[美] S.塞里 等

分布式数据库原理和系统

关英春 等译　　张仲义 徐 悦 校

水利电力出版社

分布式数据库原理和系统

[美] S.塞里 等

关英春 等译　　张仲义 徐悦 校

水利电力出版社

Ceri, Stefano.
DISTRIBUTED DATABASES
Principles and Systems
McGraw-Hill computer science series 1984
分布式数据库原理和系统
[美] S.塞里 等
关英春 等译 张仲义 徐悦 校
*
水利电力出版社出版、发行
(北京三里河路6号)
各地新华书店经售
水利电力出版社印刷厂印刷
*
787×1092毫米 16开本 16.5印张 368千字
1989年11月第一版 1989年11月北京第一次印刷
印数0001—2330册
ISBN 7-120-00958-3/TP·32
定价13.00元

内 容 提 要

本书系统地介绍了分布式数据库的原理和典型的分布式数据库系统。全书共分为十五章。第一章和第二章介绍分布式数据库概况和对数据库、计算机网络的讨论。第三章至第十章介绍分布式数据库的原理，论述了分布式数据库的体系结构、设计、查询和优化、并行控制技术、数据管理功能，以及故障恢复技术等。第十一章至第十五章介绍SDD-1和R*分布式数据库系统，以及同构型和异构型的研究样机。

本书适用于大学计算机专业的大学生、研究生和教师，可作为教材或教学参考书，也可以作为计算机科学技术的研究人员和其它专业人员的参考书。

前　言

近年来，随着计算机和通信技术的发展，数据库和计算机网络的广泛应用，出现了一个新的领域，就是分布式数据库。分布式数据库是建立在计算机网络上的，并不是建立在单个计算机上的数据库，它的数据存储在计算机网络上的不同节点。

建立和实现一个分布式数据库有许多新的问题需要解决，为此，人们已经进行了大量地研究工作，并形成一个新的学科。为了理解分布式数据库，不仅需要了解传统的数据库和计算机网络的原理，还需要学习一些新的技术。本书提供了这种新的技术原理。

本书适用于对分布式数据处理有兴趣的专业人员，例如，计算机科学专业的学生和教师、研究人员、系统管理人员、系统或应用的设计人员、分析人员和程序设计人员等。本书可以作为分布式数据库课程的教材，也可以作为数据库系统课程的一部分。这本书的最初版本已经在美国斯坦福大学（Stanford University）和意大利等地方作为分布式数据库课程讲授。

本书的第一章是分布式数据库概述，在数据处理方面有兴趣的任何人都可阅读，并不需要任何专门基础知识。第二章对数据库和计算机网络进行了某些讨论，这是为了理解本书的其它各章所需要的。当然，第二章仅是重新定义了术语、符号和基本概念，使本书自成体系。以上两章是预备性质的。

本书的第三章至第十章是分布式数据库原理部分，论述了分布式数据库的全部技术问题。第三章是从应用程序员的观点论述了分布式数据库的体系结构。第四章是关于分布式数据库的设计，它论述了在计算机网络的不同节点上划分和分配数据问题。这一章对分布式数据库的设计更为有用，它是理解分布式数据库性质的基础。第五章和第六章讨论分布式数据库的查询和优化问题。第七章、第八章和第九章论述了事务管理，第七章提出了在分布式数据库的设计和实现中事务管理，以及它们的综合的基本技术。第八章评论了分布式并行控制技术，也就是允许在不同节点的事务处理的并行执行技术。第九章评论了分布式数据库的故障恢复技术。第十章论述了数据管理功能。这一章实质上是介绍分布式数据库的目录管理和安全问题。

本书的第十一章至第十五章介绍分布式数据库管理系统。第十一章指出用现在市场上购得到的系统如何构成分布式数据库。第十二章介绍SDD-1实验性的分布式数据库管理系统，它是一个具有代表性的系统。第十三章介绍R*系统，它是开发分布式数据库的最重要的研究成果。第十四章是对其它同构型研究样机的综述。最后，第十五章介绍在异构型分布式数据库管理系统领域中主要的研究样机。

