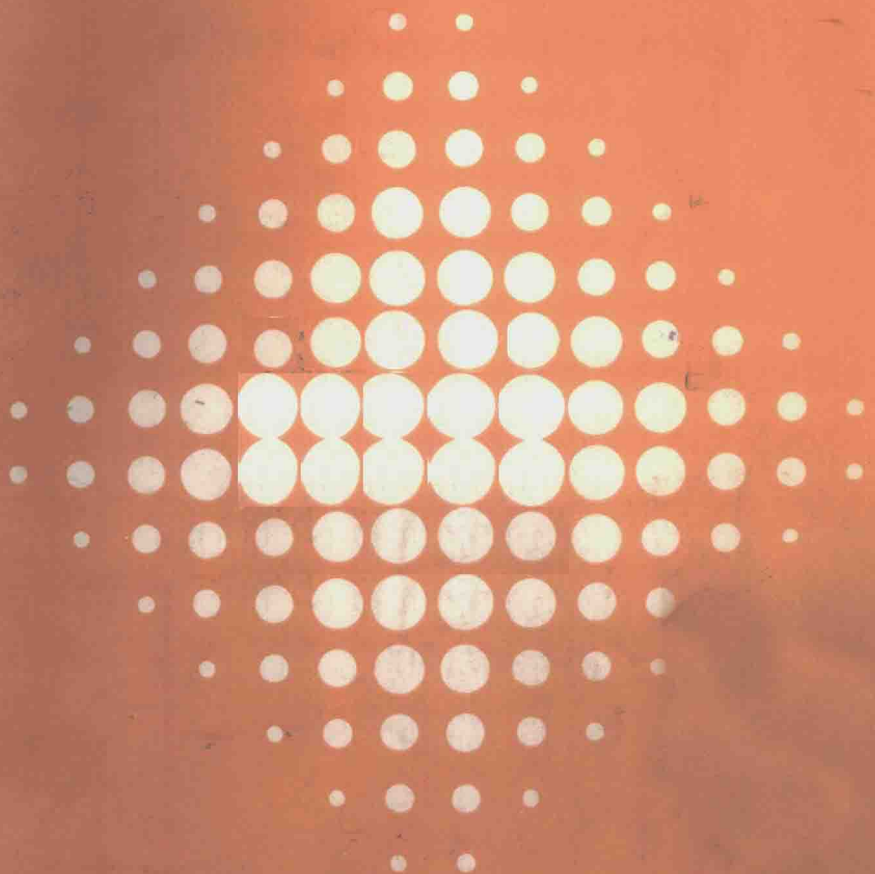
高等学校教材

分布式数据库系统

王以和　涂小平　编



電子工業出版社

高等学校教学用书

分布式数据库系统

王以和　涂小平　编

電子工業出版社

内 容 提 要

本书内容分为三大部分：第一篇是分布式数据库基础；第二篇是关于分布式数据库的原理，讨论分布式数据库的所有技术问题，是本书的主要部分；第三篇介绍了一些分布式数据库管理系统的实例，SDD-1及R* 系统等。

本书作为高等院校数据库方面传统课程的扩充，也可作为专门讲述分布式数据库课程的教材。

分布式数据库系统

王以和　涂小平 编

责任编辑：王惠民

电子工业出版社出版（北京海淀区万寿路）

新华书店北京发行所发行　各地新华书店经售

北京科技印刷厂印刷

开本：787×1092毫米 1/16 印张：17.125 字数：416千字

1988年6月第一版　1988年6月第一次印刷

印数：1—6000册　定价：2.85元

ISBN 7-5053-0082-2/TP·9

前　言

众所周知，七十年代中已经广泛地使用计算机来建立功能强大的集成式数据库系统。因此，数据库系统技术已经建立起它的理论基础，并且在大量的应用中取得了经验。与此同时，计算机网络也得到了广泛的发展，它能够把不同的计算机连接起来，并在它们之间交换数据和共享其他资源。

近几年来，数据库和计算机网络的应用已经产生了一个新的领域：分布式数据库。简单地说，分布式数据库也是一种集成数据库，不过它是建立在计算机网络之上而不是在单机之上。这种数据库中的数据存放在计算机网络的不同站点之中，而且计算机中运行的应用程序要从不同的站点存取数据。

