

(12) 发明专利

(10) 授权公告号 CN 101324877 B

(45) 授权公告日 2012. 01. 04

(21) 申请号 200810125595. 8

(22) 申请日 2008. 06. 13

(30) 优先权数据

11/763, 411 2007. 06. 14 US

(73) 专利权人 国际商业机器公司

地址 美国纽约

(72) 发明人 小威廉·G·沃多恩

安德鲁·D·瓦尔斯

(74) 专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

代理人 李向英

(51) Int. Cl.

G06F 15/163(2006. 01)

H04L 29/08(2006. 01)

(56) 对比文件

US 6725307 B1, 2004. 04. 20, 第3栏第15-17行, 第8栏第19行-第9栏第15行、附图4.

US 7065599 B2, 2006. 06. 20, 第1栏第39-45行, 第28栏第40-44行, 第20栏第45-60行、附图1, 11, 14.

审查员 杜军

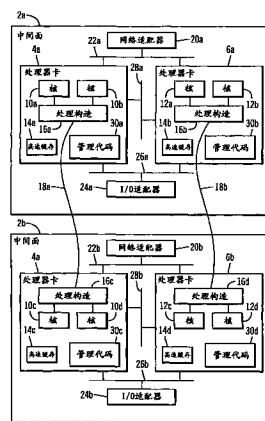
权利要求书 3 页 说明书 8 页 附图 5 页

(54) 发明名称

通过处理器构造连接的处理器卡的多节点配置的系统 and 制造方法

(57) 摘要

提供一种系统, 该系统包括第一节点、第二节点和多个通信接口, 所述第一节点包括多个处理器卡, 每一个所述处理器卡包括处理器构造, 其中所述第一节点中的所述处理器卡通过它们的处理器构造连接, 所述第二节点包括多个处理器卡, 每一个所述处理器卡包括处理器构造, 其中所述第二节点中的所述处理器卡通过它们的处理器构造连接, 其中所述多个通信接口中的每一个接口连接所述第二节点中的一个处理器卡和所述第一节点的一个处理器卡, 使所述连接的处理器卡之间能够进行通信, 从而协调所述第一节点和所述第二节点中的所述连接的处理器卡之间的处理器操作。



1. 一种通过处理器构造连接的处理器卡的多节点配置的系统,包括:

第一节点,所述第一节点包括多个第一处理器卡,每一个所述第一处理器卡包括处理器构造,其中所述第一处理器卡通过它们的处理器构造连接,

第二节点,所述第二节点包括多个第二处理器卡,每一个所述第二处理器卡包括处理器构造,其中所述第二处理器卡通过它们的处理器构造连接,

多个中间面,所述多个中间面中的每一个包括所述多个第一处理器卡之一和所述多个第二处理器卡之一,以及

多个通信接口,其中所述多个通信接口中的每一个接口连接包括在同一个中间面中的所述第二处理器卡和所述第一处理器卡,使所述连接的处理器卡之间能够进行通信,从而协调所述第一节点和所述第二节点中的所述连接的处理器卡之间的处理器操作。

2. 根据权利要求 1 的系统,其中通信接口和处理器构造使用不同的技术,其中利用它们的处理器构造进行通信的所述第一节点和所述第二节点的每一个节点中的所述处理器卡作为第一和第二对称多处理器。

3. 根据权利要求 2 的系统,其中不同节点中的所述处理器卡不利用所述处理器构造技术彼此通信。

4. 根据权利要求 1 的系统,其中每一个中间面都包括在存储服务器中。

5. 根据权利要求 1 的系统,其中每一个中间面包括能够与外部设备进行通信的输入/输出 I/O 适配器,其中使用不同于所述处理器构造接口的通信接口,将每一个中间面中的每一个处理器卡连接到所述中间面上的所述 I/O 适配器上。

6. 根据权利要求 1 的系统,还包括:

电缆,其中所述电缆连接所述第一处理器卡的所述处理器构造,并连接所述第二处理器卡的所述处理器构造。

7. 根据权利要求 1 的系统,

其中,所述第一节点和所述第二节点中每一个上的至少一个处理器卡包括:

用于检测所述第一节点或第二节点的其中一个上的处理器卡的故障的部件;

用于确定包括所述故障处理器卡的节点是否包括至少一个运行的处理器卡的部件;

用于响应于确定所述节点包括至少一个运行的处理器卡,对包括所述故障处理器卡的所述第一节点或所述第二节点进行重新配置,以在没有所述故障处理器卡的情况下进行操作的部件;

用于响应于确定所述节点不包括至少一个运行的处理器卡执行切换操作,以使用不包括所述故障处理器卡的所述第一节点或所述第二节点的部件。

8. 根据权利要求 1 的系统,

其中,所述第一节点和所述第二节点中每一个上的至少一个处理器卡包括:

用于检测所述第一节点或所述第二节点其中一个上的一个处理器卡的故障的部件;

用于执行切换操作,以使用不包括所述故障处理器卡的所述第一节点或所述第二节点的部件。

9. 一种通过处理器构造连接的处理器卡的多节点配置的系统,包括:

第一节点,包括第一处理器卡和第三处理器卡,其中每一个处理器卡包括处理器构造,并且所述第一处理器卡和所述第三处理器卡通过它们的处理器构造连接;

第二节点,包括第二处理器卡和第四处理器卡,其中每一个处理器卡包括处理器构造,并且其中所述第二处理器卡和所述第四处理器卡通过它们的处理器构造连接;

第一中间面,其包含:

所述第一处理器卡和第二处理器卡;

第一通信接口,所述通信接口连接所述第一处理器卡和所述第二处理器卡,使所述第一处理器卡和所述第二处理器卡之间能够进行通信;以及

第二中间面,包括:

所述第三处理器卡和第四处理器卡;

第二通信接口,连接所述第三处理器卡和所述第四处理器卡,使所述第三处理器卡和第四处理器卡之间能够进行通信。

10. 根据权利要求 9 的系统,其中通信接口和处理器构造使用不同的技术,其中所述连接的第一处理器卡和第三处理器卡以及所述连接的所述第二处理器卡和第四处理器卡作为第一和第二对称多处理器。

11. 根据权利要求 9 的系统,其中每一个中间面都包括在存储服务器中。

12. 根据权利要求 11 的系统,其中每一个中间面包括能够与外部设备进行通信的输入/输出 I/O 适配器,其中使用不同于所述处理器构造的通信接口,将每一个中间面中的每一个处理器卡连接到所述中间面上的所述 I/O 适配器上。

13. 根据权利要求 9 的系统,还包括:

电缆,其中所述电缆连接所述第一和第三处理器卡的所述处理器构造,以及所述第二和第四处理器卡的所述处理器构造。

14. 根据权利要求 9 的系统,

其中,至少一个处理器卡包括:

用于检测一个处理器卡的故障的部件;

用于确定所述故障处理器卡是否通过所述处理器构造连接到一个运行的处理器卡的部件;

用于响应于确定所述故障处理器卡通过所述处理器构造连接到所述运行的处理器卡,重新配置通过它们的处理器构造连接的所述故障处理器卡和运行的处理器卡,以在没有所述故障处理器卡的情况下进行操作的部件;以及

用于响应于确定所述故障处理器卡没有通过所述处理器构造连接到所述运行的处理器卡,执行切换操作,以使用通过它们的处理器构造连接的、没有通过处理器构造连接到所述故障处理器卡的处理器卡的部件。

15. 根据权利要求 9 的系统,

其中,至少一个处理器卡包括:

用于检测一个处理器卡的故障的部件;

用于执行切换操作,以使用通过它们的处理器构造连接的、没有通过处理器构造连接到所述故障处理器卡的处理器卡的部件。

16. 一种通过处理器构造连接的处理器卡的多节点配置的方法,包括:

配置包括多个第一处理器卡的第一节点,每一个所述第一处理器卡包括处理器构造,所述第一处理器卡通过它们的处理器构造连接,

配置包括多个第二处理器卡的第二节点,每一个所述第二处理器卡包括处理器构造,所述第二处理器卡通过它们的处理器构造连接,

配置多个中间面,每一个中间面包括所述多个第一处理器卡之一和所述多个第二处理器卡之一,

配置多个通信接口,其中所述多个通信接口中的每一个接口连接包括在同一个中间面中的所述第二处理器卡和所述第一处理器卡,使所述连接的处理器卡之间能够进行通信,从而协调所述第一节点和所述第二节点中的所述连接的处理器卡之间的处理器操作。

17. 根据权利要求 16 的方法,其中所述通信接口和所述处理器构造使用不同的技术,其中对所述第一节点和所述第二节点以及所述处理器构造进行配置包括配置所述第一节点和所述第二节点,以分别作为第一和第二对称多处理器。

18. 根据权利要求 16 的方法,还包括:

检测所述第一节点或第二节点的其中一个上的一个处理器卡的故障;

确定包括所述故障处理器卡的所述节点是否至少包括一个运行的处理器卡;

响应于确定所述节点至少包括一个运行的处理器卡,重新配置包括所述故障处理器卡的所述第一节点或所述第二节点,以在没有所述故障处理器卡的情况下进行操作;以及

响应于确定所述节点不包括至少一个运行的处理器卡,执行切换以使用不包括所述故障处理器卡的所述第一节点或所述第二节点。

19. 根据权利要求 16 的方法,还包括:

检测所述第一节点或第二节点其中一个中的一个处理器卡的故障;

执行切换以使用不包括所述故障处理器卡的所述第一节点或所述第二节点。

## 通过处理器构造连接的处理器卡的多节点配置的系统 and 制造方法

### 技术领域

[0001] 本发明涉及通过处理器构造 (processor fabric) 连接的多节点配置的处理器卡的制造方法、系统和产品。

### 背景技术

[0002] 存储网络中的主机系统可以通过多条路径与存储器控制器通信。存储器控制器可能由多个独立的存储群集组成, 每一个存储群集都能够访问存储器, 并提供访问存储器的冗余。如果存储群集故障, 主机还可以切换至其他存储群集来访问存储器。

[0003] 在某些系统中, 多个双机群集存储服务器可以管理对一个存储系统的 I/O 访问。在操作过程中, 每一个存储服务器被分配一个不同范围的存储器, 例如不同的逻辑单元号 (LUN)。如果一个存储服务器或存储群集接收了请求, 那么存储服务器或群集将在该目标 LUN 被分配给接收存储服务器的情况下, 对该请求进行处理。该存储服务器将把该 I/O 请求转发给管理对该目标 LUN 的访问的另一个存储服务器或群集。这种设置要求每一个存储服务器 / 群集都知道每一个其他存储服务器 / 群集所管理的 LUN。如果一个存储服务器出现故障, 那么作为切换操作的一部分, 该存储服务器管理的节点被分配给正常工作的存储服务器。如果有存储服务器恢复, 那么可以进行倒切换 (failback), 以便给恢复的存储服务器重新分配一些 LUN。

[0004] 存储服务器可以封装电子元件以容纳多个处理器卡, 包括总线接口和其他电子元件以支持在系统中实现不同节点的多个处理器卡。这种结构要求服务器包括多个冗余的元件, 例如插槽、总线接口等。在很多情况下, 如果在设有存储服务器的环境中没有包括或要求允许的最大数目的处理器卡, 未必所有的服务器电子元件, 如插槽、总线接口等都被使用。

[0005] 在本技术领域存在着改进在存储系统中配置多个节点的技术的需要。

[0006] 发明内容

[0007] 提供一种通过处理器构造连接的处理器卡的多节点配置的系统, 该系统包括第一节点、第二节点和多个通信接口, 所述第一节点包括多个处理器卡, 每一个所述处理器卡包括处理器构造, 其中所述第一节点中的所述处理器卡通过它们的处理器构造连接, 所述第二节点包括多个处理器卡, 每一个所述处理器卡包括处理器构造, 其中所述第二节点中的所述处理器卡通过它们的处理器构造连接, 其中所述多个通信接口中的每一个接口连接所述第二节点中的一个处理器卡和所述第一节点中的一个处理器卡, 使所述连接的处理器卡之间能够进行通信, 从而协调所述第一节点和所述第二节点中的所述连接的处理器卡之间的处理器操作。

[0008] 在另一个实施例中, 所述通信接口和处理器构造使用不同的技术, 其中利用所述第一处理器构造和第二处理器构造进行通信的所述第一节点和所述第二节点中的所述处理器卡, 作为第一和第二对称多处理器。

[0009] 在又一个实施例中,不同节点中的所述处理器卡不利用所述处理器构造技术彼此通信。

[0010] 在又一个实施例中,存在多个中间面 (midplane),其中每一个中间面包括至少两个处理器卡,并且其中所述第一节点和所述第二节点的每一个至少包括一个来自每一个中间面的处理器卡。每一个处理器卡包括处理器构造接口,以连接包括所述处理器卡的所述第一节点或所述第二节点中的其他处理器卡。

[0011] 在又一个实施例中,每一个中间面都包括在存储服务器中。

[0012] 在又一个实施例中,每一个中间面包括能够与外部设备进行通信的输入 / 输出 I/O 适配器,其中使用不同于所述处理器构造接口的通信接口,将每一个中间面中的每一个处理器卡连接到所述中间面上的所述 I/O 适配器上。