本书由关英春翻译定稿。参加初稿翻译工作的，还有田盛丰、戴晓英、吴旭丽和毓钧等同志。在本书翻译过程中，解凯、车红、何京翔和富秀琴等同志给了许多帮助，并进行

了部分译稿、图表的整理和抄写工作；全书由张仲义、徐悦两位同志校订，在此均表示感谢！

限于译者水平，书中错误和不妥之处，请广大读者批评指正。

译　者

1988年8月

目　　录

第一章　分布式数据库概述

近几年来，分布式数据库已成为信息处理的一个重要领域，并且可以预见它的重要地位还将迅速上升，这是由结构和技术两方面原因决定的：分布式数据库消除了集中式数据库的许多缺点，而且它更适合于各种分散结构。

分布式数据库一个典型却又相当含糊的定义是：分布数据库是一种数据的集合，这些数据逻辑上属于同一系统，但却分布于一个计算机网络的各个节点。这个定义强调了分布式数据库的两个同等重要方面：

（1）分布性。即数据并非存储在同一个节点上。根据这一点我们可以对分布式数据库与单一的集中式数据库加以区分。

（2）逻辑相关性。即数据具有相互关联的特性。根据这一性质我们可以将分布式数据库与一组局部数据库或存于计算机网络不同节点的文件区分开来。

上述定义存在的问题是：分布性和逻辑相关性这两个性质定义非常含糊，它不能确切地辨别哪些情形真正属于分布式数据库。为了给出一个更明确的定义，让我们来看几个例子。

【例 1-1】 某银行在不同地点设有三个分行。在每一个分行，计算机控制它的出纳终端和帐目数据库（见图1-1），每台计算机连同分行的局部帐目数据库构成分布式数据库的一个节点，各计算机是通过通信网络连接起来的。在一般情况下，由分行终端发出的应用请求只需访问本分行的数据库，这些应用请求完全由发出请求的分行计算机来执行，因此叫做局部应用请求。分行办理只存于本分行那些帐目的借贷业务就是局部应用请求的一个例子。

如果我们打算用分布式数据库的定义来解释上述情况，就会感到难以断定逻辑相关这一性质在此是否适用。各分行只容纳涉及本行帐目的信息是否足够？上述例子应该属于一个分布式数据库还是一组局部数据库？

为了回答这些问题，让我们集中考察一下使用一组局部数据库与使用存放相同数据的分布式数据库的差别究竟在哪里。从技术观点来看，最重要的差别是由于存在某些应用请求，这些存取数据的请求是在一个以上的分行提交的，应用请求叫做全局应用请求或分布应用请求　全局应用请求的存在被认为是分布式数据区别于一组局部数据库的辨别特性。

从一个分行的帐目到另一个分行帐目的现款汇兑就是一种典型的全局应用请求。这种应用请求要修改两个不同分行的数据库。请注意这种应用请求不仅仅是在两个独立分行执行局部数据的修改（借方或贷方），还必须确认两个修改都被执行或都没被执行。要做到对这种应用请求的确认可不是件容易的事。