要建立和实现分布式数据库会遇到全新的问题，为了解决这些问题已经进行了大量的研究工作。这些研究工作形成了一个新的学科，它具有自己的原理和方法。为了了解分布式数据库，我们必须在掌握传统的数据库和计算机网络的原理以外，能把这些知识与这种新技术的独特之处结合起来。

本书将介绍这一新技术的工作原理，其对象是对分布式数据处理感兴趣的计算机科学方面的大学生和教师，研究人员、系统管理员、系统或应用的设计者、分析师和程序人员。本书可用作数据库方面传统课程的扩充教材，也可作为关于分布式数据库课程的专门教材。

本书内容分为三大部分，即：基础、原理和系统举例。第一篇是分布式数据库的基础，其中第一章一般地介绍分布式处理系统的特点和要求。第二章介绍分布式数据库的特性和问题。第三章讨论本书其余部分所需的关于数据库和计算机网络方面的有关知识，主要是介绍一些概念、术语和基本原理。

第二篇是关于分布式数据库的原理，由第四章至第十一章组成，讨论分布式数据库的所有技术问题，因而是本书的主要部分。其中第四章讨论从应用程序员观点来看的分布数据库的体系结构，用例子来说明并定义了分布式数据库管理系统能够提供的不同层次上的可见度。第五章是关于分布式数据库的设计，它研究在计算机网络不同站点中数据的分割和分配问题。本章对设计者很有用，而且，了解数据为什么要分布和如何分布是了解分布式数据库性质的一个基本问题。上面这两章适合于自学，比较容易掌握。

第六、七章讨论访问分布式数据库的效率问题，这方面也可称为分布式查询的优化，但是这些技术不仅可专用于查询，而且也可用于设计重复执行的应用。第六章介绍如何把应用所需的数据库操作变换成相应于访问分布式数据库的语义上等效的操作。第七章处理更为专门的优化问题。相对来说，这两章比前面几章难懂一些。

第八、九、十章讨论事务的管理，也就是说，支持事务的高效、可靠和并发的执行。第八章介绍事务管理的基本方法，并且把这些方法结合在分布式数据库的设计和实现之中。第九章讨论分布式并发控制的方法，也即允许事务在不同站点平行地执行。第十章讨论分布式数据库从故障进行恢复的方法。第九、十两章和第六、七章一样较难懂一些，而且

也包含了较完整的综述。

第十一章处理数据的管理功能，它基本上是关于分布数据库的目录管理和安全性的问题。

第三篇介绍了一些分布式数据库管理系统的例子，重点是在把第二篇中描述的方法结合到具体的系统中去。第十二章说明了在当前商品化系统中如何来建立分布式数据库，以及可达到何种程度。第十三章介绍一个分布式数据库管理系统的原型 SDD-1，它代表了这个领域中的一个里程碑。第十四章介绍 R* 系统，这是分布式数据库的发展中正在进行的最重要的一项研究工作。第十五章综述了其他一些匀质研究原型所采用的方案。最后，第十六章介绍了在异质分布式数据库系统方面的几个主要的研究原型，这里所谓的"异质"，就是把不同的本地数据库管理系统集成起来。

本书的第四、五、六、及第十二章由涂小平同志编写，其余各章由王以和同志编写，并由王以和同志对全书进行了审校和文字统一的工作。

分布式数据库是一个较新的领域，涉及的知识面较广，目前尚在迅速地发展，有些术语及译名尚无统一的标准，加上编者经验有限，时间仓促，错误或疏漏之处，敬请读者指正。

编　者

一九八七年四月

目　　录

第一篇　分布式数据库系统基础

第一章　分布式处理系统

1.1 引言

分布式处理系统代表了数据处理领域中发展最快的一个分支，它迎合了大学和研究环境、商品化厂家以及军事方面各种用户的需要。大家都对"分布处理"系统提出了性能方面的种种要求，下面仅列出其中的一部分：

