

Distributed
Databases

分布式数据库
—— **原理与系统**

Principles and Systems

分布式数据库

—原理与系统

[美] 斯蒂芬诺·塞里 吉乌塞普·皮拉盖蒂 著

宝鸡师范学院计算中心

一九八六年十一月

编 译 说 明

随着分布式计算机系统的迅速发展,数据库技术开始从传统的集中式转向分布式,分布式数据库已成为分布式计算机系统开发和应用的关键技术之一,同时也是我国“七五”期间的重点研究开发项目之一。但是目前国内尚无一本详细介绍分布式数据库原理与系统的著作。我院计算中心结合院内科研课题,组织编译了这本在分布式数据库方面具有权威性的专著,以满足院内外有关方面的急需。

本书以教材格式编排,适合于计算机科学和工程领域中的教师、研究生、高年级学生、科研工作者、系统管理人员、系统分析和程序设计人员阅读。

在本书的编译出版过程中,我们始终得到我院各级领导的热情鼓励和支持。杨异军副院长在百忙中审阅了本书,并提出了一些有价值的意见。施凤飞、郑存库、油朝坤、刘武等同志,也给予了大力支持和帮助,在此一并表示感谢。

由于时间仓促,加之编译者水平所限,因此错误之处在所难免,恳望读者批评指正。

编译者

DISTRIBUTED DATABASES

Principles and Systems

Stefano Ceri

Giuseppe Pelagatti

1984 by McGraw-Hill Book Inc.

原作者序

众所周知，二十世纪七十年代人们曾广泛地利用计算机来建立强有力的集中式数据库系统。数据库系统技术已经建立了它自己的理论基础，并且在大量的应用中经过了实践检验。与此同时，计算机网络也一直在迅猛地发展。计算机网络能把不同的计算机互连起来，允许在互连的计算机之间交换数据和其它资源。

近几年来，数据库和计算机网络的深入发展和广泛应用导致了一个崭新的领域：分布式数据库。简单地说，分布式数据库是建立在计算机网络之上的一种数据库，而不是建立在单个计算机上的数据库。构成分布式数据库的数据存贮在计算机网络的不同地点，由网络中计算机所运行的应用程序可访问不同地点的数据。

建立和实现分布式数据库面临着许多完全崭新的问题。为了解决这些问题已经进行了大量的研究工作。这种研究工作构成了一门新的学科，这门新的学科有它自己的基本原理。为了搞清楚分布式数据库，仅了解传统的数据库和计算机网络的基本原理是远远不够的，我们还必须把这些基本原理与分布式数据库这门新学科的特殊问题的研究综合起来。

本书论论分布式数据库这门新学科的基本原理。本书特别适合那些在专业上对分布式数据处理的最新进展感兴趣的读者，例如，计算机科学领域中的教师、研究生和高年级学生，科研工作者，系统管理人员，系统或应用设计人员，系统分析人员和程序设计人员。既可将本书用作有关数据库系统传统教程的扩充教程，又可将本书用作专门的分布式数据库教程。本书的初版已经在斯坦福大学作为分布式数据库教程使用。授课时间为一学期（每学年分为四学期制）。也在意大利作为分布式数据库专业人员的教程。授课时间为一周。

本书的编写方法与内容

本书由两部分组成：原理和系统。前两章讨论一些预备知识，第1章是关于分布式数据库的特点和问题的综述。对数据处理有兴趣的人无需任何专门预备知识都可读懂本章。第2章讨论数据库和计算机网络的一些问题。这是理解本书其它部分所必需的。当然，第2章并没有完整地、详细地讨论数据库和计算机网络这两个领域中的问题，而是重新定义一些术语、符号和基本概念，以使本书自成体系。然而，为了充分理解分布式数据库中的问题，读者应具备数据库和计算机网络方面的一些基础知识。

第一部分包括第3章到第10章，既讨论了分布式数据库的基本原理，也讨论了分布式数据库的基本技术问题。第3章从应用程序员的角度讨论了分布式数据库的体系结构，定义并举例说明了分布式数据库管理系统所提供的不同的分布式透明层次。本章具有指导性，且易于读。第4章讨论分布式数据库设计，重点强调在网络的的不同地点分割和分配数据的问题。本章对分布式数据库的设计者是非常有用的。为了从总体上搞清楚分布式数据库的特性，弄明白如何分布数据和为什么要分布数据是非常必要的。本章在风格上也是带有指导性的。

第5章和第6章讨论如何有效地访问分布式数据库（人们常把这个领域叫做分布式查询优化；然而我们要强调指出的是这些技术既可以用于设计一些重复执行的应用，也可以用于分布式查询优化）。这两章较前几章更难理解一些。这两章用综合的、但是系统的方式讨论了人们在这个领域中所研究出来的方法。第5章讨论如何把应用所要求的数据库操作变换成为语义上等价的对访问分布式数据库来说是适当的操作。第6章专门讨论查询优化问题。

