定的限制，第二个问题就是缺乏合适的操作系统对透明的支持，从而不能完全透明，使得分布式数据库的管理很困难。在使用透明方式访问地理上分布的数据库的应用程序具有可管理性、模块性和信息性差的三大缺点。

II. 分布式事务的可靠性

充分利用分布式事务的冗余特性和完整性，提高分布式事物的处理数据的可靠性，避免由于通讯或其他问题导致的操作失败而造成全系统的瘫痪或数据出错的问题。

III. 分布式并发控制

分布式并发控制要解决两方面的问题：一是当多个用户同时读写相同数据时系统必须作出协调，保证结果的正确性和数据库的完整性，并尽可能地提高并行性；另一方面，在系统中包含冗余数据的情况下，在处理用户的更新操作时，系统也必须以最小的开销保持各冗余副本的一致性。当前分布式数据库产品多数采用锁机制或时间印方法来实现并发控制，但这样会引入死锁检测机制或事务重启机制，加重系统的负担。因此，总的来说，衡量分布式并发控制算法的标准应为正确性与性能的统一。

IV. 性能的提高

分布式数据库分割了概念数据库，使得数据区域化，并减少 CPU 和 I/O

的开销以及远程访问延时。在查询能力上，通过分布式数据库的并发控制和提交协议来实现查询间并行和查询内并行。

V. 异构数据库的转化

实现分布式数据库各子数据库之间数据的转化及统一。

1.5 本文的主要工作

本文主要以 ORACLE9i 为基础来研究分布式数据库管理系统。并在 ORACLE9i 提供的平台上对分布式数据库仍需解决的一些问题做出探讨。

在课题的研究过程中，主要针对以下几个方面进行探讨：

1. 对 ORACLE9i 的分布式数据库管理系统的研究；
2. 对分布式数据库并发控制的研究；
3. 对分布式数据库数据分配的研究。

第二章 ORACLE 的分布式数据库管理

基于 ORACLE 的分布式数据库是物理上存放在网络中的多个 ORACLE 数据库，在逻辑上可把它们看做单个的数据库。用户可以通过网络对异地数据库中的数据同时进行存取，而服务器之间的协同处理对于工作站用户及应用程序而言是完全透明的；使用者无需关心网络的连接细节，无需关心数据在网络接点中的具体分布情况，也无需关心服务器之间的协调工作过程。

2.1 ORACLE 分布式数据库的体系结构

ORACLE 分布式数据库系统是一个客户机/服务器的网络体系结构，在网络环境中，每一个具有多用户处理能力的硬件平台都可以成为服务器，也可以成为工作站。运行在不同服务器上的数据库彼此共享数据。每个服务器也许是在物理上分离，但是在逻辑上却是一个统一的数据库系统，数据一致性、完整性及安全性都是对这一逻辑上统一的数据库进行控制的。图 2-1 就是一个典型的 ORACLE 分布式数据库系统。

ORACLE 分布式数据库系统允许用户透明地操作远程数据库的数据，可以用多个数据库满足多个工作组、部门或地区的需求，并把它们连接在一起，使应用程序看起来只有一个大型数据库。用户对此单一逻辑的数据库中数据的任一请求，都被自动分解、自动寻址、自动转换为网络请求，并在相应结点上实现相应的操作。

ORACLE 支持同构分布式数据库系统和异构分布式数据库系统。对于同构的分布式数据库，每个场地上的数据库都是 ORACLE 数据库系统，ORACLE 中的 SQL*NET 允许多台运行 ORACLE 产品的计算机相互通讯。在 ORACLE 的异构分布式数据库系统中，对于应用程序来说，整个系统就像一个单独的、本地的 ORACLE 数据库，本地 ORACLE 数据库服务器隐藏了数据的分布和异构的情况。