高系统性能、快速响应、高吞吐能力；

高可用性；

高可靠性；

降低网络费用；

故障软化能力、故障时降级使用；

易于模块化、逐步增长和结构灵活；

资源共享；

自动负载均分；

工作负载变化时的高适应性；

硬件和软件组成部分能逐步替换或升级；

易于扩充容量与功能；

易于适应新的功能要求；

对瞬时过载的良好响应，等等。

根据多处理机系统的当前水平是很难达到上述这些目标的。为此，分布式处理系统必须要有某些新的东西，从而开辟了广阔的研究领域。

本章将讨论仍处于研究阶段的这种新型系统的基本特性。

在一个系统中至少有四个物理组成部分才可以"分布"开来：硬件或逻辑线路，数据，处理本身以及控制（例如操作系统）。有一种说法认为只要有上述四种之一是分布的话，该系统就是分布式系统。但是，只根据系统的某些部分的物理分布来下定义是注定要失败的。正确的定义还必须包含这些分布部分互相作用的概念。存在物理分布而不认为是分布处理系统的例子如有的系统组成只是把输入/输出处理硬件和功能在物理上进行分布。同样，如果处理硬件没有分布则处理本身一定也没分布，并且反过来，把硬件分布而处理不分布也是难以想象的。

我们这里对分布处理系统的定义指的是能实现上面所列主要要求的"新"的系统，因而可以认为是一种"研究与发展"的定义。分布处理系统的这个定义有五个组成部分：

1. 通用资源部分（包括物理资源和逻辑资源这两方面）的重复性，可以在动态基础上把它们分配给具体的任务。物理资源是否匀质（homogeneity）在这里不是本质的。

2. 系统的这些物理部件和逻辑部件在物理上是分布的，并且通过通讯网络进行交互作用。

3. 有一高级的或高层的操作系统来对分布的各组成部分进行统一而综合的控制。各个处理机都有自己本地的操作系统，而且可以是各不相同的。

4. 系统透明性，即可以只用名字来请求种种服务，而不需要指出服务者。

5. 协作性自治 (cooperative autonomy)，表示了物理资源和逻辑资源的操作和交互作用的特征。

在许多系统中只是在不同程度上具有这些性质和工作特性，因而只能提供前面所列优点之中的一部分。只有把全部这些准则结合起来，才能够唯一地定义分布式处理系统。

1.2 重复性

在系统中，如果提供服务的可分配资源没有重复性的话，那么在整个系统中就不能有处理的分布。前面定义中"通用"这个词常常用来描述这些"可分配的"资源，例如通用计算机系统，或通用处理机等。系统要能在任何给定时间动态地重构那些资源以提供指定的服务。资源的这种重构或重分配必须能不影响不直接涉及的那些资源的工作。如果系统的目的是提供一种需要通用处理机的服务的话，那么该系统必须要有多个重复的通用处理机。如果系统的目的是满足某种别的功能(例如事务处理或在线控制)的话，那么该系统可以满足所有这五个准则，而可分配的硬件资源可以在广义上认为是专用的。但是，在该系统内它们是通用的和可分配的。一般情况下，系统同时具有通用和专用(固定功能)两种成分，或许专用成分中的某些部分是供专门使用的(不可分配的)。

一个极端情况是系统中每种资源只有一个，它不满足上述定义的第一项准则；而另一极端是每种资源都有重复，在这两个极端之间可以有许多的变型，这些结构中有的资源只有一个，而有些资源却有重复多个。因为在不完全重复的系统内，可自由地分配的多个重复资源的管理几乎与完全重复的情况一样困难，而且所提供的好处也几乎相同，所以下面的一般讨论同时包括了这两类系统。

为了达到高可用性、整个系统的可靠性和故障软化的目的，多个资源的可用性和有效地利用它们的能力是很重要的。这些特性也直接有助于灵活性、适应性和逐步扩充的能力。

1.3 物理分布与互相通讯

虽然可以有许多方法来定义一个网络，可以考虑它的许多不同的方面，但是在这里最重要的一个问题是信息的传送。两个物理资源之间协同工作的一个最好的例子就是通过网络进行信息的物理传送。这种传送遵从一种协议，其中通讯双方必须协作才能成功地完成传送。这种传送机构与门控式传送不同，后者主控者完全有权来迫使受控者物理上接受一报文。但在双方协作的协议中，目的地可以回送一指示信息（例如"未准备好"、"忙"、或"未确认")来物理上拒绝接受这个报文。信息传送的门控或主从控制方法会妨碍自治式工作(后面会讨论)，而所有资源之间的高度自治对于系统的高度可用性是很重要的。