在[例1-1]中，计算机分布在地理上不同的地点；但是，分布式数据库也能建立在局部网络上。

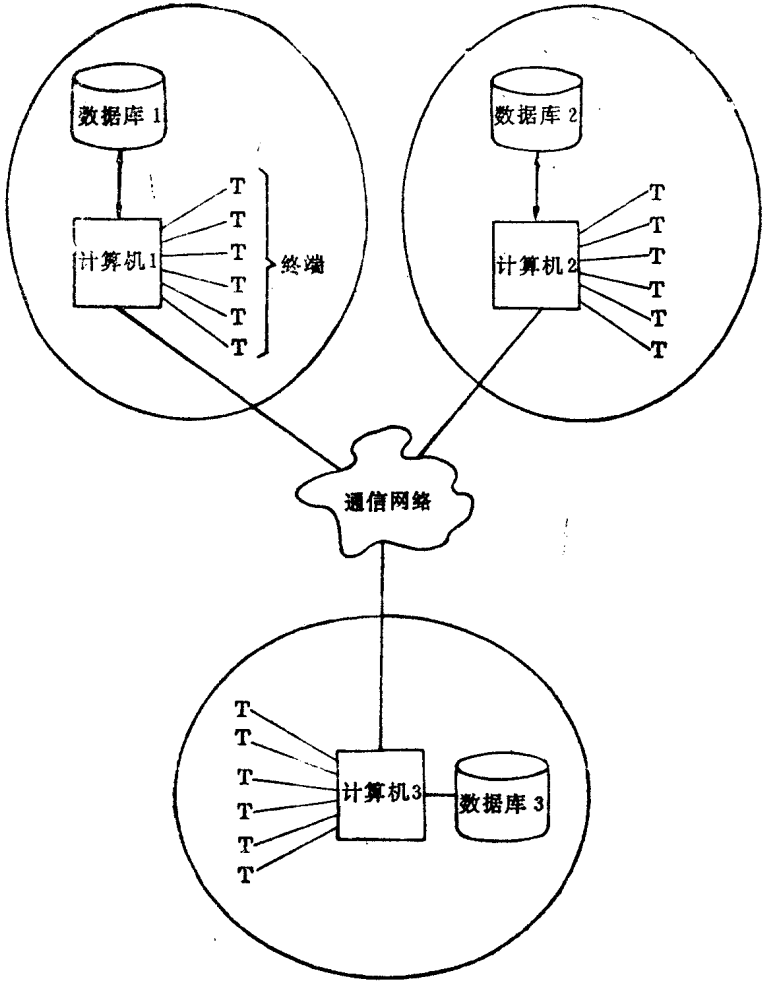


图 1-1 地区分散网络上的分布式数据库

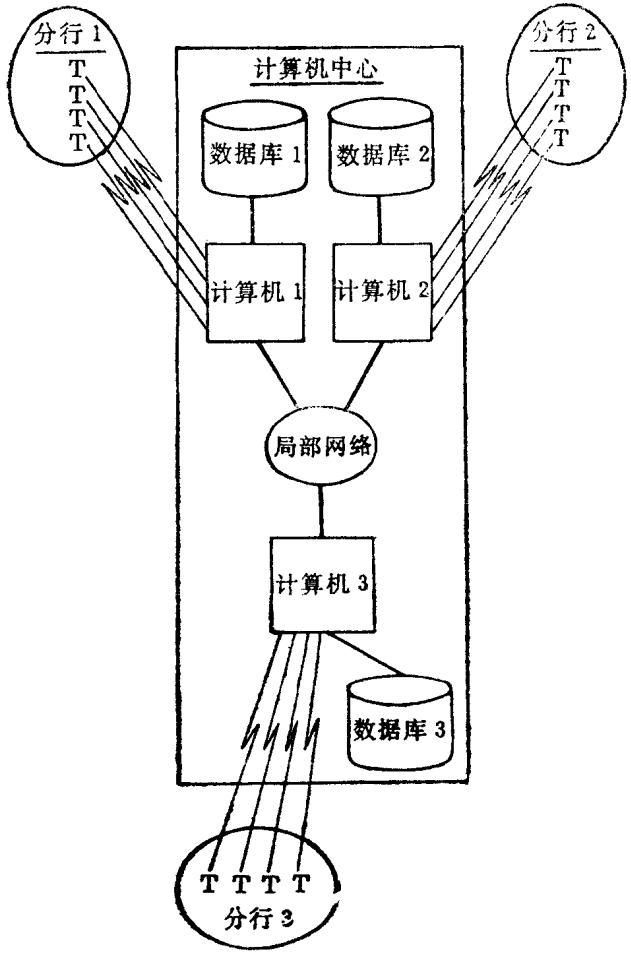


图 1-2 局部网络上的分布式数据库

【例 1-2】 还是上例中的银行，其应用请求不变，但是系统结构如图1-2所示。同样的处理机连同它们的数据库已经从分行移进二座公共大楼内，并且与高带宽的局部网络相连接。各分行的出纳终端用电话线连接到相应的计算机上，每台处理机和它的数据库构成局部计算机网络的一个节点。

我们可以看到连接方式的物理结构与[例1-1]相比有所变化，然而，系统结构方面的特性仍然相同。仍然是同样的计算机访问同样的数据库，执行同样的应用请求。如果局部性不是根据它的计算机地理位置的分布性来定义，而是指每台计算机只控制它本身数据库的话，那么前例中的局部应用请求在这儿仍然是局部性的应用请求。

