



[12] 发 明 专 利 说 明 书

[21] ZL 专利号 99104521.1

[45] 授权公告日 2004 年 3 月 24 日

[11] 授权公告号 CN 1143228C

[22] 申请日 1999.3.25 [21] 申请号 99104521.1

[30] 优先权

[32] 1998.4.14 [33] GB [31] 9807683.9

[71] 专利权人 国际商业机器公司

地址 美国纽约州

[72] 发明人 A·D·J·班克斯

R·S·马克斯韦尔

D·N·米尔伍德

审查员 谭 毅

[74] 专利代理机构 中国专利代理(香港)有限公司

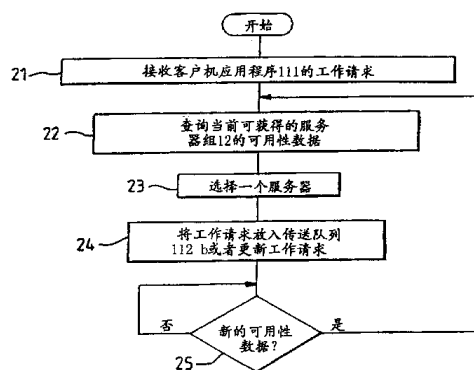
代理人 王 勇 张志醒

权利要求书 2 页 说明书 6 页 附图 1 页

[54] 发明名称 对服务器进行负荷管理的数据处理设备和方法

[57] 摘要

一种对异步客户机/服务器计算系统中的一组服务器进行工作负荷管理的数据处理设备,它具有:一个从客户机应用程序接收工作请求的装置;一个查询关于服务器组的可用性数据,并根据所查询可用性数据选择该组中特定服务器用于满足所接收工作请求的装置;及一个将所接收工作请求放入传送队列,等待向该组中该特定服务器传送的装置。



1. 一种用于对异步客户机/服务器计算系统中的一组服务器进行工作负荷管理的数据处理设备, 该设备的特征在于包含:

5 从客户机应用程序接收工作请求的装置;

查询关于服务器组的可用性数据, 并根据所查询的可用性数据选择该组中特定服务器用于满足所接收工作请求的装置, 其中, 为了从该组选择一个服务器, 总是使用在工作请求被接收时的当前状态的所查询的可用性数据, 而不管是否已经收到该组中每个服务器的可用性数据; 以

10 及

将所接收的工作请求放入传送队列, 等待向该组中该特定服务器传送的装置。

2. 权利要求 1 的设备, 进一步包含:

判定在收到工作请求后可用性数据是否已经改变的装置;

15 在可用性数据发生改变时查询改变后的可用性数据、并根据改变后的可用性数据选择该组中某个服务器用于满足所接收的工作请求的装置; 以及

更新传送队列中工作请求的装置, 使得如果根据改变后的可用性数据选择的服务器与特定的服务器不同, 则将工作请求发送到该组中与所述特定的服务器不同的服务器。

20 3. 权利要求 2 的设备, 其中, 改变后的可用性数据包括向该组最初收到工作请求时没有可用性数据的服务器添加的可用性数据。

4. 权利要求 2 的设备, 其中, 改变后的可用性数据包括对该组最初收到工作请求时有可用性数据的服务器修改的可用性数据。

25 5. 一种用于对异步客户机/服务器计算系统中的一组服务器进行工作负荷管理的数据处理方法, 该方法的特征在于包含下述步骤:

从客户机应用程序接收工作请求;

30 查询关于服务器组的可用性数据, 并根据所查询的可用性数据选择该组中特定服务器用于满足所接收工作请求, 其中, 为了从该组选择一个服务器, 总是使用在工作请求被接收时的当前状态的所查询的可用性数据, 而不管是否已经收到该组中每个服务器的可用性数据; 以及

将所接收的工作请求放入传送队列，等待向该组中该特定服务器传送。

6. 权利要求 5 的方法，进一步包含下述步骤：

判定在收到工作请求后可用性数据是否改变；

- 5 在可用性数据发生改变时查询改变后的可用性数据、并根据改变后的可用性数据选择该组中某个服务器用于满足所接收的工作请求；以及更新传送队列中工作请求，使得如果根据改变后可用性数据选择的服务器与特定的服务器不同，则将工作请求发送到该组中与所述特定服务器不同的服务器。