[0013] 在又一个实施例中,电缆连接所述第一节点上的处理器卡,并连接所述第二节点上的处理器卡。

[0014] 在又一个实施例中,所述第一节点和所述第二节点中每一个上至少一个处理器卡包括:用于检测所述第一节点或所述第二节点的其中一个上的一个处理器卡的故障的部件;用于确定包括所述故障处理器卡的节点是否包括至少一个运行的处理器卡的部件;用于响应于确定所述节点包括至少一个运行的处理器卡,重新配置包括所述故障处理器卡的所述第一节点或所述第二节点,以在没有所述故障处理器卡的情况下进行操作的部件;用于响应于确定所述节点不包括至少一个运行的处理器卡执行切换操作,以使用不包括所述故障处理器卡的所述第一节点或所述第二节点的部件。

[0015] 在又一个实施例中,所述第一节点和所述第二节点中每一个上的至少一个处理器卡包括:用于检测所述第一节点或所述第二节点其中一个上的一个处理器卡的故障的部件;用于执行切换操作,以使用不包括所述故障处理器卡的所述第一节点或所述第二节点的部件。

[0016] 在又一个实施例中,一种通过处理器构造连接的处理器卡的多节点配置的系统包括第一中间面和第二中间面,第一中间面包含第一处理器卡和第二处理器卡和第一通信接口,其中每一个处理器卡包括处理器构造,所述第一通信接口连接所述第一处理器卡和所述第二处理器卡,使所述第一处理器卡和所述第二处理器卡之间能够进行通信。第二中间面包括第三处理器卡和第四处理器卡和第二通信接口,其中每一个处理器卡包括处理器构造,其中所述第一处理器卡和所述第三处理器卡通过它们的处理器构造连接,并且其中所述第二处理器卡和所述第四处理器卡通过它们的处理器构造连接。第二通信接口连接所述第三处理器卡和所述第四处理器卡,使所述第三处理器卡和第四处理器卡之间能够进行通信。

[0017] 还提供了一种通过处理器构造连接的处理器卡的多节点配置的方法,用于对第一节点、第二节点和多个通信接口进行配置,所述第一节点包括多个处理器卡,所述多个处理器卡的每一个包括处理器构造,所述多个处理器卡通过它们的处理器构造连接;所述第二节点包括多个处理器卡,所述多个处理器卡的每一个包括处理器构造,所述多个处理器卡通过它们的处理器构造连接;其中所述多个通信接口的每一个接口连接所述第二节点中的一个处理器卡和所述第一节点中的一个处理器卡,使所述连接的处理器卡之间能够进行通信,从而协调所述第一节点和所述第二节点中所述连接的处理器卡之间的处理器操作。

[0018] 在又一个实施例中,所述通信接口和所述处理器构造使用不同的技术,其中对所述第一节点和所述第二节点以及所述处理器构造进行配置包括配置所述第一节点和所述第二节点,以分别作为第一和第二对称多处理器。

[0019] 在又一个实施例中,对所述第一节点和所述第二节点以及所述处理器构造进行配置进一步包括对多个中间面进行配置,其中所述每一个中间面包括至少两个处理器卡,并且其中所述第一节点和所述第二节点中的每一个至少包括一个来自每个中间面的处理器卡。

[0020] 在又一个实施例中,在所述第一节点或第二节点的其中一个上检测一个处理器卡的故障。确定包括所述故障处理器卡的所述节点是否至少包括一个运行的处理器卡。响应于确定所述节点至少包括一个运行的处理器卡,重新配置包括所述故障处理器的所述第一节点或所述第二节点,以在没有所述故障处理器的情况下进行操作。响应于确定所述节点不包括至少一个运行的处理器卡,执行切换以使用不包括所述故障处理器卡的所述第一节点或所述第二节点。

[0021] 在又一个实施例中,在所述第一节点或第二节点的其中一个上检测一个处理器卡的故障。执行切换以使用不包括所述故障处理器卡的所述第一节点或所述第二节点。

[0022] 附图说明

[0023] 图 1 示出了多节点计算环境的实施例。

[0024] 图 2 示出了多节点计算环境的附加实施例。

[0025] 图 3 和图 4 示出了在节点上配置处理器卡的操作实施例。

[0026] 图 5 和图 6 示出了执行切换和倒切换的操作实施例。

[0027] 具体实施方式

[0028] 图 1 示出了计算环境实施例。中间面 2a、2b 每个都包括两个处理器卡 4a、6a 和 4b、6b,其可包括插入到中间面 2a、2b 的扩展插槽中的扩展卡。作为替换,处理器卡 4a、4b、6a、6b 可以在中间面 2a、2b 板上实现。每个处理器卡 4a、4b、6a、6b 包括一个或多个核 10a、10b、10c、10d、12a、12b、12c、12d。尽管在每个处理器卡 4a、4b、6a、6b 上示出了两个核,每个处理器卡 4a、4b、6a、6b 还可以包括一个或两个以上的核。每个处理器卡 4a、4b、6a、6b 还可以包括高速缓存 14a、14b、14c、14d,其可以为它们的处理器卡 4a、4b、4c、4d 提供板上(on-board)存储器。处理器卡 4a、4b、6a、6b 上的处理器构造 16a、16b、16c、16d 分别提供了总线、端口、逻辑部件、判优器和队列等,使在同一处理器卡上的核 10a、10b、10c、10d、12a、12b、12c、12d 之间和在不同处理器卡 4a、4b、6a、6b 上的核之间能够互相通信。处理器构造 16a、16b、16c、16d 还包括构造控制器(未示出),其通过各种附加核 10a、10b、12a、12b、10c、10d、12c、12d 对构造总线进行争用调停(mediate contention),并通过其处理器构造 16a、16b、16c、16d 连接的所有处理器卡 4a、4b、6a、6d 的构造端口,规定对称多处理器(SMP)拓扑配置。电缆 18a、18b,如软电缆,被用于连接处理器构造 16a 和 16c,以及处理器构造 16b 和 16d。电缆 18a 和 18b 可以直接连接到处理器卡 4a、4b、6a、6b。

[0029] 所有连接到互连的处理器构造 16a、16b、16c、16d 的相同和不同的处理器卡 4a、4b、6a、6b 上的核 10a、10b、10c、10d、12a、12b、12c、12d,用作单个 SMP。例如,连接到互连的处理器构造 6a 和 16c 上的核 10a、10b、10c、10d 形成单个 SMP。互连构造 16a 和 16c 提供高速链接以便于在连接的核 10a、10b、10c、10d 和在连接的处理器卡 4a、4b 上的其他元件,如