为了支持进程的自治操作概念，必须把"网络"相互通讯的概念从物理资源推广到逻

辑资源的信息传送，例如逻辑资源之间的状态、服务请求和同步等。

双方协作协议的最重要一个特性就是它允许任何资源，不管是物理资源或逻辑资源，根据它对自己状态的了解来拒绝或接受一个报文的传送。网络的这个定义对于互连路径的长度没有任何限制，它们甚至可以是在单个集成电路芯片中两个资源之间的极短连线。这个定义的本质性特征是利用了一种双方协作的协议。

1.4 系统工作的统一性

为了把一分布处理系统中的各物理的和逻辑的组成部分集成为一功能性的整体，必须要实现一高层的操作系统。各个处理机可以有自己的操作系统，但是还要有一组完善定义的策略来管理整个系统的集成操作。实现这些策略的机构可以是与每个个别的操作系统相同的，也可以只是本地策略的逻辑扩充。在高层操作系统和本地操作系统之间必须不存在任何强烈的（strong）层次结构，因为那样会违反自治操作的准则。此外，本地操作系统不一定要是匀质的，虽然存在各种各样的接口肯定会使系统设计问题复杂化。

正如有人指出的那样，在一群独立工作的计算机和一群计算机集成成工作平稳而价格可取的分布式处理系统之间的主要差别是在系统软件上。

一个分布式系统具有多个控制地点和多个处理活动。为了满足我们定义中的全部五个准则，必须同时存在这些重复性以及控制地点的动态变化。此外，控制地点或处理活动在物理上的约束必须减至最小。对于每个处理机来说很重要的操作系统的核心必须减至最小，并且整个系统控制的地点在运行时必须是动态的。固定性约束的程度必须减至最小，并且系统必须没有任何"临界的路径"或"临界的部件"，例如对一资源的任何一个拷贝的固定性约束以包含全局的状态信息。在固定指派承担系统控制功能的一资源故障或过载时系统的运行性能不得受到严重影响或极大地降级。对于处理活动也有类似的要求，虽然有些情况下专用的资源可能要求固定束缚于一种处理功能（例如数据库或某一 I/O 设备的专门传送）。

这个高层操作系统必须具有的其他特征为：处理机之间的有效传送带宽比一般简单地通讯的计算机之间的带宽高；一处理机从该网络断开时仍能继续工作并有效地利用其本地的资源，并且能容易地重新连接和结合入系统而不影响正在进行的其他操作；能支持在"细微"级上的进程和处理机的交互作用（当需要时）。

这个高层操作系统不管是有一组独立的代码还是仅为一种设计思想都能把可用的资源从一简单的硬件集合转变为一能有条理地进行工作的系统。显然，高层操作系统的设计是整个系统设计中的关键，它提出了一系列的问题要求进行广泛的深入研究来解决它们。后面还会谈到其中的一些问题。

1.5 系统透明性

如果任一系统能够提供在引言中所列的能力的话，那它提供给用户的界面（或接口）必须是一种服务而不是服务者（服务器）。用户必须能用指定要做什么事的方法来请求一个动作，而不要求说明提供此服务的物理部件或逻辑部件。

分布式系统的一个重要特性是它的存在对用户是完全透明的，除非用户为了某些特殊的效率问题而自愿去了解它。用户应该能够象与单一个集中式系统进行通讯那样来开

发程序和处理数据库的操作。事实上，分布式系统的用户接口必须要甚至比当前的集中式系统的还要简单。这主要是因为用户是与高层操作系统进行通讯的；而这个控制软件的一个功能是处理系统中提供的所有各种命令语言和数据定义。非习质网络系统的初期工作已经强烈地推动更好的(甚至是标准化的)操作系统命令语言的研究。由于分布系统的组成对用户完全透明，所以系统的状态信息也都看不到，因而用户不可能知道当前哪些资源可用，或最好用哪个资源来执行此任务。因此，用户必须能用服务的名字而不用指出服务者来请求服务。如果有的应用以指定的资源来服务较好的话，则用户也应能请求这种形式的服务。

分布式系统用户接口的性质决定了这种系统最重要特性之一，即由一分布处理系统提供的服务也可以由一单处理机系统来提供，只要存在必要的硬件并可如此组织的话。但是，这样一种单处理机系统，不能够提供分布处理系统中象可靠性、适应性和模块化等的其他优点。