第7章、第8章和第9章讨论事务处理的管理，即支持事务处理的有效的、可靠的和并行的执行。第7章讨论事务处理管理的基本技术以及这些基本技术在分布式数据库系统设计和实现中的应用。第8章讨论分布式并发控制技术，即允许在计算机网络的地点并行执行事务处理的技术。第9章讨论分布式数据库从故障中恢复的技术。与第5章和第6章一样，第8章和第9章也是比较难于理解的。

第10章是第一部分的最后一章，讨论数据管理问题。本章简单介绍了分布式数据库的安全性问题和目录管理问题。

第二部分讨论分布式数据库管理系统。重点介绍如何把第一部分中所讨论的技术应用到特殊的系统中去。第11章讨论如何用目前现有的系统来建立分布式数据库。以及建立什么样的分布式数据库。第12章讨论SDD-1，SDD-1是分布式数据库管理系统的原理样机，该样机是分布式数据库领域中的里程碑。第13章讨论R*系统。R*是最重要的。目前正在努力开发的分布式数据库系统。第14章概述其它同构型分布式数据库原理样机中所采用的方法。第15章讨论异构型分布式数据库领域中的主要研究原理样机。异构型分布式数据库是那些综合不同的局部数据库管理系统的系统。

如何阅读本书

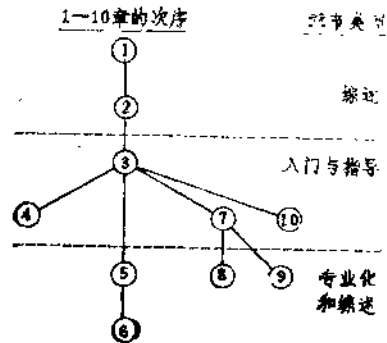
为了便于不同的读者阅读本书，我们是这样编排本书的：不同的读者可以用不同的方式阅读本书。既可以作为一种概述阅读本书，也可以作为一种关于分布式数据库的技术问题的指南阅读本书，也可以作为一种关于解决分布式数据库的问题所研究出来的技术综述阅读本书。

不同的读者可以按照右图来找出适合自己情况的阅读顺序。

图的左边示出了各章之间的优先顺序，图的右边示出了每一章的类型。例如，优先树表明，在阅读第8章之前，必须先阅读第7、3、2和1章。

当然也可按顺序从头至尾阅读本书。作为专门的分布式数据库教程来说，这种阅读方法是正确的。作为短期的、对传统的数据库管理教程的补充教程来说，则可跳过第一部分和第二部分中任何专门化的章节。

系统管理人员和专业人员也可跳过更专门化的章节。例如 仅对分布式数据库综述感兴趣，并且对已有的原理样机和系统感兴趣的读者只要阅读第1、2、3、4、7章和第二部分中的任一章即可。



斯蒂劳诺·塞里 吉乌塞普·皮拉盖蒂

目 录

编译说明	(I)
原作者序	(VII)
第 1 章 分布式数据库综述	(1)
1.1 分布式数据库与集中式数据库特点的比较	(3)
1.2 为什么要开发分布式数据库	(7)
1.3 分布式数据库管理系统(DDBMS)	(8)
结论	(10)
参考文献	(10)
第 2 章 数据库与计算机网络综述	(11)
2.1 数据库综述	(11)
2.2 计算机网络综述	(17)
参考文献	(21)
第一部分 分布式数据库原理	(22)
第 3 章 分布式透明性的层次	(23)
3.1 分布式数据库的参考体系结构	(23)
3.2 数据段存贮的类型	(25)
3.3 只读应用的分布式透明性	(29)
3.4 更新应用的分布式透明性	(35)
3.5 分布式数据库的访问基元	(38)
3.6 分布式数据库中的完整性约束	(42)
结论	(43)
练习	(43)
参考文献	(44)
第 4 章 分布式数据库的设计	(45)
4.1 分布式数据库设计的结构	(45)
4.2 数据库段存贮的设计	(48)
4.3 段分配	(56)
结论	(59)