- 10 7. 权利要求 6 的方法，其中，改变后的可用性数据包括向该组最初收到工作请求时没有可用性数据的服务器添加的可用性数据。

8. 权利要求 6 的方法，其中，改变后可用性数据包括对该组最初收到工作请求时有可用性数据的服务器修改的可用性数据。

对服务器进行负荷管理的 数据处理设备和方法

- 5 本发明涉及的领域是计算机系统,更具体地说是客户机/服务器(亦称“分布式”)计算系统,其中的一个计算设备(“客户机”)请求另一个计算设备(“服务器”)执行该客户机的部分工作。

过去几年中,客户机/服务器计算在信息技术世界已经变得越来越重要。这种分布式计算允许一台计算机将其某些工作委托给另一台例如可能更适合执行该工作的计算机去执行。例如,服务器可能是一台高性能计算机,运行着管理大量数据的存储的数据库程序,而客户机只是一台从数据库请求数据以用于其一个本地程序的桌面计算机(PC)。

10 报文排队(亦称报文代理)数据处理技术在当今的客户机/服务器计算机网络中已经越来越流行。这项技术允许客户机计算机系统与服务器计算机系统进行通讯,即使这两个系统之间在操作系统、数据格式和通讯协议方面存在很大差别。此外,由于这项技术的异步性质,在客户机能向服务器发送报文,服务器可以将报文存储在一个队列中,在晚些时候对报文作出处理与响应。这与要求客户机和服务器实时对话的同步客户机/服务器模型(例如,客户机要等待服务器响应后才能执行其它任务)有相当的不同。

20 关于报文排队和市场上的报文排队产品的介绍文章有《使用 MQI 的报文发送和排队》(“Messaging and Queuing Using the MQI”,作者 B.Blakeley、H.Harris&R.Lewis、McGraw-Hill,1994 年),以及 IBM 公司的下列出版物:《报文发送和排队介绍》(“An Introduction to Messaging and Queuing”,IBM 文件号 GC33-0805-00)和《MQSeries—报文排队接口技术参考》(“MQSeries-Message Queue Interface Technical Reference”,IBM 文件号 SG33-0850-01)。IBM 和 MQSeries 是 IBM 公司的商标。IBM 公司的 MQSeries 报文发送软件产品按照一个报文发送协议提供事务处理报文发送支持、逻辑工作单元内的同步报文,该协议保证,即使发生系统或通讯故障,也只进行一次并且仅仅一次报文投递。MQSeries 产品提供投递保障的方法是,在确定接收系统安全地存储了报文之前,不将发送系统中存储的该报文最终删除,并且还使用

复杂的恢复手段。在确认存储成功时履行报文传输之前，从发送系统的存储器中删除该报文和将该报文插入接收系统的存储器这两种操作都处于疑问状态，并能在发生故障时原子地放弃。对这种报文传送协议和相关的事务处理概念及恢复手段的描述，见于国际专利申请 WO 95/10805 5 和美国专利 5465328，它们被引用为本文的参考文献。

在异步客户机/服务器网络例如使用报文和排队软件体系结构的网络中，越来越普遍地提供一组服务器，其中组中的每个服务器都能满足客户机的请求。这样，客户机的请求就能得到更迅速的服务，这是因为，当某个服务器忙时，请求能被传递到另一个服务器。一般来说，位于客
10 户机侧的工作负荷管理单元负责决定应当指定哪个服务器来处理特定请求，它采用每个服务器的可用性数据作为决定的依据。如果所有这种可用性数据一时还得不到，工作负荷管理单元就等待，直到能收到这种数据，然后用这组完整数据来决定选择哪个服务器。然后，客户机请求与附加的工作负荷管理单元选择的服务器的指示，被放置到一个传输队
15 列，等待传输到选定服务器。客户机也将有关该排队请求的某些信息、包括选定服务器的指示信息，写入本地存储器，以便如果发生传输问题，服务器未收到传送的客户机请求时，该请求能再次从客户机向服务器发送。

这个领域当前的技术状态的缺点在于工作负荷管理单元必须用复杂的软件部分来连续地检查每个服务器的服务器可用性数据的状态，以判
20 定什么时候完整地服务器收到所有这种数据。这种复杂的软件编程大大地增加了总体系统的附加费用。

此外，如果尚未收到所有的服务器可用性数据，就不能向服务器组传送请求，这样就在客户机侧暂停数据传送过程。这一点非常不好，在
25 快速传送速度非常重要的环境中尤为如此。

本发明的第一个方面是，提供一种对异步客户机/服务器计算系统中的一组服务器进行工作负荷管理的数据处理设备，它具有：一个从客户机应用程序接收工作请求的装置；一个查询关于服务器组的可用性数据，并根据所查询的可用性数据选择该组中特定的服务器用于满足所接
30 收的工作请求的装置，其中，为了从该组选择一个服务器，总是使用在工作请求被接收时的当前状态的所查询的可用性数据，而不管是否