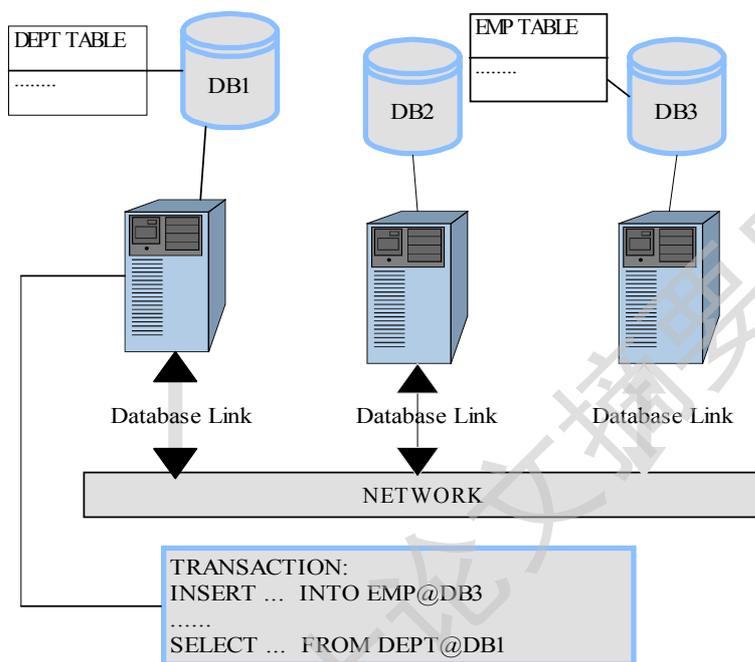


图 2-1 一个典型的 ORACLE 分布式数据库系统

2.2 ORACLE 分布式数据库中的数据库链接

数据库链接是一个定义从一个 ORACLE 数据库服务器到另一个 ORACLE 数据库服务器的单向的通信路径的指示器。链接指示器实际上是一个数据字典入口。

数据库链接允许本地用户通过每个数据库在网域中唯一的全局名来访问远程数据库。在连接远程数据库时可以有三种数据库链接方式：

Private 数据库链接：只有私有数据库链接的拥有者或该模式中的 PL/SQL 子程序，可以使用该链接来访问相应的远程数据库中的数据库对象。

Public 数据库链接：数据库中所有的用户和 PL/SQL 子程序，都能使用该链接来访问相应的远程数据库中的数据库对象。

Global 数据库链接：任何数据库中的用户和 PL/SQL 子程序，都能使用该链接来访问相应的远程数据库中的数据库对象。

当创建数据库链接时，要确定有哪个用户应该连接到远程数据库以访问数

据。该类用户也有三种。

已连接用户：这种方式是指必须在远程数据库上有一个账号，而且这个账号与在本地数据库上的用户名相同。

固定用户：用户使用链接中已引用的用户名和密码。

当前用户：用户作为一个全局用户登陆。

创建数据库链接时使用“CREATE DATABASE LINK”语法。

当本地数据库通过数据库链接与远程数据库进行连接时，每一方都可以有独占或者共享的服务模式(如表 2-1)。

表 2-1 数据库链接可能的服务模式

| 本地数据库模式 | 远程数据库模式 |
|---------|---------|
| 独占服务 | 独占服务 |
| 独占服务 | 共享服务 |
| 共享服务 | 独占服务 |
| 共享服务 | 共享服务 |

采用数据库链接的最大优点就是，它允许用户访问远程数据中的其他用户的对象，所以这些用户就具有了这些对象的拥有者的权限。即一个本地用户可以访问一个到远程数据库的链接，而不必是该远程数据库中的一个用户。同时数据库链接允许将有限的访问远程数据库的权限授予给本地用户。通过使用当前的用户链接，就能创建集中管理的全局用户。

2.3 分布式数据库系统的事务管理

2.3.1 分布式事务的特性

分布式事务包括一个或多个的语句，这个语句单独或成群出现，在一个分布式数据库两个或更多的节点上访问、更新、删除数据。分布式事务与集中式

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士学位论文摘要库