练习	(59)
参考文献	(61)
第5章 全局查询到段查询的变换	(62)
5.1 用于查询的等价变换	(62)
5.2 把全局查询变换成段查询	(68)
5.3 对分布式组合与聚集函数的评价	(77)
5.4 参数型查询	(80)
结论	(82)
练习	(83)
参考文献	(84)
第6章 访问策略的优化	(85)
6.1 查询优化的结构	(85)
6.2 连接查询	(96)
6.3 一般查询	(111)
结论	(114)
练习	(114)
参考文献	(116)
第7章 分布式事务处理的管理	(118)
7.1 事务处理管理的组织	(118)
7.2 支持分布式事务处理的原子性	(123)
7.3 用于分布式事务处理的并发控制	(136)
7.4 分布式事务处理的体系结构	(140)
结论	(144)
练习	(145)
参考文献	(145)
第8章 并发控制	(147)
8.1 分布式并发控制基础	(147)
8.2 分布式死锁	(154)
8.3 以时标为基础的并发控制	(161)
8.4 用于分布式并发控制的优化方法	(166)
结论	(170)
练习	(171)
参考文献	(172)

第9章 可靠性	(174)
9.1 基本概念	(174)
9.2 非中断提交协议	(177)
9.3 可靠性和并发控制	(185)
9.4 确定网络的一致性视图	(190)
9.5 不一致性的检测和消除	(192)
9.6 检查点和冷重新启动	(194)
结论	(196)
练习	(196)
参考文献	(197)
第10章 分布式数据库的管理	(199)
10.1 分布式数据库中目录的管理	(199)
10.2 认可和保护	(203)
结论	(205)
参考文献	(205)
第二部分 分布式数据库系统	(207)
第11章 商品化系统	(208)
11.1 Tandem 公司 ENCOMPASS 分布式数据库系统	(208)
11.2 IBM 内部系统通信	(213)
结论	(217)
参考文献	(218)
第12章 SDD—1: 一种分布式数据库管理系统	(219)
12.1 体系结构	(219)
12.2 并发控制 (读阶段)	(220)
12.3 查询的执行 (执行阶段)	(222)
12.4 可靠性事务处理提交 (写阶段)	(224)
结论	(228)
第13章 R* 系统	(229)
13.1 R* 系统体系结构	(230)
13.2 查询的编译、执行和重新编译	(231)
13.3 视图管理	(234)
13.4 R* 中的数据定义协议与特许协议	(235)
13.5 事务处理管理	(238)

13.6 终端管理.....	(240)
结论.....	(241)
参考文献.....	(241)
第14章 其它同构型分布式数据库系统.....	(242)
14.1 DDM: 基于 ADAPLEX 语言的分布式数据库管理系统.....	(242)
14.2 Distributed—INGRES	(246)
14.3 POREL 系统.....	(248)
14.4 SIRIUS—DELTA	(251)
结论.....	(253)
参考文献.....	(255)
第15章 异构型分布式数据库系统.....	(257)
15.1 异构型分布式数据库的几个问题.....	(257)
15.2 MULTIBASE系统	(259)
15.3 DDTs: 分布式测试台系统.....	(268)
15.4 异构型SIRIUS—DELTA	(273)
结论.....	(274)
参考文献.....	(275)

第1章 分布式数据库综述

最近几年，分布式数据库已经成为信息处理学科的重要领域，可以预料其重要性还会迅速地增加。这种趋势有组织机构上的和技术上的两种原因：分布式数据库克服了集中式数据库的许多缺点，并且更加适应多个机构的分散式结构。

典型的、但还不很明确的分布式数据库的定义如下：分布式数据库是逻辑上属于同一系统、但又分布在计算机网络的多个地点的数据库的集合。该定义强调了分布式数据库的两个同等重要的方面：

1. **分布** 也就是说数据不是驻留在同一地点（处理机），这样我们就能够把分布式数据库与单个集中式数据库区分开来。

2. **逻辑相关** 也就是说数据具有某些相互联系的特性，因此我们能够把分布式数据库与一些驻留在计算机网络不同地点的局部数据库或文件集合区分开来。

上述定义所存在的问题是分布和逻辑相关这两种特性没有明确地定义出来，以致于常常不能把确实属于分布式数据库的情况和不属于分布式数据库的情况区分开来。下面通过对几个例子的分析与研究推导出更确切的分布式数据库定义。

例子1.1

研究一下在不同位置上拥有三个分行的银行。在两个分行，计算机控制分行的出纳终端和分行的账目数据库（图1.1）。在一个分行里，每个具有局部帐目数据库的计算机构成了分布式数据库的一个地点；计算机由通信网络连接。在正常操作期间，一个分行终端所要求的各种应用只需要访问该分行的数据库。这些应用完全由提出它们的分行计算机进行处理。因此，这些应用被称做局部应用。局部应用的例子是借贷应用和信用应用，由存贮在要求该类应用的同一分行的帐目完成。

如果我们试图把分布式数据库的定义用于迄今所讨论的场合，则发现很难说逻辑相关性是否成立。各分行包括有关同一分行账户的信息这一事实充分吗？应该把这看作是一个分布式数据库的例子呢，还是看作局部数据库的集合的例子呢？