1.6 协作自治性

协作自治性（cooperative autonomy）是最后一个、或许是最重要的一个要素。分布处理系统必须设计得所有组成部分或资源(物理的或逻辑的)的操作都有极高程度的自治性。在物理级上，这可以利用网络的传输协议来实现，其中报文的传输需要发送者和接收者两方面的协调动作。在逻辑级上，在进程之间也必须存在同样程度的协作。再者，任何资源即使在它接受了该物理报文以后仍要能拒绝该服务的请求。这是由于在此系统内部控制是没有任何层次结构的。

但是，这不是无政府的混乱状况。所有的组成部分遵循一"总计划"（master plan），它反映在高层操作系统的思想之中。这种工作模式应称为协作自治性而不是简单的自治性。为了获得前面列出的许多好处，在所有组成部分之间的高度自治是极重要的，只有满足了定义中的全部五个准则的要求，在系统操作和组成部分交互作用方面才能具有这一特性。

1.7 某些排除在外的系统

在新的系统设计中已经进行了大量的工作来获得引言中所列的部分优点，但是极少有系统在满足所有这些准则方面得到较大的进展。

我们定义中的大多数准则是否满足可用在一特定方向上与一阈值相交来表示。这个定义不是一组二进制的判据，通过研究被此定义排除的一些系统能够对这些准则和它们的阈值有更好的了解。

例如，它排除了在一台主机内部的分布。某些现代处理机系统在结构上的特点是包含了独立的 I/O 通道，有人认为这些通道是协作性的分布式处理机，因为它包含了独立的 I/O 处理机、运算逻辑处理机和可能有的诊断处理机。这样一种分类方法很少用处而且也未被广泛接受。显然，在这类系统组织中各个组成部分固定地束缚于某些任务。

控制与主机进行通讯的前端处理机肯定不属于这里定义的分布式系统。虽然它可能满足了某些准则，但它也是专用于一种功能而且不能自由地进行分配。

在硬件和软件控制中有很多主从式的例子。关键之点是传送信息的接收者不能决定

它是否要接受这次传送并按此进行动作。在硬件控制中，这个概念叫做门控式传送（gated transfer）。在软件控制系统中，在多计算机操作系统和基本的多处理机操作系统中经常会遇到主从式关系的问题。但是，由于不是执行协作性协议，都不属于分布处理系统。

随着硬件价格的不断下降，人们对新的多处理机系统的兴趣日益增加，这种系统组织中包含了象向量乘法器、浮点运算部件、或快速富里哀变换部件等专用的功能部件。在操作的一般概念上，这种专门功能的处理与主从关系只稍有不同而已。主要差别在于主从控制关系也从我们的定义中排除了许多包含多个通用处理部件的硬件系统。对于这些结构造成术语上混淆的原因是这些专门的服务常常是由一通用部件（例如可编程微处理机）来提供的。这个功能部件可以用一微程序来使其"专门化"，或者它完全是通用的，不过在一专门的功能地位中使用而已，例如在一大型系统中使用一通用小型机来控制输入/输出。这一类被排除的系统的区别性特点是资源专用于某一个或某一组功能。只要在控制它自己的活动时，它就工作在主从模式，这就违反了自由分配和自治性这两条准则。

单台宿主处理机带有许多收集和发送数据的远程终端并不认为是一分布处理系统，即使这些终端是智能的或还可做某些编辑和格式化的工作。这个结论除了某些推销或广告人员以外是普遍公认的。

甚至在一复杂的网络互连结构中存在多个宿主机也不一定使得系统成为分布式的。从信息交换的观点看它可以是分布式的；但是从整个操作和控制的观点来看，它一般仍是集中式的。这些系统在硬件出故障时并没有动态重新分配任务的能力。

在广告中常常把智能终端系统叫做分布处理系统。但是，必须仔细地研究具有智能终端或本地处理机的系统的工作才能判断其处理在什么程度上真正是分布的。这种系统由接至一本地处理机的若干终端组成，该本地处理机还有磁盘、磁带等外部存贮能力。它提供智能式的数据录入——通过运行本地处理机中的程序来进行现场编辑和类似的功能。它能进行共享的文件访问，但只限于本地的文件。它可与一主处理机进行通讯，不过这时的本地处理机要模仿一"笨"的终端来使用一般的协议。最后，它还能进行远程作业录入。但是没有表明控制功能有任何的分布，因为工作的分布是固定的，本地的终端不能来影响它。