已经收到该组中每个服务器的可用性数据；以及一个将所接收的工作请求放入传送队列，等待向该组中该特定服务器传送的装置。

- 5 该设备最好还具有：一种判定在收到工作请求后可用性数据是否改变的装置；一种在可用性数据发生改变时查询改变后的可用性数据、并根据改变后的可用性数据选择该组中某个服务器用于满足所接收工作请求的装置；及一种更新传送队列中工作请求的装置，使得如果根据改变后的可用性数据选择的服务器与特定的服务器不同，则将工作请求发送到与该特定的服务器不同的该组中的服务器。

- 10 进一步最好是，改变后的可用性数据包括向该组中最初收到工作请求时没有可用性数据的服务器添加的可用性数据。另一种可选方案是，改变后的可用性数据包括对该组中最初收到工作请求时有可用性数据的服务器修改的可用性数据。

本发明的第二个方面是，提供一种执行第一个方面所述工作负荷管理设备的功能的方法。

- 15 本发明的第三个方面是，提供一种计算机程序产品，它存储在计算机可读存储介质中，当在计算机上运行时，执行第一个方面的功能。

- 20 这样，本发明能使工作负荷管理单元软件大大简化，这是因为没有必要连续检查服务器可用性数据是否全部都已收到。相反，可以使用一种简单的软件结构，其中的所有客户机请求都按同样方式对待，即一旦某客户机发出请求，就用服务器可用性数据的当前状态来作出决定，将请求发送到传送队列。并且，作为记录客户机请求本身的同一操作的一部分也将目的地记录在持久存储器中。

- 25 倘若在请求从传送队列发往选定服务器之前，有了新的可用性数据，则可以重新选择服务器，将排队的请求更新以发往不同的服务器。然而，如果在请求从传送队列发往选定服务器之前，没有新的可用性数据，则排队的请求被发往选定的服务器，而不管是否组中的每个服务器的可用性数据都有了。这样就增加了一个优点，即不会因为工作负荷管理单元不得不等待每个服务器的可用性数据都有而产生的传送延迟。本发明用在刚刚收到工作请求时的当时所能得到的可用性数据，对应该选择组中的哪个服务器作出初始决定。只有在不同寻常的情况下才会出现新的可用性数据，从而要求建立新的目的地记录。
- 30

现在结合以下附图说明本发明的一个实施例：

图 1 是表示本发明优选实施例的操作所在的基本客户机/服务器环境的框图;

图 2 是表示按照本发明的优选实施例, 报文发送与排队单元所采取步骤的流程图;

5 图 3 是表示按照本发明的优选实施例, 报文发送与排队单元的工作负荷管理单元所查询的可用性数据的表。

在图 1 中, 客户机数据处理设备 11 与一组 12 服务器数据处理设备 (121、122、123) 在通讯网络 13 上通讯。客户机 11 和服务器组 12 运行报文发送与排队软件产品例如 IBM 的 MQSeries, 这样, 客户机和服务器之间进行的是异步通讯。就是说, 当客户机向服务器发送工作请求时, 该服务器不必是可操作的。请求将被放入队列, 服务器可以晚些时候在服务器可操作时检索排队的请求。报文发送与排队产品也执行必要的通讯协议转换, 以便允许客户机 11 与服务器组 12 在网络 13 上对话。

15 尽管在优选实施例中, 客户机和服务器位于由网络 13 连接的不同处理单元中, 本发明绝非限于此种结构。客户机和服务器组可以位于相同的数据处理单元。

客户机 11 含有客户机应用程序 111 和报文发送与排队软件系统 112, 后者包括工作负荷管理单元 112a 和传送队列 112b。客户机应用程序 111 是向服务器发出工作请求的消费软件部分。一旦服务器执行了所请求的工作, 就会将执行结果返回给客户机应用程序 111。传送队列 112b 是报文发送与排队软件结构的一部分, 用于在将工作请求通过网络 13 发送到服务器组 12 之前, 对从客户机应用程序 111 收到的工作请求进行排队。工作负荷管理单元 112a 也是报文发送与排队软件的一部分, 它是一个软件成分, 负责为每个从客户机应用程序 111 接收的工作请求, 从服务器组 12 中的服务器 121、122、123 中选择一个服务器。