为了回答这些问题，让我们集中讨论是什么因素使得处理一组局部数据库的数据，不同于处理被认为是一种分布式数据库的同一数据。从技术观点上讲，好象真正重要的方面是能在几个分行上访问数据的某些应用的存在。这些应用被叫做**全局应用或分布式应用**。就一组局部数据而言，全局应用的存在可被认为是分布式数据库的

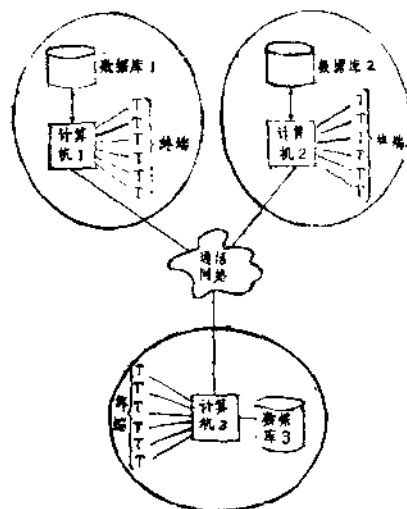


图1.1 以地区分散网络为基础的分布式数据库

判别特征。

一个典型的全局应用是从一个分行的账户到另一个分行的账户的现金汇兑。这种应用需要在两个不同分行上更新数据库。注意，这种应用不仅仅是在两个单独分行（借方和贷方）完成两个局部更新，因为它还必须保证或者两个均更新，或者两个都不更新。正如我们将要看到的那样，实现全局应用的这种要求是一件困难的任务。

在例子1.1中，计算机处于不同的地理位置，但分布式数据库也可以建立在局部网络上。

例子1.2

研究一下先前例子的同一银行，它具有相同的应用，但其系统结构如图1.2所示。拥有数据库的同一处理机已经从各分行移到公共办公楼中去了，而且用宽带技术把它们互连为局部网络。各分行的出纳员终端利用电话线连接到其各自的计算机上。每个子处理机和其数据库组成一个局部计算网络地点。

我们看到，和例子1.1相比，连接的物理结构已发生了变化，但在构造性方面仍然相同，特别是同样的计算机执行访问同样数据库的同样应用更是如此。如果不是根据执行该应用的计算机地理分布来确定局部性，而是根据仅涉及一个拥有自己的数据库的计算机来确定局部性，则在前一个例子中曾是局部应用的应用仍是局部应用。

如果存在全局应用，则可把这个例子看作为分布式数据库，因为刻划前一个例子的大部分特性仍是有效的。但是，分布式数据库是在局部网络上、而不是在地区网络上实现的，并且具有更高的吞吐量和可靠性，这将在一些情况下改变解决某些问题所采用的方法。让我们研究一个系统的例子，本书中不把这种系统看成是分布式数据库。

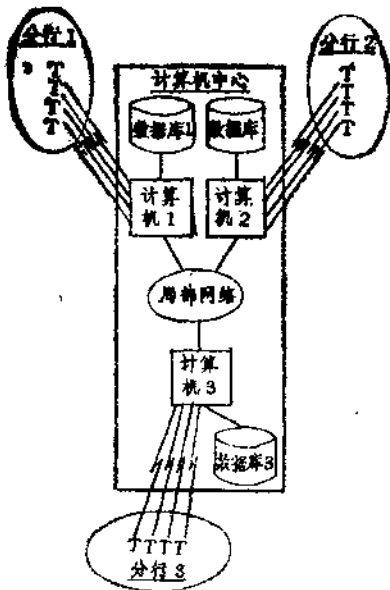


图1.2 以局部网络为基础的分布式数据库

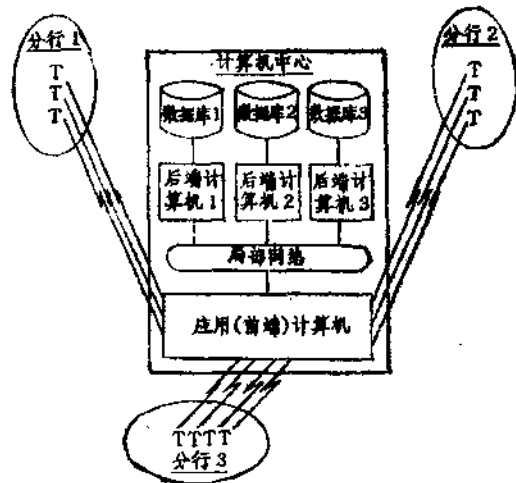


图1.3 多处理机系统

例子1.3

研究一先前例子中的同一银行，但该银行具有图1.3所示的系统结构。不同分行的

数据分布在三个“后端”计算机上，这些计算机执行数据库管理功能，应用程序由另外的计算机执行，当需要时，它们可以要求对后端数据库进行访问。

不把这种类型的系统看作是分布式数据库的原因是，尽管数据确实分布在不同的处理机上，但从应用观点来看，其分布是不相关的。这里系统整体化已达到其中一个计算机本身不能够处理一个应用的程度，这时我们失掉的是局部应用的存在。