这个例子如果存在全局应用请求，那么就可以认为它是一个分布式数据库，因为它仍然具有描述前例的大部分特征。但是，分布式数据库不在地区网络上，而在局部网络上实现会有高得多的吞吐量和可靠性，而且在某些情况下对一些问题的解决方法也有所变化。下面让我们来考察在本书中认为不是分布式数据库系统的例子。

【例 1-3】 与前例相同的银行，它的系统结构如图1-3所示，各分行数据分散在3个后端计算机上，这些计算机执行数据库管理功能。应用请求程序由另一台计算机执行，当需要时，它便向后端计算机请求数据库存取服务。

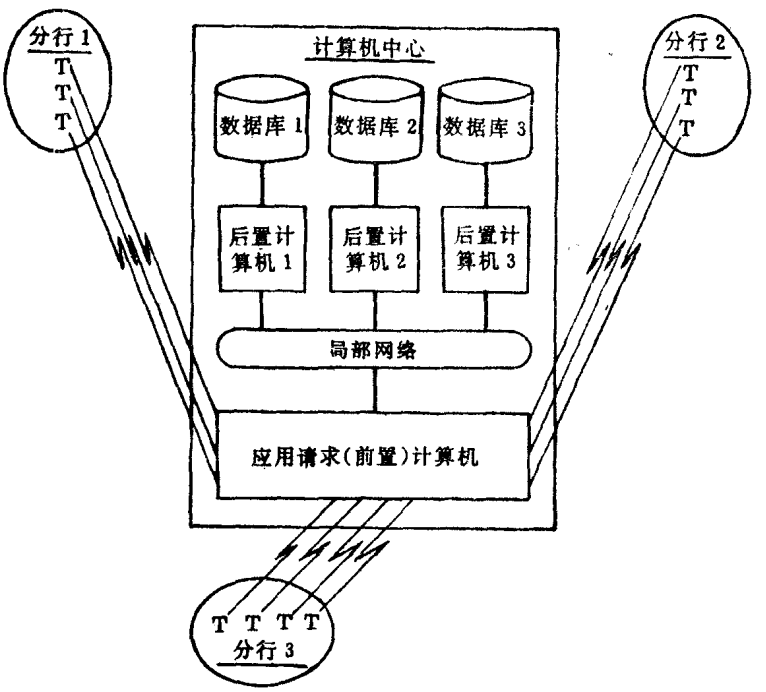


图 1-3 多处理机系统

这类系统不是分布式数据库。其理由是：虽然从物理上数据分布在不同的处理机内，但从应用请求的观点看数据的分布不是相关的。这里我们找不到局部应用请求，因为整个系统没有一台计算机可以单独地完成一个应用请求。

现在，我们可以把以上几个例子中所得到的观点归结为下面的定义：

分布式数据库是数据的集合，这些数据分布于一个计算机网络的各个计算机内，网络的每个节点有独立的处理能力，并且能执行局部应用请求，每一个节点也要参与至少一个全局应用请求，这种应用请求需要在多个节点上使用通信子系统来存取数据。

分布式数据库最重要的技术问题是自主节点间的协同作用。上面的定义强调了这一点。

第一节　分布式数据库的特点

分布式数据库不是集中式数据库的简单分布的实现，因为它们的系统设计所体现的特点与传统的集中式系统不一样。因此，了解一下传统数据库的典型特点，并将它们与分布式数据库相对应的特点进行比较是很有用的。传统数据库方法的特点是：集中控制、数据独立、低冗余度、实现有效存取、完整性、恢复、并行控制、保密和安全性的复杂物理结构。

一、集中控制

能够对整个企业或机构的信息资源提供集中控制，是主张采用数据库最强有力的动机之一。数据库是随着信息系统的演变而发展起来的，在这些信息系统中，每个应用程序有它自己的专用文件。数据库管理员（DBA）的基本任务是保证数据的安全；数据本身已被当作需要集中任务的企业的重要投资。

在分布式数据库中，强调集中控制的思想要少得多，这也取决于系统结构。[例1-2]比起[例1-1]来更适合于集中控制。一般来说，在分布式数据库中可以认为存在一种全局数据库管理员和局部数据库管理员的分层控制结构。全局数据库管理员负责整个数据库，而局部数据库管理员只负责各自的局部数据库，但是，局部数据库管理员可能有更高的自主性，甚至节点间的协调仅由局部DBA自己来完成，从而不再需要全局数据库管理员。这一特性通常叫做节点自主性。分布式数据库在节点自主性的程度上可能有很大的不同，没有任何集中数据库管理员能够从完全节点自主性到几乎完全集中控制。