一个终端带有驻留的文本编辑器（不管它是由硬件还是由软件提供的）不是分布处理系统的例子。为了满足定义的要求，这个终端必须首先要能做某些真正的工作，其次要能知道自己不能完成所分配的任务时把它传递给另一合适的服务部件。当本级的利用比较充分而简单地把工作卸载给一更高级别的能力就正是过渡至完全分布处理的开始。如果这个终端能够不要人的干预自动地协调几个并发的远程作业，在不同地点给每个作业提供不同类型的服务的话，那么它就更接近于一分布系统。当本地的控制系统能够根据本地负载和能力的分析来决定某项工作应该在本地完成或者传递给系统中其他部分时就达到了这个准则的阈值。分布式处理肯定不只是相当于"把设备搬至一系统的周围以便在数据源处获得和处理数据。

智能终端在分布处理系统的发展中或许确实起了一定的作用。它能方便于"无痛苦"地过渡至对于硬件、数据存贮和控制来说更为分散的组织形式。其实现的方法就是在建立高层系统联系和完整的全局功能以前给本地系统增加一些特点和作某些改进以增加本地的功能。

1.8 表征分布的维数

从系统实现的观点来看，可以使用三维来表征一个系统的分散程度，也就可以用它来定义分布式系统。这三维分别代表硬件组成、控制点组织和数据库组织上的分散程度。

沿着硬件组成的方向可以指出若干点来，按增加分散度的次序，它们分别如下：

1. 具有一个控制器、一个运算器（ALU）和一个中央存贮器的单一中央处理机。

2. 多个执行部件：系统含有一个控制器、多个相同的运算器(例如 Illiac IV 中的处理单元)，以及或许还有多个独立的中央存贮器。

3. 独立的专用功能部件：系统含有一个通用控制器和多个 ALU 或处理部件，在处理部件中某些可以是专用部件，例如通道或 I/O 处理机、快速富里哀变换处理机、向量部件、或浮点运算部件。与这种专用功能部件相联系的可以有一些附加的控制部件，但是它们是固定的或能力有限的控制部件。与此同时，附加的 ALU 中有一些可以是相同的通用部件。

4. 多处理机：系统由多个控制器、多个 ALU (通用部件)以及或许有多个独立的中央存储器组成，但是只有一个合作的输入/输出系统。

5. 多计算机：每个计算机包含一通用的 CPU (控制器、ALU 和中央存储器)以及输入/输出系统。

除了最后两种组成以外，五个一般准则排除了其他所有组织形式归为分布式系统这一类的可能性。

与此相似，控制组织也是一个坐标，按照分散度增加的次序可分出下列各点：

1. 单个固定控制点，或者是物理的或者是概念上的。

2. 固定的主从关系：可以有多级的主从关系，而从属者之间的关系可以是非对称的。重要之点是这种主从关系是固定的，直到由完全是外部的动作来修改以前保持不变。

3. 动态的主从关系，可由软件来修改。

4. 多个(或者重复的)系统控制地点完全自治地工作。一个例子是两个独立的计算机仅在 I/O 级通过传送完整的文件来交互作用。

5. 在执行一任务时多个控制点协同工作，这个任务已被分成若干子任务。

6. 在执行一任务时重复的、相同的控制点协同工作。

7. 在执行一任务时多个控制点(不一定是匀质的)完全协同工作。

前四种排除在分布式处理之外。

沿着数据库轴上的各点的特征和次序就不那么简单，这些点的正确次序也不明显。我们在这里给出一种次序；但是，分布式数据库的各种组织的使用和维护目前尚未充分了解，本书的后面部分将在这方面进行较深入的探讨。

数据库本身有两个成份可被分布，即文件和对那些文件进行编目的目录。这两个之中哪一个都可被分布而不管另一个怎样，虽然某些组合很明显是无意义的或不实际的。除了分布的问题以外，还有一个文件与(或)目录的重复性问题。在我们图中轴上各点是已经真正实现的或已在实现上作了周密考虑的一些概念：

1. 集中式数据库，在文件及目录上只有单一拷贝放在外存储器中。

2. 同上，不过是放在主存中。

3. 分布式文件而有单一个集中式目录，且无本地目录。所有的访问都必须通过该单一个中央目录来处理。

4. 重复的数据库，在每个处理节点有一份全部文件的完整拷贝以及一份系统的目录。

5. 分片的数据库（partitioned data base），每个节点最直接需要的数据就放在该节点，而在其他节点则视需要存放其拷贝，但在主节点上则存放全部文件的一份完整的重复拷贝。在这种安排下，在一本地节点的任何事务处理发生例外情况时（即数据不在本地时）将由该主节点来进行处理。