总结上面几个例子，现在我们可以这样来定义分布式数据库：

分布式数据库是分布在计算机网络中不同计算机上的数据的集合。该网络的每个地点都具有自主处理能力，并能够执行局部应用。每个地点，至少参与处理一个全局应用，这需要利用通信子系统在几个点上访问数据。

分布式数据库的最重要的技术问题来源于把它们看做是一种“自主地点之间合作”的结果；上述定义突出了这个特点。

1.1 分布式数据库与集中式数据库特点的比较

分布式数据库不是集中式数据库设置的简单分布，因为它们允许设计与传统的集中式系统相比较具有不同特点的系统，因此有必要了解一下传统数据库的典型特点，并把它们与分布式数据库的相应特点比较一下。传统数据库处理所具有的特点是集中控制、数据独立性、减少冗余、用于有效访问的复杂物理结构、完整性、恢复、并发控制、保密性和安全性。

集中控制 在整个企业或机构范围的信息资源上提供集中式控制的可能性，被认为是采用数据库最重要的动力之一；它们是根据信息系统的演变开发出来的。在这种信息系统中，每一应用均具有其自己的专用文件。数据库管理员（DBA）的基本作用曾是保证数据的安全；数据本身被认为是一种要求集中负责的企业的重要资源。

在分布式数据库里，几乎不强调集中式控制概念。这也是由这种体系结构所决定，就像从例子1.1和1.2中所能看出那样。例子1.2比例子1.1可能更适合集中式控制。一般来讲，在分布式数据库中，能够识别出以一个集中负责整个数据库的全局数据库管理员为基础的分层控制结构。然而，必须强调指出，局部数据库管理员可能具有高度的自主权，最高可达到全局数据库管理员完全监听不到的程度，内部协调由局部管理员自己执行。这种特性通常被叫做**地点自主权**。分布式数据库在地点自主权程度上可能有极大的差别：从没有任何集中式数据库管理员的整个地点自主权到几乎完全的集中式控制。

数据独立性 数据独立性也曾被认为是采用数据库方法的主要动力之一。实际上，数据独立性意味着数据的实际结构对应用程序员来讲是透明的，利用数据“概念”图即所谓的**概念模式**编写程序。数据独立性的主要优点是程序不受数据物理结构变化的影响。

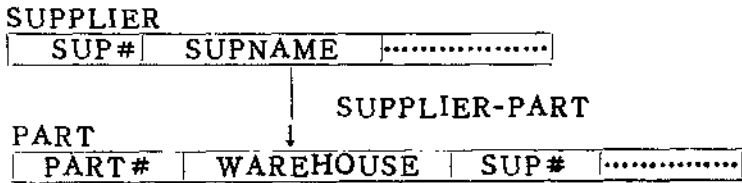
在分布式数据库中，数据独立性具有与传统数据库同等的重要性；然而一种新的概念加进了数据独立性的一般观念中，也就是**分布式透明性**。所谓分布式透明性，我们指的是在编写各种程序时就好像数据没有被分布一样。这样把数据从一个地点转移到另一个地点不会影响程序的正确性；但其执行速度会受到影响。

通过具有不同的数据描述和它们之间的映射的多层次体系结构，曾为传统的数据库提供了数据独立性；为此目的，开发了概念模式、存储模式和外部模式的概念，利用类

似的方法，通过采用新的层次和模式，在分布式数据库中获得了分布式透明性。第3章叙述了用于实现分布式透明性的可能途径。

降低冗余 在传统的数据库中，尽可能降低冗余的原因有两个：首先，通过只用一个副本自动地避免同一逻辑数据中几个副本之间的一致性。其次，用消除冗余来节省存储空间。通过共享数据方式，即通过允许几个应用访问同一文件和记录达到降低冗余的目的。