二、数据独立

数据独立也被认为是推动数据库方法的主要动力之一。实质上，数据独立意味着实际的数据结构对于应用程序员是透明的。程序根据概念模式来编写。所谓概念模式就是从用户观点出发来看待数据的一种概念化的视图。数据独立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，数据独立与在传统数据库中具有同样重要地位，并被赋予了新意，即分布透明性。分布透明性的含义是，编写程序时好象数据库不是分布式的，因此，程序的正确性不受数据从一个节点到另一个节点移动的影响；但是，程序的处理速度要受些影响。

在传统数据库中，数据独立性是通过多层次体系结构实现的，不同层次具有不同数据描述，层次之间具有映象关系。概念模式、存储模式以及外部模式等术语就是用来描述这

种结构的。用类似的方法，分布式数据库通过提供新的层次和模式实现分布透明性。第三章将讨论实现分布透明性的可能方法。

三、低冗余度

在传统数据库中，要求尽可能减少数据冗余的原因有两个：第一，同一种逻辑数据只有一份副本从而自动避免了因将数据重复存放在好几份副本上引起的数据不一致；第二，消除冗余可以节约存储空间。低冗余度通过数据共享得到，就是说让几个应用请求存取相同的文件和记录。

然而，在分布式数据库中，由于某些原因把数据冗余看成为一种理想的特性：首先，如果数据在应用请求需要它的所有节点都重复存储，那么应用请求的局部性可以增加；其次，系统的有效性可以提高，因为如果数据重复的话，一个节点出了故障不至于中止执行其它节点的应用请求。一般来说，在传统数据库环境下反对冗余的同样理由仍然是适用的。因此，最佳冗余度的确定需要进行相当复杂的折衷和评价。我们可以非常概括地认为，随着应用请求的执行检索存取与修改存取比率的提高，复制某个数据项带来的方便将增加。如果我们有一个数据项的几个副本，检索操作就可以在任意一个上进行（但修改操作必须在所有的副本上一致进行），所以提高了数据重复的方便。对于这个问题将在论述分布数据库设计（见第四章）时详细讨论。

四、复杂的物理结构和有效存取

象二次索引、文件连接等复杂的存取结构是传统数据库的主要方面，对这些结构的支持是数据库管理系统的最重要部分。提供复杂存取结构的目的是为了获得数据的有效存取。

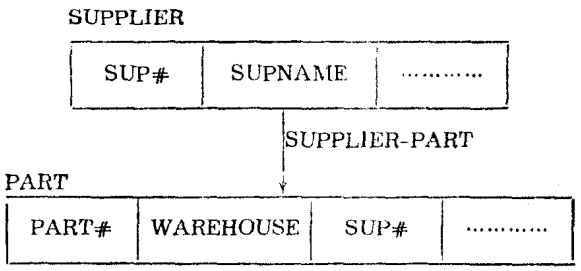
在分布式数据库中，复杂的存取结构不是获得有效存取的正确方法。因此，尽管有效存取是分布式数据库的一个主要问题，但物理结构却并不是相关的技术问题。分布式数据库的有效存取不能通过节点间的物理结构提供，因为建立和维护这样的结构是非常困难的，并且在分布式数据库中，按导航方式查找记录很不方便。让我们通过例子来说明这个观点。

【例 1-4】 如图1-4(a)所示的Codasyl-like数据库模式有两个记录类型——SUPPLIER和PART，类型SUPPLIER-PART连接了SUPPLIER记录和PART记录。应用请求“查找由供应商S1提供的全部PART记录”的Cadasyl-likeDML语句，如图1-4(b)所示。

我们假设上述数据库是分布在一个计算机网的三个节点上，如图1-4(c)所示，供应商文件存于节点1（中央管理），而PART文件分为两个不同的子文件，并设置在2和3的两个节点上（1货栈）。我们进一步假设有一个分布式执行的Codasyl系统，这样我们能够在分布数据库上运行同一个如图1-4(b)上的程序。假设应用请求来自节点1、很明显，对“repeat until”，每次迭代，系统将必须存取一个远程的PART记录，因此对一个记录的每次存取，不仅必须传送记录本身，而且要交换几个报文。