6. 分片的数据库，本地数据（文件和目录）都保存在处理的节点，不在主节点重复存放。但是，主节点中保持一完整的目录，可以为引起本地例外情况的事务指引路由。

7. 分片的数据库，但没有任何主文件或主目录——对于一指定文件的位置来说没有任何单一的信息源。

上述头上三种数据库组织表现了一个公共的特性，即它们对于数据库的任何访问来说都只有单一个入口点——单一个集中式目录。显然，这种情况直接影响到系统的脆弱性（vulnerability）、可靠性和吞吐能力。

另一种表征数据库的方法是把数据的命名和存放作为不同的维。这也可能有几种组合，例如一公共名字空间（文件目录）和数据本身的分布式存储，与此相反的是在单一的共享存储设备上保持各完全分开的名字空间。

在前面给出的定义中，并考虑到这三维上排除的区域，就可以在一三维空间中表示出分布式系统所占的位置，如图 1-1 所示。

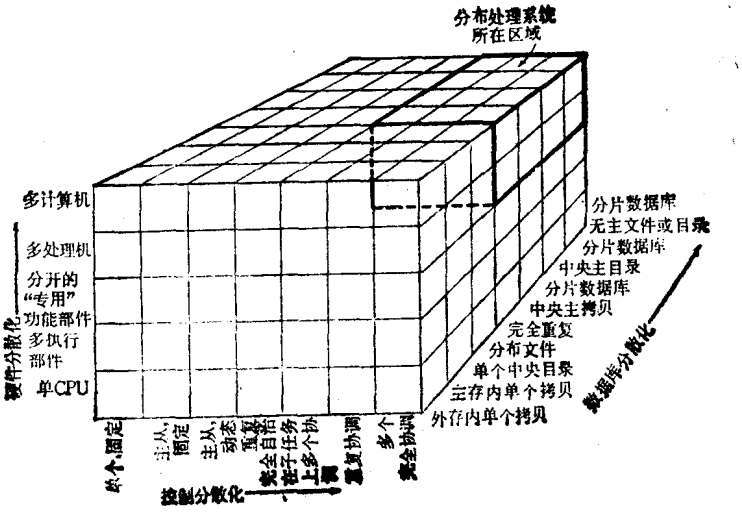


图 1.1　表征分布性的三维图

1.9　高层操作系统

高层操作系统是分布处理系统的一个关键因素，它的设计必须要考虑几个特性和问题。

操作系统的经典设计假定有大量系统的信息可用。虽然有关用户提出的工作的信息

是否完整和有效是有疑问的，但是一般假定操作系统能取到其工作环境的完整而准确的信息。但是，在分布处理系统中就不是这种情况；这时永远也得不到关于该系统的完整的信息。资源提供一个服务，但是它们可能有意无意地不让外界来检查其信息。

在分布系统中，在收集关于系统组成部分的状态信息时总有一定的时间延迟。这些时间延迟的不一致是极为重要的。在一般的集中式处理机中，操作系统可以请求状态信息，而保证被询问的部件不会改变状态而等待根据这些状态信息所作的决定，因为只有这一个询问的操作系统能发出命令。但在分布处理系统中，发生的时间滞后可能很显著；结果，由于自治的部件按自己的路径向前进行，所以可能发送出不准确的(过期的)信息。如果你曾使用过输入/输出设备处理程序的话，那你肯定常常要问所获得的信息是否准确。因此，对于分布处理系统来说，必须要设计成甚至在错误或不准确状态信息情况下也仍能工作。

关于系统信息有效性方面进一步复杂化的原因是提供给各系统控制器的信息可能不相同。这种差别可能是时间延迟和不同控制器信息屏蔽上的差别造成的。在时间和空间这两方面缺乏唯一性这一点具有非常重要的影响。