但是，在分布式数据库中，数据冗余被看成是所需要的特性有几个原因：首先，如果在应用需要冗余的所有地点都复制数据，则可以增加应用的局部性。其次，可以增加系统的有效性，因为如果复制数据，则地点故障不能停止执行其它地点上应用。一般来说，为传统集中式环境规定的克服冗余的原因仍是有效的。因此，最佳冗余度的评价要求一种比较复杂的折衷选择评价。作为最一般的说明，如复制一份数据项的方便程度随应用所执行的检索访问与更新访问的比率而提高。数据复制便利程度的提高是因为我们具有一个项目的多个副本，检索可以在任一副本上进行，而更新必须在所有副本上统一进行。我们打算在第4章中讨论分布式数据库的设计时，对这个问题再进行较为详细的讨论。



(a) Codasyl数据库模式

Find SUPPLIER record with SUP# =S1;

Repeat until "no more members in set"

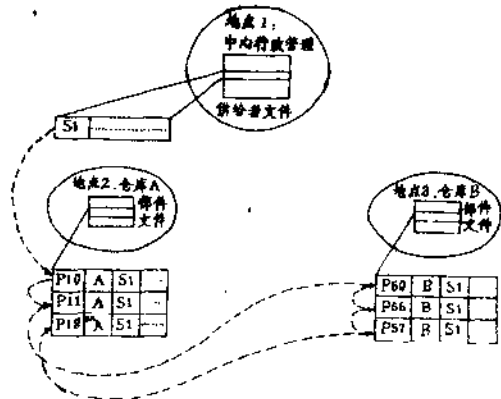
Find next PART record in SUPPLIER-PART set;

Output PART record;

(b) 用于找出由Supplier S1提供的部件的类Codasyl-DBMS程序

复杂的物理结构和有效访问 复杂的访问结构，如辅助索引、文件间的链接等，都是传统数据库的重要方面。对这些结构的支持是数据库管理系统（DBMS）最重要的部分。提供复杂访问结构的原因是为了获得对数据的有效访问。

在分布式数据库中，复杂的访问结构对有效访问来讲不是合适的工具。因此，当有效访问是分布式数据库的一个主要问题时，物理结构就不是一个相关的技术问题了。通过采用内部地点的物理结构不能保证有效地访



(c) SUPPLIER-PART集合的分布

图1.4 类似于Codasyl的分布式数据库

问分布式数据库。因为建立和保持这种结构是极困难的，同时也因为它不便于在分布式数据库记录层上“引导”(navigation)。我们可以用一个例子来说明这一论点。

例子1.4

研究一下图1.4a所示的类似于Codasy1数据库的数据模式。有两种记录类型，即SUPPLIER与PART和系型SUPPLIER-PART，这两种记录类型能把供给者记录连接到由它们所供给的部件的记录上。“find all PART records supplied by supplier S1”的应用按类Codasy1的DML编码，如图1.4b所示。

现在，我们假设上面所述的数据库分布在计算机网络的三个地点，如图1.c所示；供给者的文件存贮在地点1（中心行政管理），而部件文件又分成两个分别贮存在地点2和3（仓库）的子文件中。我们进一步假设我们拥有一个分布式的Codasy1系统工具，因此我们能够在分布式数据库上运行图1.4b给出的同一（引导）程序。假设该应用从地点1开始运行。很清楚，对“repeat until”循环的每次迭代来说该系统可能必须访问远程PART记录；因此就每次对记录的访问来讲，不仅必须传送记录本身，而且还可能必须交换几个信息。

同一应用的更有效的实现由分组构成，并尽可能包括所有远程访问，如图1.5所示。比较图1.4b和1.5中的过程；在图1.4中，子句“find next”需要在逐个记录的基础上进行访问；在图1.5中，子句“find all”能组合同一地点执行的所有访问。图1.5所示的过程由两种操作组成：局限在单个地点上程序的执行和两个地点之间文件的传送。类似上面的过程叫做**分布式访问计划**。

- 1) At site 1
Send sites 2 and 3 the supplier number SN
- 2) At sites 2 and 3
Execute in parallel, upon receipt of the supplier number, the following program:

Find all PARTS records having
SUP#=SN;
Send result to site 1.
- 3) At site 1
Merge results from sites 2 and 3;
Output the result.

图1.5 访问计划的例子

分布式访问计划可以由程序员编写，也可以由优化程序自动编写。在程序员规定怎样访问数据库的意义上讲，编写分布式访问计划类似于集中式数据库中的引导程序设计。然而，内部地点的引导应在几组记录层上完成，而对每一地点的局部处理来讲，可以执行常用的一次一个记录的引导。因此，和引导语言相比非过程面向集合的语言更适用于建立访问计划。

在设计优化程序时，仍要解决几个问题。优化程序能自动生成类似图1.5中所示的访问计划。把这些问题分成两类是方便的：**全局优化**和**局部优化**。全局优化旨在确定必须在哪些地点访问哪些数据，并确定哪些数据文件必须在两个地点之间传送。