对相同应用请求的更有效的实现途径是尽可能地集聚所有远程存取，如图1-5所示。比较图1-4(b)和图1-5的程序：前者，“find next”语句需要一个记录一个记录地存取；后者，“find all”语句集聚了在同一节点执行的所有存取。如图1-5所示的过程由两种

类型操作组成：位于单个节点程序的执行和节点之间文件的传送。象上面那样的过程被称作分布存取计划。



(a)

```
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)

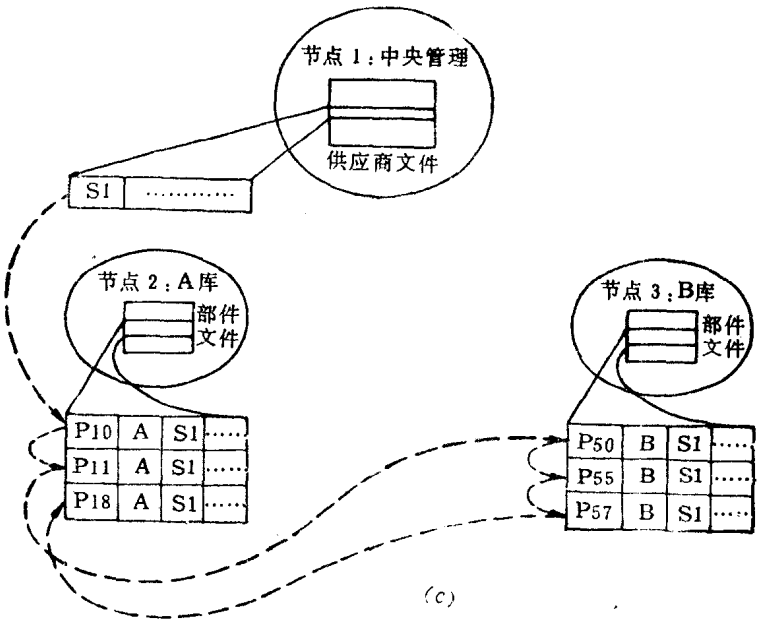


图 1-4 类Codasyl分布式数据库

(a)Codasyl数据库模式；(b)寻找由供应商S1提供部件(part)的类Codayl-DBMS程序；(c)SUPPLIER-PART集的分布

(1)在节点 1，发送供应商号码SN给节点 2 和节点 3 。

(2)在节点 2 和节点 3 ，依据接收供应商号码，按下述程序并行执行：寻找所有的PARTS记录，则有SUP# =SN；结果送到节点1。

(3)在节点1，归并由节点 2 和节点 3 的结果；输出结果。

图 1-5 存取计划例子

一个分布存取计划可由程序员编写或由优化程序自动生成。在程序员必须详细规定如何存取数据的意义上，编写一个分布存取计划和在集中式数据库中编写导航程序是相似的。但是，节点之间导航必须在记录组级执行，而在每个节点的局部处理可以执行每次一个记录的导航。因此，导航语言不如非过程的、面向集合的语言更适合于制定存取计划。

有几个问题在优化程序设计中需要加以解决，这种优化程序能自动生成如图 1-5 所示的存取计划。这些问题可分为两大类，全局优化和局部优化。全局优化就是决定哪些数据必须存在哪些节点和哪些文件必须在节点间传送。虽然在某些情况下，存取局部数据库的费用也要考虑，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依赖于通信费用和磁盘存取费用的比率，还依赖于通信网络的类型。局部优化就是决定如何执行每个节点上局部数据库的存取；局部优化问题是典型的传统非分布式数据库的问题，在本书中将不进行深入研究。

第五章和第六章对全局存取计划的问题进行了介绍。全局优化的研究已经取得了卓有成效的结果，即使存取计划还不能自动产生，但它有助于了解一个分布式数据库如何进行有效地存取。

五、完整性、恢复和并行控制

在数据库中，完整性、恢复和并行性控制虽然涉及不同的内容，它们却有很强的内在联系。在很大程度上，对这些问题的解决就是提供事务处理。事务是一个执行的原子单位；就是说：它是一系列操作，或者全部执行，或者一个都不执行。在[例1-1]中给出的“现款汇兑”应用请求是一个全局应用请求，它必须是一个原子单位：借方和贷方要么执行，要么都不执行，只执行其中一方是不允许的。因此“现款汇兑”应用请求也是一个全局事务。

很明显，在分布式数据库中事务的原子性问题有特殊的意义：当需要“现款汇兑”时，如果“借方”节点是可操作的而“贷方”节点是不可操作的，那么系统应该如何动作呢？是异常终止该事务呢？还是有一个灵活的系统，在两个节点完全不能同步操作的情况下仍试图去正确执行现款汇兑呢？当然，如果采用后一种方法，用户受故障的影响将会少些。