高速缓存和存储器之间以连贯的方式传输数据,从而支持连贯性事务处理和并仲裁对互连构造 16a 和 16b 的访问。同样,核 12a、12b、12c、和 12d 连接到互连处理器构造 16a 和 16d,以形成单 SMP。因而,使用不同的中间面 2a、2b 上的处理器卡的组合可以形成独立的 SMP,其中每个中间面 2a、2b 给每个 SMP 提供一个处理器卡 4a、4b、6a、6b。

[0030] 每个中间面 2a、2b 还包括分别连接处理器卡 4a、6a 和 4b、6b 的网络适配器 20a 和 20b,以提供到网络的连接。每个中间面 2a、2b 上分别设有总线 22a、22b,其包括了网络适配器 20a、20b 和处理器卡 4a、6a 和 4b、6b 之间的通信接口。总线 22a 和 22d 可以利用本领域中公知的接口技术,如外围部件互连 (PCI)、PCI express 等。

[0031] 每个中间面 2a、2b 还包括分别连接到处理器卡 4a、6a 和 4b、6b 的输入 / 输出 (I/O) 适配器 24a 和 24b,以提供到本领域公知的如存储设备、打印机或其他 I/O 设备的附属外部设备的连接。每个中间面 2a、2b 上分别设有总线 26a、26b,其包括 I/O 适配器 24a、24b 和处理器卡 4a、6a 和 4b、6b 之间的总线通信接口。总线 26a 和 26b 可以利用本领域中公知的总线技术,如外围部件互连 (PCI), PCIexpress 等。

[0032] 群集总线 28a 和 28b 分别提供了在同一中间面 2a、2b 上的处理器卡 4a、6a 和 4b、6b 之间的通信接口,或提供在两个不同群集之间的连接。群集总线 28a 和 28b 可以利用本领域中公知的总线技术,如外围部件互连 (PCI), PCI express 等。分别在相同的中间面 2a 和 2b 中的处理器卡 4a 和 6a 上的核 10a、10b、12a、12b 和 10c、10d、12c、12d 可以通信对当一个处理器卡出现故障时到正常运行的处理器卡的切换和到使用恢复的处理器卡 4a 或 6a 的倒切换进行协调。

[0033] 在一个实施方式中,每个处理器卡 4a、6a、4b、6b 包括管理代码 30a、30b、30c、30d 以管理节点上的处理器卡 4a、6a、4b、6b 的配置,并按节点管理的切换和倒切换操作。管理代码 30a、30b、30c、30d 可以包括硬件和 / 或由处理器卡 4a、6a、4b、6b 上的一个或多个核 10a、10b、10c、10d、12a、12b、12c、12d 执行的计算机可执行代码。在一个替代实施例中,管理代码可以由处理器卡外部的配置为节点的元件来执行。

[0034] 在一个实施方式中,中间面 2a 和 2b 可以包括在独立存储服务器 或框中。此外,尽管图 1 示出了两个中间面 2a 和 2b,通过在不同中间面的处理器卡上互连处理器构造形成的对称多处理器,可以通过进一步对附加中间面中的处理器卡上的处理器构造进行互连而得到扩展。

[0035] 图 2 示出了包括以某种方式连在一起的多个存储服务器 50a、50b、50c 用作单个存储图象的网络计算环境。关于存储系统 56a 和 56b,存储服务器 50a、50b、50c 通过网络 54 接收 I/O 请求。存储系统 56a 和 56b 可以耦接到一个或多个存储服务器 50a、50b、50c。每个存储服务器 50a、50b、50c 可以实现图 1 的结构,其中每个服务器至少包括两个处理器卡 58a、60a、58b、60b 和 58c、60c。每个存储服务器 50a、50b、50c 可以有两个以上的所示处理器卡。此外,存储服务器 50a、50b、50c 上的处理器卡 58a、58b 和 58c 的处理器构造通过电缆 62a、62b、62c 互连,以形成包括来自所有三个存储服务器 50a、50b、50c 的处理器卡 58a、58b、58c 的一个对称多处理器或存储系统中的第一节点。此外,通过电缆 64a、64b 和 64c 对存储服务器 50a、50b、50c 上的处理器卡 60a、60b 和 60c 的处理器构造进行互连,以形成包括来自所有三个存储服务器 50a、50b、50c 的处理器卡 60a、60b、60c 的另一个对称的多处理器或存储系统的第二节点。



[0036] 在一个实施例中,每个处理器卡包括与相同节点上每个其他处理器卡连接的点对点处理器构造。在一个替代实施例中,并非在相同节点上所有的处理器卡(例如 58a、58b、58c 或 60a、60b、60c)都直接连接,而是在相同节点上的处理器卡可以通过连接在相同构造上的相同节点中的一个或多个其他处理器卡进行通信。以这种方式,多个存储服务器 50a、50b、50c 框通过一种方式连接,以提供两个对称的多处理器或节点,其中每个节点都在存储服务器上扩展。因此,节点包括通过处理器构造接口连接的处理器卡。在操作中,可在节点间进行负载平衡,以将主机请求分发给相同节点和跨节点的处理器卡,从而对节点间的工作负荷进行平衡。通过连接到中间面 2a、2b 中每个节点上的一个处理器卡 4a、6a、4b、6b,可以将所有的适配器 20a、20b 和 24a、24b 连接到一个系统中的两个节点上。

[0037] 存储服务器 50a、50b、50c 可以包括本领域公知的存储控制器或存储服务器,如企业级存储服务器或低端服务器产品。中间面 2a、2b 可以包括在不同的存储服务器框中或单个存储服务器框的刀片中。主机 52 可以包括本领域公知的能够与存储服务器 50a、50b、50c 进行 I/O 请求通信从而访问存储器 56a、56b 的计算设备,如服务器、大型机、工作站、个人电脑、手提电脑、膝上型电脑、电话设备、网络应用设备等。网络 54 可以包括存储区域网络(SAN)、局域网(LAN)、内部网、因特网、广域网(WAN)或对等等。存储器 56a、56b 可以包括如磁盘驱动器的单个存储设备、如磁盘阵列(JBOD)的存储设备阵列、直接访问存储设备(DASA)、独立冗余磁盘阵列(RAID)、虚拟化设备、磁带存储器或闪存等。

[0038] 图 3 示出了对中间面 2a、2b 上的节点进行的配置的操作实施例。当开始启动配置操作时(在方框 100),用户可以调用一个中间面/存储服务器上的处理器卡执行的代码以配置第一节点包括第一处理器卡,以及配置第二节点包括第二处理器卡(102)。用户也可以使用电缆连接不同中间面上的处理器卡(104)。

[0039] 图 4 示出了由处理器卡执行的配置处理器卡在节点中运行的操作。当检测到处理器卡 4a、4b、6a、6b 通过处理器构造 16a、16b、16c、16d 连接时(方框 120),检测到的处理器卡与连接在处理器构造 16a、16b、16c、16d 上的检测处理器卡进行通信(方框 122),以配置检测到的处理器卡在包括检测处理器卡的节点上运行。