25 工作负荷管理单元 112a 通过网络 13 接收关于服务器组中每个服务器 121、122、123 的数据 (以下称作“可用性数据”)。这个数据表示各服务器当前的可用性, 即各服务器对从客户机接收新的工作请求的准备程度。工作负荷管理单元用这个可用性数据来决定应选择服务器 121、122、123 的哪一个来满足特定的工作请求。

30 下面结合图 2 的流程图来说明报文发送与排队系统 112 的操作。

步骤 21, 报文发送与排队系统 112 接收客户机应用程序 111 的一个工作请求。步骤 22, 工作负荷管理单元 112a 查询数据可用性表 (见图 3), 以便确定组 12 中服务器对处理所接收工作请求的可用性。

图 3 中, 工作请求刚刚被接收的时间点称为时间 1。在时间 1, 可用性表表示服务器 1 (121) 可用于接收工作请求 (如在构成时间 1 与服务器 1 的交点的框中代表“可用”的字母 A 所指示的)。在时间 1 上, 可用性表指示服务器 2 (122) 不可用于接收工作请求 (如在构成时间 1 与服务器 2 的交点的框中代表“不可用”的字母 NA 所指示的)。在时间 1 上, 注意在图 3 的表中对服务器 3 没有项目, 从而表示服务器 3 尚未向客户机 11 的报文发送与排队软件 112 提供其可用性数据。

假设时间 1 上这一可用性数据的状态, 工作负荷管理单元 112a 选择 (步骤 23) 服务器 1 (121) 分配给这一工作请求, 由于服务器 1 可用, 服务器 2 不可用及服务器 3 是否可用未知。该工作请求然后被放入 (步骤 24) 传送队列 112b, 并向工作请求加入一个控制指示, 指示该工作请求将被发往服务器 1 (121)。工作请求于是在队列 112b 中等待, 直到服务器 1 (121) 指出其已经准备好接收请求 (服务器 1 可能与客户机处于不同的时间区, 在时间 1 是不可操作的)。

在稍晚于时间 1 的时间 2, 客户机 11 已经接收到更新的服务器可用性数据, 这种接收是报文发送与排队软件 112 在步骤 25 判定的。于是, 控制循环返回步骤 22, 工作负荷管理单元 112a 再次查询可用性数据, 判定现在的情况: 服务器 1 仍然可用 (在时间 2 所在行与服务器 1 所在列交叉的方框中的字母 A 表示“可用”); 服务器 2 现在变得可用 (在时间 2 所在行与服务器 2 所在列交叉的方框中的字母 A 表示“可用”); 服务器 3 不可用 (在时间 2 所在行与服务器 3 所在列交叉的方框中的字母 NA 表示“不可用”)。于是, 在时间 1 到时间 2 期间内, 从服务器组 12 向客户机 11 发送了新的可用性数据, 服务器 2 的可用性发生变化, 服务器 3 提供了可用性数据, 而在这之前它并没有提供这种数据。

于是, 在步骤 23, 工作负荷管理单元 112a 用这个更新的可用性数据来从三个服务器中选择一个服务器。尽管可以使用任何工作负荷平衡方案, 但是最为有效的方案是, 如果有两个服务器均可用, 而其中的一个已经在上一遍通过流程中被选定, 则保留对该服务器的选择, 以免为修改在队列 112b 中等待的工作请求而付出处理能力。这也省去了向本

地存储器的数据写操作，因为为每个排队工作请求选定的服务器都必须
在本地存储器记录，以便万一发生传送故障时作恢复用。于是在这种情
况下，步骤 24 的结果是不对该排队工作请求进行更新。

然后在晚于时间 2 的时间 3，工作请求仍然在队列 112b 中等待，接
收到新的服务器可用性数据，这种数据的接收是由报文发送与排队软件
112 在步骤 25 通知的。于是，控制循环返回步骤 22，工作负荷管理单
元 112a 再次查询图 3 的可用性数据表，以确定是否需要根据新的可用
性数据来改变选定的服务器。如图 3 所示，服务器 1 现在不可用，服务
器 2 仍然可用，服务器 3 现在可用。于是，在步骤 23，工作负荷管理单
元 112a 选择服务器 2 或服务器 3。同样可以使用任何常用的工作负荷平
衡算法来决定应选择这两个服务器中的哪一个。按照循环调度法，选择
服务器 2，因为它在队列中排在服务器 1 之后（在最后一轮中，服务器 3
将是队列中下一个被选择对象）。

然后在步骤 24，工作请求在传送队列 112b 中被更新，使得该请求
现在包含的目的地指示指向服务器 2 而不是服务器 1。这也涉及必须执
行一次对本地存储器的写操作，以提供一个必要时用于数据恢复的记录。