显然，原子事务是获得数据库完整性的方法，因为它们确保数据库要么从一个相容状态转换到另一个相容状态的所有动作都被执行，要么初始的相容状态保持不变。

事务原子性有两个危险的敌人，就是故障和并行性。故障可以导致系统在事务执行中间停止，从而破坏原子性的要求。不同事务的并行执行，可能会使一个事务监测到另一个事务执行过程中产生不相容的瞬时状态。

数据库恢复主要处理发生故障时维护事务原子性的问题。在分布式数据库中，这点尤其重要，如前面的例子所示，事务执行的有关节点往往会发生故障的。分布式数据库恢复的问题将在第九章中讨论。

并行控制在并行执行事务的情况下确保事务的原子性。这个问题可认为是一种典型的同步问题。象所有的分布式系统一样，在分布式数据库中解决同步问题比在集中系统中要困难些。这个问题将在第八章讨论。

六、保密性和安全性

在传统的数据库中，数据管理员掌握集中控制权，他能够确保只有经过许可的数据存取才能被执行，但是要知道集中式数据库方法本身，如果没有特殊的控制过程，其保密性和安全性比基于单个文件的早期方法更易遭到破坏。

在分布式数据库中，局部管理员基本上面临着传统数据库中数据库管理员所遇到的同样问题，然而，分布式数据库两个特殊方面值得提一下。首先，在具有高度节点自主性的节点的分布式数据库中，局部数据拥有者感受到更强的保护，因为他们可以不依赖于中心数据库管理员而实行他们自己的保护；其次，安全问题对分布式系统来说是最根本的问题，因为通信网络对提供保护来说是个薄弱环节，分布式数据库的保密性和安全性将在第十章中讨论。

第二节　为什么需要分布式数据库

分布式数据库的发展有以下几个主要原因。

一、组织上和经济上的原因

许多组织是非集中的，分布式数据库方法非常自然地适合于这些组织的结构。分布式组织结构和相应的信息系统的问题是一些书和文章的主题。随着近年来计算机技术的发展，从经济的规模因素考虑，追求大型集中式计算机中心的想法越来越站不住脚。这里，我们不进一步讨论这个问题，但是，组织上和经济上的刺激很可能是发展分布式数据库的最重要的原因。

二、现有数据库的互连

当一个组织已经有几个数据库而且执行全局应用请求的要求增加时，分布式数据库是一个自然的解决方法。这种情况下，分布式数据库由已经存在的局部数据库自底向上生成。这种处理可能需要一定程度的局部调整，然而，这种调整所需要的努力比建立一个全新的集中数据库所需的努力要小得多。

三、增量的增长

如果一个组织因增加新的相对独立的组织单位（新分行、新仓库等）而扩大，那么，分布式数据库可以使这种增长平稳地进行，对现有单位的影响减小到最低程度。用集中式的方法，或者最初建立系统时就考虑今后的扩充，这往往是难以预见的。实现费用也很大，或者系统增大时不仅对新的应用，而且对已经存在的应用都带来很大的影响。

四、减少通信开销

在一个象[例1-1]那样地理上分散的数据库中，许多应用请求是局部应用请求，相对于集中数据库来说明显地减少了通信开销。因此，尽可能提高应用请求的局部性是分布式数据库设计的主要目标之一。

五、性能上的考虑

几个独立处理机的存在，通过高度并行性使性能得到提高，这种想法不仅适用于分布