[0040] 图 5 示出了在其中一个节点上进行切换和倒切换操作的操作实施例。操作可以由一个或两个节点上的处理器卡中的代码执行,在其中一个处理器卡发生故障的节点中的处理器卡和/或在另一个节点上的处理器卡通过处理器间群集总线 28a、28b 被警告故障。当利用跨多个存储服务器 50a、50b、50c 延伸的双节点存储系统执行操作时(图 2)(方框 150),检测在一个节点上的一个处理器卡 58a、58b、58c、60a、60b 或 60c 的故障(方框 152)。故障可以由发生故障的节点中的可用处理器卡和/或另一个节点中的处理器卡检测出并进行管理。如果在具有故障处理器的节点上有一个正常工作的处理器卡(方框 154),那么对该具有故障处理器卡的节点进行重新配置(方框 156),以将故障处理器卡从该节点移除,从而只使用该具有故障处理器的节点中的正常运行的处理器卡。可以对该故障处理器卡进行修理(方框 157)。另外,如果在具有故障处理器卡的节点中没有正常运行的处理器卡(方框 154),那么执行切换只使用不包括故障处理器的节点(方框 158)。

[0041] 当检测到先前发生故障的处理器卡恢复时(方框 160),如果两个节点都是运行的(方框 162),如在方框 156 之后的操作,那么对具有恢复处理器卡的节点进行重新配置(方框 164),以在节点操作中包括该恢复处理器卡。另外,如果只有一个节点是运行的(方框

162),如在方框 108 之后的操作,那么对两个节点都进行重新配置(方框 166),以倒切换到使用两个节点,这两个节点包括该恢复的处理器以使用该恢复的处理器卡和分配给那个节点的任何其他可用的处理器卡。由于 I/O 设备 20a、20b、24a、24b 连接到两个节点联合体,如果一个处理器卡 4a、6a、4b、6b 出现故障,没有 I/O 设备会丢失。相反只是通过具有该故障处理器卡的一个节点的连接会丢失,而通过具有正常运行的处理器卡的节点的连接不会丢失。

[0042] 图 6 示出了在其中一个节点上执行切换操作和倒切换操作的附加实施例。该操作可以由一个或多个节点的处理器卡上的代码执行,在其中一个处理器卡出现故障的节点上的处理器卡和 / 或另一个节点上的处理器卡通过处理器间群集总线 28a、28b 被警告故障。当利用跨多个存储服务器 50a、50b、50c 延伸的双节点存储系统执行操作时(方框 200),检测到一个节点中一个处理器卡的故障(方框 202)。响应于检测到的故障,执行从包括故障处理器卡的节点到不包括故障处理器的其他节点的切换,以从其他节点运行(方框 204)。因而,在该节点中的一个处理器卡的故障会引起到另一个节点的切换,即便是具有故障的节点具有其他运行的处理器卡。

[0043] 当检测到先前出现故障的处理器卡恢复时(方框 206),那么对两个节点都进行重新配置(方框 208),以倒切换到使用两个节点,这两个节点包括恢复的处理器卡,以使用该恢复的处理器卡和分配给那个节点的任何其他可用的处理器卡。

[0044] 所述实施例提供了在相同中间面或存储服务器提供跨处理器卡实现节点的技术,其中该节点中的处理器卡通过处理器卡上的处理器构造接口连接。所述实施例提供了一种在多个存储服务器框的存储服务器中实现不同节点的技术,以为允许在节点中和节点间的切换或倒切换操作提供冗余。

[0045] 附加实施例细节

[0046] 所述操作可以实现为使用标准编程和 / 或工程技术产生软件、固件、硬件或它们的任意组合的方法、装置或产品。所述操作可以实现为保存在计算机可读介质中的代码,其中处理器可以读取和执行计算机可读介质中的代码。计算机可读介质可以包括如磁性存储介质(例如硬盘驱动器、软盘、磁带等)、光学存储器(CD-ROM、DVD、光盘等)、易失性和非易失性存储设备等(例如 EEPROM、ROM、PROM、RAM、DRAM、SRAM、闪存、固件、可编程逻辑器件等)。执行上述操作的代码还可以实现为硬件逻辑器件(例如,集成电路芯片、可编程门阵列(PGA)、特定用途集成电路(ASIC)等)。此外,执行上述操作的代码可以在“传送信号”中实现,其中传送信号可以通过空间或传送介质传播,如光纤、铜线等。其中对代码和逻辑进行编码的传送信号还包括无线信号、卫星传送、无线电波、红外信号、蓝牙等。其中对代码和逻辑进行编码的传送信号能够通过发送基站被发送,并在接收基站被接收,其中在传送信息中被编码的代码和逻辑可以在接收或发送基站或设备上被解码和存储到硬盘或计算机可读介质中。“产品”包括其中代码可以被执行的计算机可读介质、硬件逻辑器件和 / 或传送信号。在其中对执行所述操作实施例的代码进行编码的设备,可以包括计算机可读介质或硬件逻辑器件。当然,本领域技术人员将认识到,在不背离本发明的范围内可以对这些设置进行修改,并且产品可以包括本领域公知的合适的信息承载介质。

[0047] 图 1 和图 2 中的所述实施例示出了在每个中间面 2a、2b 和存储服务器 50a、50b、50c 上的两个处理器卡,其中连接到不同中间面 / 存储服务器的一个或多个处理器卡的相

同中间面上的每个处理器卡形成节点对 I/O 请求进行操作。在一个替代实施例中,在每个中间面 / 存储服务器上存在两个以上的处理器卡,以允许形成两个以上节点,其中每个节点包括来自不同中间面 / 存储服务器的处理器卡。此外,在一个附加实施例中,节点可以包括在相同中间面 / 或存储服务器上的两个处理器卡。此外,不是直接连接到中间面 2a、2b,处理器卡 4a、4b、6a、6b 可以在具有 I/O 接口的单独集中实现,以与连接该处理器卡框与另一个处理器框或中间面的电缆匹配。

[0048] 在所述实施例中,节点管理对 I/O 设备的访问,如存储。在一个替代实施例中,节点可以执行处理操作而不是 I/O 操作,如计算操作或管理对外部设备如打印机、发送器等而不是存储器的访问。

[0049] 术语“一实施例”、“实施例”、“多个实施例”、“该实施例”、“这些实施例”、“一个或多个实施例”、“一些实施例”和“一个实施例”意味着“一个或多个(但不是全部)本发明的实施例”,除非另有说明。