尽管在某些情况下也应该考虑访问局部数据库的费用，但全局优化的主要优化参数是**通信费用**。这些因素的相对重要性取决于通信费用与磁盘访问费用之比，而这个比又取决于通信网络的类型。

局部优化包括确定如何在每一地点上完成局部数据库的访问；局部优化的问题是传统非分布式数据库所特有的，本书中不准备再对这些问题做进一步的讨论。

第5和6章专门研究开发全局访问计划的问题，以便在分布式数据库中实现应用。对全局优化的研究已经取得了一些有效的结果，即使访问计划不能自动生成，这些结果也是有效的，因为它有助于理解如何才能有效地访问分布式数据库的问题。

完整性、恢复和并发控制 在数据库中，尽管完整性、恢复和并发控制等问题涉及到不同的问题，但它们之间的相互联系还是很多的。在很大程度上这些问题的解决在于所提供的事务处理。**事务处理**是执行的**原子单位**，即它或者完全执行，或者完全不执行操作序列，例子1.1中给出的“资金转账”应用是一种必须是原于单位的全局应用；或者执行借方部分和贷方部分两者，或者都不执行；只执行其中一方是不合适的。因此，资金转账应用也是全局性事务处理。

显然，在分布式数据库中，转账事务处理的原子性问题具有特殊意义：当需要资金转账时，如果“借方”地点正在工作，“贷方”地点不工作，那么该系统应如何工作呢？即使两个地点不同步工作，也应该异常终止事务处理吗（取消在地点故障之前已经完成的所有工作）？或者有效系统应该试图正确进行资金转账吗？当然，如果采用后一种方法，用户可能极少受到故障的影响。

很清楚，原子事务处理是获得数据库完整性的方法。因为它们可以保证或者执行把数据库从一个一致状态转换到另一个一致状态的所有动作，或者使初始一致状态保持原样。

对事务处理的原子性有两个危害很大的问题：**故障**和**并发性**。故障可能会使系统在事务处理中间停止工作，这样会违背原子性要求。不同事务处理的并发执行可能使一个事务处理在其执行期间观察到由另一个事务处理产生的不一致性瞬时状态。

恢复在很大程度上涉及到存在故障时保持事务处理原子性的问题。在分布式数据库中，这个方面特别重要，因为事务处理执行时所涉及的某些地点可能会发生故障，就像前面那个例子所展示的那样。分布式数据库的恢复问题将在本书第9章中进行讨论。

并发控制涉及到在存在事务处理的并行执行时确保事务处理的原子性。这个问题可以看作是一个典型的同步化问题。在分布式数据库中，像在所有的分布式系统中那样，同步化问题比集中式系统中的同步化问题更难解决。这个问题将在第8章中加以讨论。

保密性和安全性 在传统数据库中，具有集中式控制的数据库管理员能保证只执行那些经授权的数据访问。但是请注意，集中式数据库方法本身（无专门的控制过程），比基于单独文件的较老式方法更难以实现保密性和安全性。

在分布式数据库中，局部管理员面临的一些主要问题与传统数据库的数据库管理员

面临的问题一样，但是，分布式数据库的两个特殊的方面值得一提：首先，在具有极高度地点自主权的分布式数据库中，局部数据的拥有者认为比较安全，因为他们能实施其自己的安全措施，以代替集中式数据库管理员；其次，安全性问题一般是分布式系统所固有的，因为保护是通信网络中可能出现的一个弱点。分布式数据库的保密性问题将在第10章中加以讨论。

1.2 为什么要开发分布式数据库

开发分布式数据库有几个原因。下面我们列举一下其中的一些主要原因。

组织和经济原因 许多组织不集中在一起，分布式数据库方法能更切合实际地满足组织机构的要求。分布的组织机构和相应的信息系统是一些书籍和论文讨论的课题。根据计算机技术中最新的发展，再建立大型集中式计算机中心从经济观点讲是不合算的。本章里我们不准备进一步讨论这个课题。但是，组织机构和经济上的考虑可能是开发分布式数据库最重要的原因。

现有数据库的互连 当在一个组织中已经存在几个数据库，且实现全局应用的必要性增加时，分布式数据库就是合乎道理的解决方法。在这种情况下，就可由先前的局部数据库自底向上构成分布式数据库。这个过程可能需要一定程度的局部重建。然而，这种重建所需要的工作量要比建成一个全新的集中式数据库所需要的工作量少得多。

扩充 如果一个组织机构需要增加新的相对自主的组织单位(新分支、新存储室等)来扩充组织机构，则分布式数据库方法能在对现有单位影响最小的条件下支持稳步的扩充。如果采用集中式方法，则在系统最初设计时就要考虑到将来的扩充，而这种扩充是难以预测的和代价昂贵的，同时这种扩充不仅对新应用影响很大，而且对现有应用也影响很大。