在这个操作期间的某个时刻，选定服务器将从队列 112b 提取工作
请求，该请求将经网络 13 被传送到服务器 2，这样就结束了图 2 中对某
特定工作请求的循环流程。

应当注意，本发明的操作所基于的基本假设是，服务器可用性数据
的变化不是很频繁，并且很少发生需要在步骤 25 选择 YES 分支、循环
返回步骤 22 再次查询图 3 的表的情况。这样，在客户机应用程序发出
请求后，很早就能对服务器作出初始选择。这样就能尽快地将工作请求
放入传送队列 112b，与此同时利用当前能获得的服务器可用性数据，使
得请求能为向服务器组传送而得到迅速有效的处理。只有当服务器可用
性数据改变时，才有必要修改对服务器的初始选择。

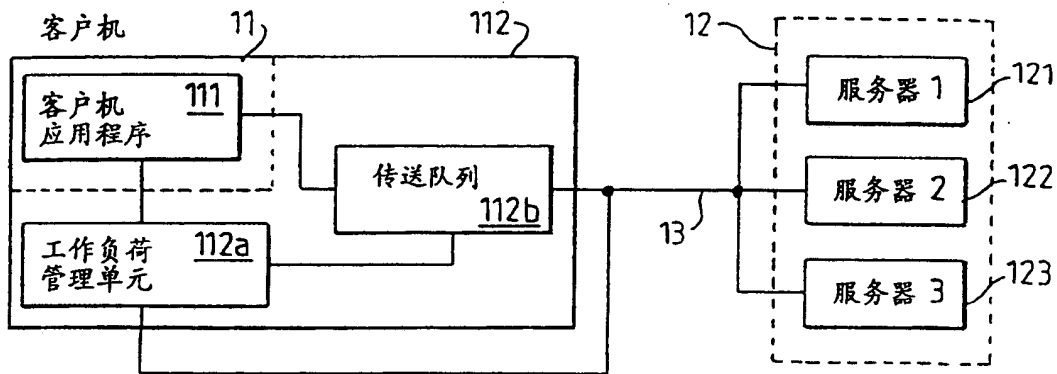


图 1

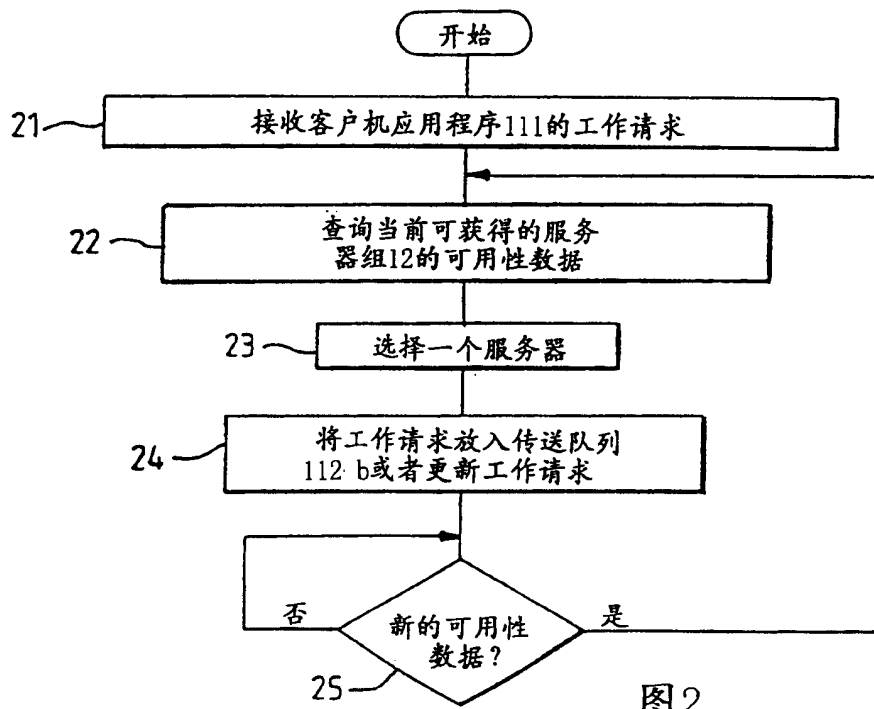


图 2

时间 \ 服务器	服务器 1	服务器 2	服务器 3
	服务器 1	服务器 2	服务器 3
1	A	NA	
2	A	A	NA
3	NA	A	A

图 3