[0050] 术语“包括”、“包含”、“具有”和它们的各种变形都意味着“包括但不限于”,除非另有说明。

[0051] 所列举的术语列表并不意味着任何或所有的术语都是相互排斥的,除非另有说明。

[0052] 术语“一”、“一个”、“该”都指“一个或以上”,除非另有说明。

[0053] 彼此通信的设备不需要彼此持续通信,除非另有说明。此外,彼此通信的设备可以通过一个或一个以上的中介介质直接或间接通信。

[0054] 对彼此相互通信的多个元件的实施例的描述并不意味着需要所有这些元件。相反,描述了多个可选元件,说明本发明可以有各种各样的可能实施方式。

[0055] 此外,尽管处理步骤、方法步骤、算法等可以以顺序地方式被描述,但这些处理、方法和算法可以配置成以交错的次序运行。换句话说,任何描述的序列或次序都不一定意味着步骤需要按照该次序执行。这里描述的处理步骤可以以任何实际的次序被执行。此外,一些步骤可以同步执行。

[0056] 当这里描述单个设备或产品时,显而易见的是可以使用一个以上的设备 / 产品(不管它们是否协作)来代替单个设备或产品。类似的,当在此描述了一个以上的设备或产品时(不管它们是否协作),显而易见的是可以使用单个设备 / 产品代替一个以上的设备或产品,或以不同于示出数量的设备或产品代替示出的数量的设备或程序。该设备的功能和 / 或特点可以可选择地由并未在此明示出的具有这种功能和特点的一个或多个其他设备来实现。因而,本发明的其他实施例不需要包括设备本身。

[0057] 图 3、图 4、图 5 和图 6 示例的操作示出了按某一次序发生的某一事件。在一个替代实施例中,某些操作可以以不同的次序被执行、修改或删除。此外,可以在上述逻辑中增加步骤,并仍然符合所述的实施例。此外,自此描述的操作可以顺序发生,某些操作也可以并行处理。此外,可以由单个处理单元或分布式处理单元执行操作。

[0058] 前述对本发明的多个实施例的描述是出于解释和说明的目的。并不意在对本发明进行穷尽或将本发明限制在描述的精确形式。根据上述教导,可以有很多的修改和变化。本发明的范围不应当被细节描述所限制,而应当由附加的权利要求进行限定。上述说明、例子和数据提供了本发明产品和构成使用的完整描述。由于在不背离本发明的精神和范围的情