减少通信额外开销 在类似例子1.1数据库的按地理划分的分布式数据库中，与集中式数据库相比，许多应用是局部的这一事实可显著地降低通信额外开销，因此应用局部性的最大化是分布式数据库设计中要实现的主要目标之一。

性能考虑 几个自主式处理机的存在能通过高并行处理来提高性能。这种考虑可适用于任何多处理机系统，不仅仅适用于分布式数据库。例如，例子1.3的体系结构也允许并行处理，但该体系结构不能看成是分布式数据库。分布式数据库的优点在于数据的分解反映了使应用局部性达到最大的应用相关准则：采用这种方法可使不同处理机之间的相互干扰降到最低。负载可在不同处理机之间分担，并可避免临界瓶颈问题，如通信网络本身或整个系统的公用服务。就分布式数据库定义中所叙述的局部应用来讲，这种影响是自主处理能力要求的结果。

可靠性和可用性 分布式数据库方法，特别是具有冗余数据的分布式数据库方法也可以用来获得较高的可靠性和可用性，但是要达到这个目标是不容易的，需要采用目前还没有完全解决的各种技术。不同地点的自主处理能力不能单独地保障系统具有较高的总可靠性，但它能保证一种工作可靠，但性能下降的特性。换句话说，分布式数据库中的故障可能比集中式数据库中的故障出现得更频繁，因为它具有相当多的部件，但每个故障的影响仅限于使用故障地点数据的那些应用，极少影响整个系统。建造可靠的分布式数

数据库的各种技术将在第9章中予以介绍。

上面谈到的开发分布式数据库的各种动机都不是新东西。那么为什么现在才开始开发分布式数据库呢？这两个主要原因：首先，小型计算机最新发展能以较低的价格提供以前曾由大型计算机提供的许多功能，为开发分布式信息系统提供了必要的硬件支持；其次，分布式数据库的技术基于另外两种在七十年代已打下牢固基础的技术，计算机网络技术和数据库技术。在计算机网络上建立分布式数据库，并在每个地点建立一套局部数据库管理系统是一项复杂的任务；没有这些积木式组件，分布式数据库是不可能实现的。

1.3 分布式数据库管理系统 (DDBMS)

分布式数据库管理系统支持分布式数据库的建造和维护。在分析DDBMS特性时，区分商品化的系统和先进的研究样机是不难的。这种区分是基于目前的技术工艺水平。很清楚，目前在某些先进的研究样机中试验的特性可能会应用到将来的商品化系统中。在本节中给出一些提示，这些提示表明这些问题解决起来是多么困难，以及用什么样的方法才能使新的技术应用在商品化的系统中；在本书的第二部分，我们叙述了一些商品化的系统（第11章）和大部分研究样机（第12—15章）。

几种已经商品化的分布式系统是由集中式数据库管理系统的销售商开发出来的。这些分布式数据管理系统包括了扩充集中式DBMS功能的附加部件。这种功能扩充是通过安装在计算机网络不同地点的DBMS的几个例子之间提供通信和合作来实现的。在这种场合建造分布式数据库通常所必需的软件部分是：

1. 数据库管理部分 (DB)
2. 数据通信部分 (DC)
3. 数据字典 (DD)，它能用来表示有关网络中数据分布的信息
4. 分布式数据库部分 (DDB)

对于两个地点的网络来讲，这些部分按图1.6所示的那样连接。

我们将用“分布式数据库管理系统”这个术语来表示上面四个部分的所有内容，而DDB只表示专用的分布式数据库部分。同样，我们将用“数据库管理系统”这个术语来表示用于管理非分布式数据库的各部分的内容，即DB、DC和DD部分。

由上述系统类型所支持的各种服务一般是：

1. 通过应用程序进行远程数据访问；这个特性是最重要的特性。它可由具有分布式数据库的所有系统提供。
2. 某种程度的分布式透明性；这个特性可由不同的系统支持到不同的程度，因为在分布式透明性和系统性能之间存在一个妥善的折衷方案。
3. 对数据库管理和控制的支持；这个特性包括几种工具，即用于监控数据库的工

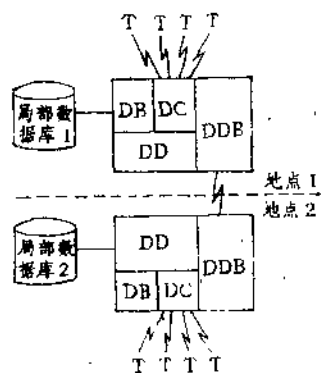


图1.6 商品化的DDBMS的各个部分