1.10 一般控制问题

这里所说的高层操作系统是高度非层次结构的，这就是说，它是单层的并且没有任何内部的主从关系。这一特性再加上部件的自治性，就大大恶化了控制问题。即使多个自治的部件正在协同工作，同时发生冲突的动作的可能性要比分层次系统中的高得多。而且，由于存在显著的时间延迟，对系统中各个控制器的动作进行同步要困难得多。最后，系统内死锁或无限循环的问题与其他系统的情况很不相同。有些建议请求一仲裁人(外界的第三者)来解决这个问题；但是，这种仲裁人必须是瞬时的，因为如果存在一固定的仲裁人的话就表示有了不可接受的分层控制。

从分布处理系统的工作特性来看，可以得出有关系统通讯的一些结论。我们定义的第二个准则要求在进程之间和处理机之间的通讯中，不论是物理的或逻辑的所有传送都要采用报文式协议。其中不得有任何全局变量，在系统部件之间也不得有任何直通"隧道"。所有参数都要通过完善定义而且严格执行的接口。

在单处理机和多处理机环境中通讯方面所做的大量工作在这里也可应用，但是要考虑到分布系统中系统部件的自治性质来加以扩充。

用户必须使用只包含服务名的命令来与系统进行通讯。我们关于系统透明性的准则使得用户指定一系统部件来提供所需的服务是不必要的或者不可能的。然而，这个要求引起了系统故障和用户差错检测方面的新问题，因为没有一个处理机能够确定所请求的服务是否可在系统中某个地方来提供，或者甚至只是确定这个服务是否合法也不可能。

分布处理系统中的资源管理是一多维的任务。因此至今对专用于分布系统的资源管理方面很少进行研究。但是，低层的功能与单处理机中执行的非常相似，例如物理资源的分配，以及在把进程安排在一具体系统部件上以后所需的各种设施的管理。然而，在能这样进行以前可能先要把所需的资源在一个地点汇集起来，或者建立起链接机构以便远程使用这些资源。在此过程中要涉及的问题是对资源进行定位，决定哪些部件是合适的，以及决定把资源移至所选地点的最佳方法。在较高层上是调度问题，即决定何时应启动或

终结一个功能。

采用单层的自治控制的任何系统在系统调度方面都会出现全新的问题。在非层次系统中对服务的请求可能一开始就被所有的物理资源拒绝。在那种情况下，此提出服务请求的实体可能要对此新的请求和当前正执行的任务进行相对优先级的评价。高效地执行这个过程是高层操作系统最重要的功能之一。

当把所有这些问题及解决办法与单处理机系统中遇到的类似问题及解决办法进行比较时，看来使分布系统控制问题复杂化的主要因素是分布处理系统内部的通讯，它相对于功能的详细执行来说是异步的，并且它在通讯处理本身的时间以外还表现出一定的时间延迟。单处理机使用信号灯、标志、加锁门或超时来解决许多问题。但要在相当复杂的分布系统中使用这些方法需要的时间太多，因而大大降低系统的吞吐能力。要记住，在一般网络中发送信号灯的传送时间大约在100毫秒数量级上。除了降低运行性能以外，单处理机中的大多数办法的可靠性和健壮性（robustness）也是有疑问的，因为象TEST AND SET（测试置位）那样的系统操作不能够象单一不可分割的机器指令那样在下一机器周期中立即执行。

时间问题进一步复杂化的原因在于，作为解决传送时间引起的困难的大多数办法（例如选举和软件同步）甚至要求系统中每个部件进行更多的处理。

小结

分布处理系统是一类新的组织和操作，它在所有各维中都表现出高度的分布性，并且在其整个操作和交互作用中具有高度的协作自治性。已经设计了大量的系统来满足分布处理系统定义中的一个或几个准则；但是在我们能够实现一真正的和完善的分布系统以前仍旧还有大量的课题需要加以研究解决。

参考文献

[1.1] P. H. Enslow, Jr, "What is a 'Distributed' Data Processing System?" COMPUTER, Jan. 1978, pp. 13—21.

[1.2] G. A. Champine, "Six Approaches to Distributed Data Bases", DATAMATION, May 1977, pp. 69—72.

[1.3] S. A. Kallis, Jr., "Networks and Distributed Processing", MINIMICRO SYSTEMS, March 1977, pp. 32—40.

[1.4] Arthur Lynch, "Distributed Processing Solves Mainframe Problems", DATA COMMUNICATIONS, Dec. 1976, pp. 17—22.

[1.5] H. Lorin, Aspects of Distributed Computer Systems, Wiley, 1981.

[1.6] G. M. Booth, The Distributed System Environment, McGraw-Hill, 1981.