---

况下,可以有多种实施方式,因此本发明由下文附属的权利要求来进行限定。

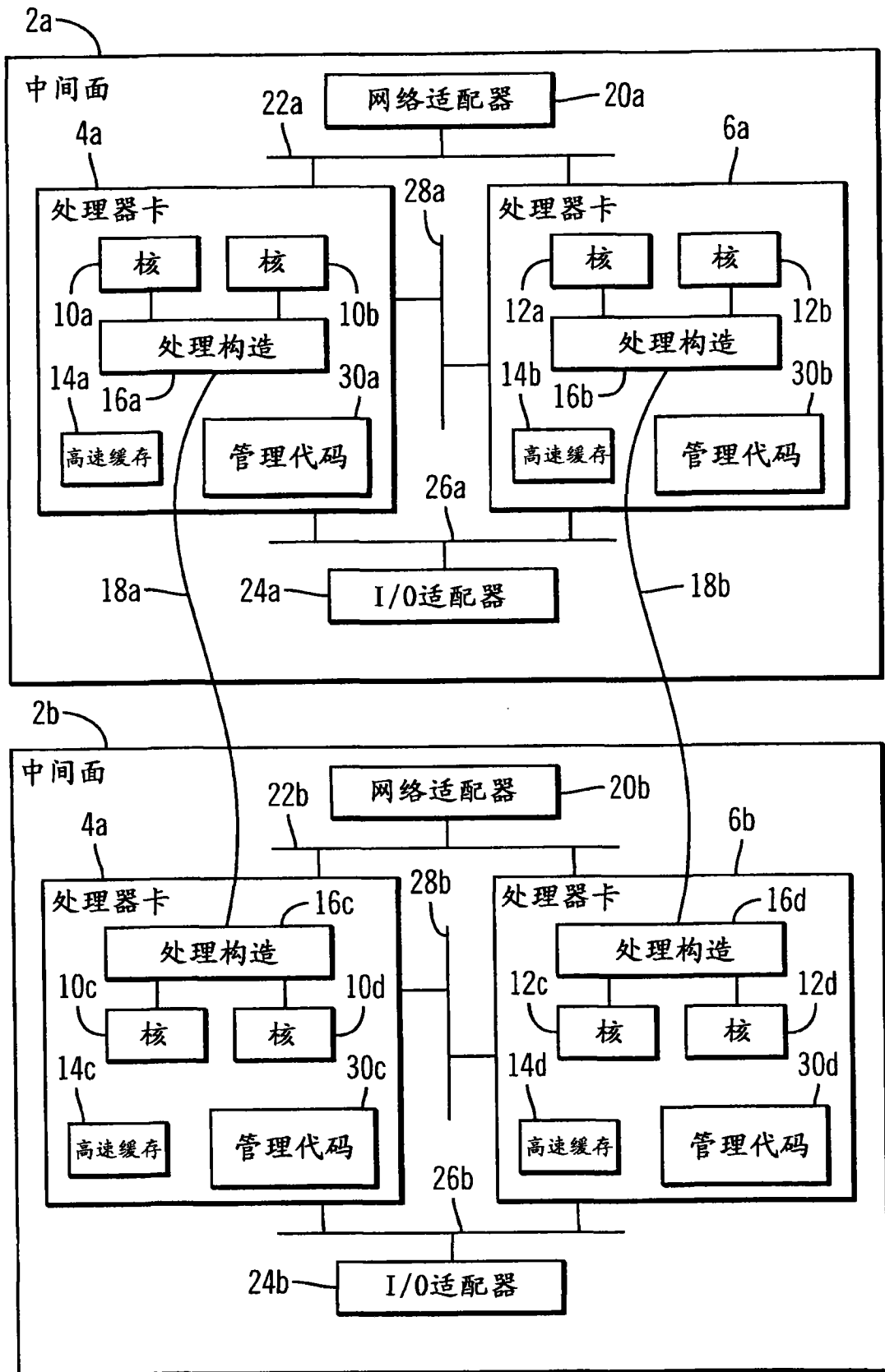


图 1

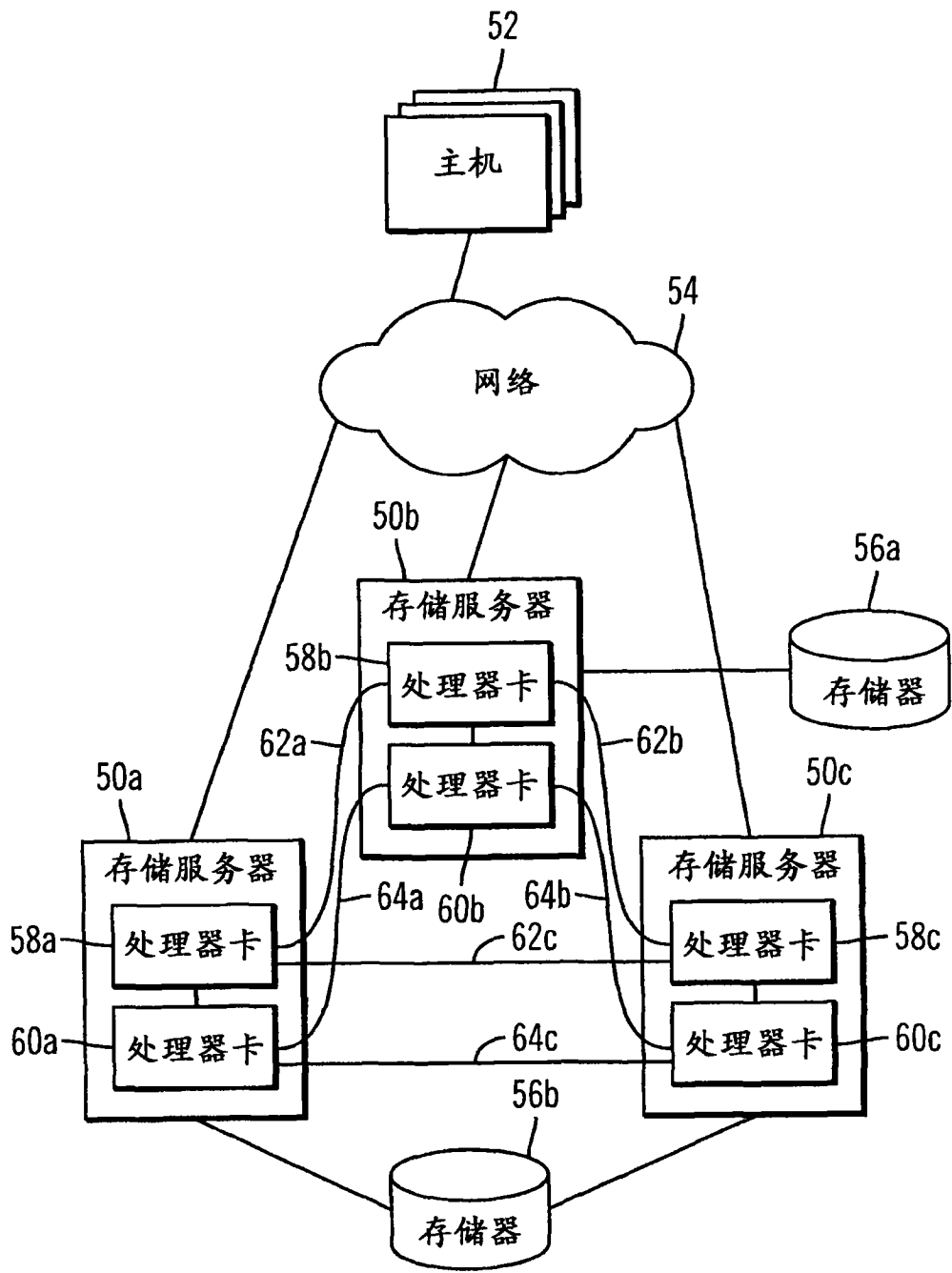


图 2

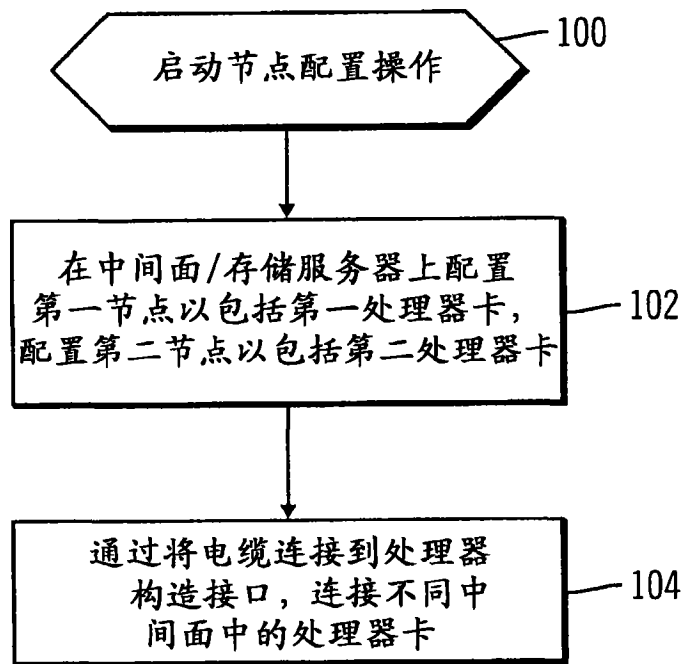


图 3

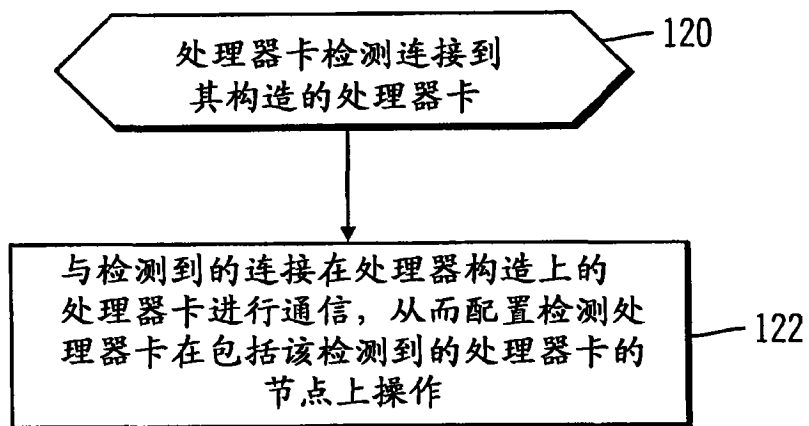


图 4

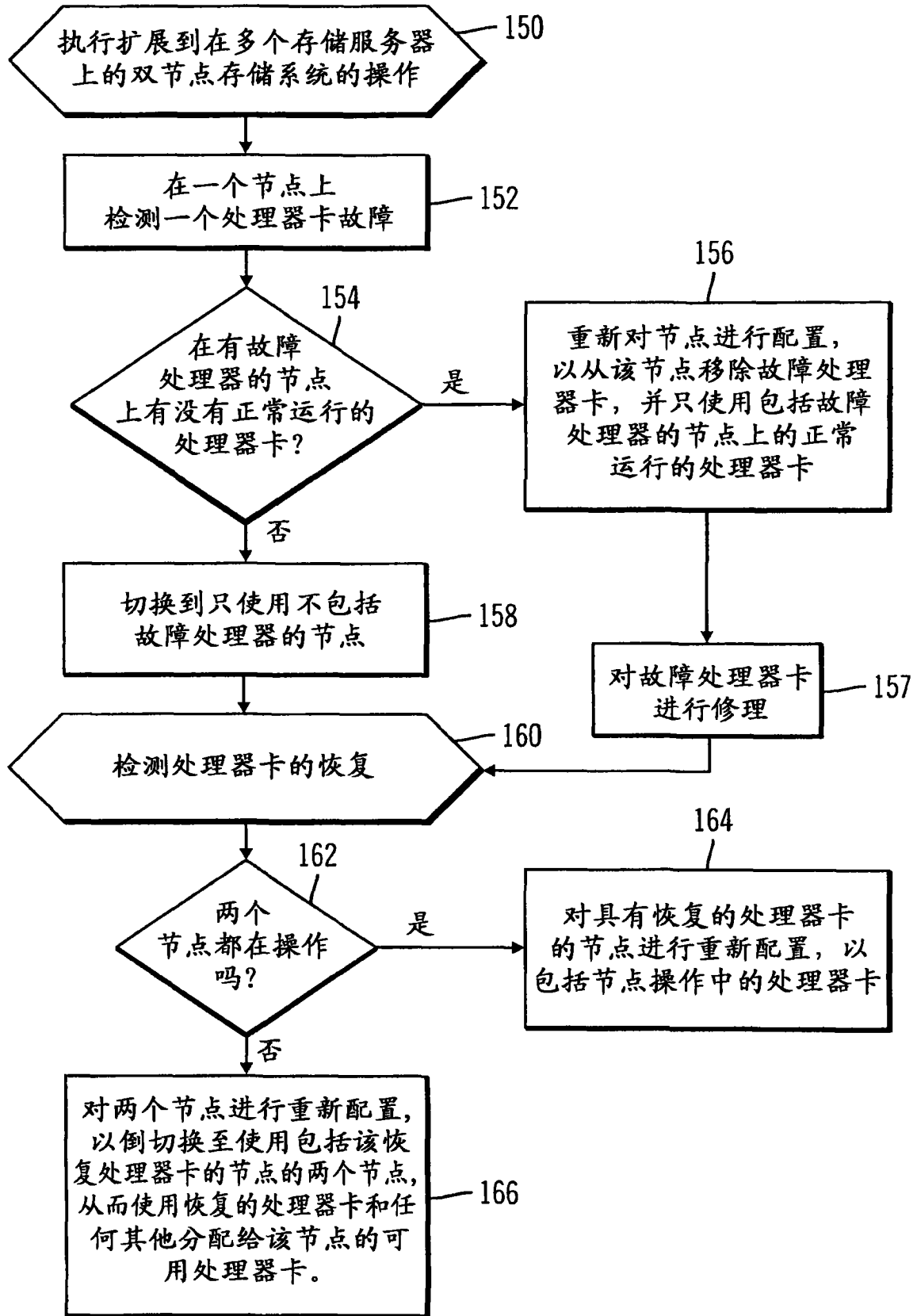


图 5



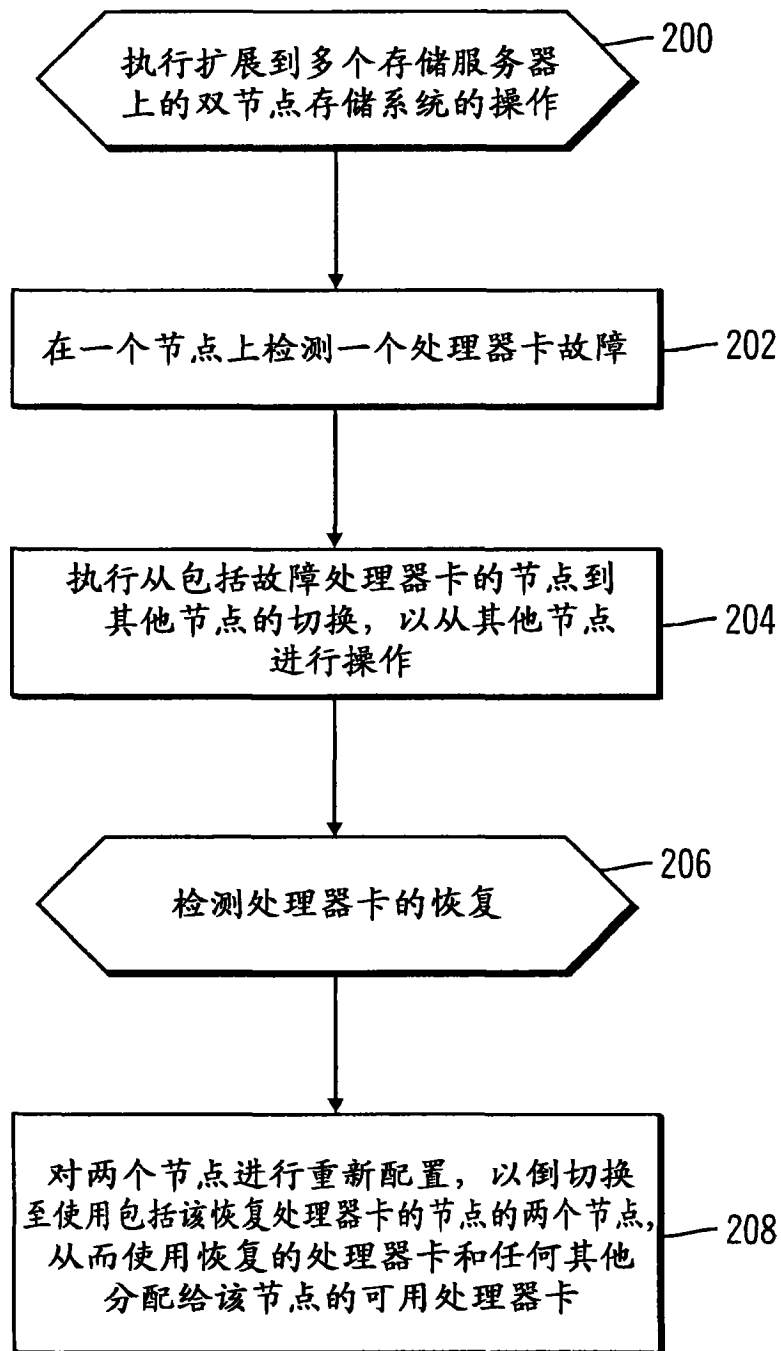


图 6