

(12) 发明专利

(10) 授权公告号 CN 101556493 B

(45) 授权公告日 2012. 05. 23

(21) 申请号 200810185018. 8

H04L 29/08 (2006. 01)

(22) 申请日 2008. 12. 26

(56) 对比文件

(30) 优先权数据

2008-100927 2008. 04. 09 JP

CN 1020972 C, 1993. 05. 26,

JP 昭 63-186313 A, 1988. 08. 01,

JP 特开平 10-031536 A, 1998. 02. 03,

(73) 专利权人 株式会社日立制作所

地址 日本东京都

审查员 刘佳

(72) 发明人 加藤猛 中岛忠克 齐藤达也

冲津润 志贺阳子 三木良雄

(74) 专利代理机构 北京银龙知识产权代理有限公司

公司 11243

代理人 许静

(51) Int. Cl.

G06F 1/20 (2006. 01)

G06F 1/32 (2006. 01)

G06F 9/50 (2006. 01)

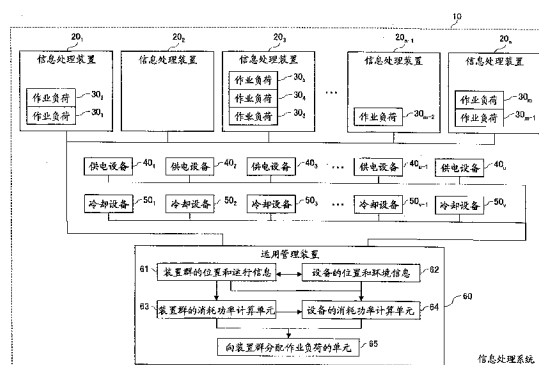
权利要求书 4 页 说明书 20 页 附图 12 页

(54) 发明名称

信息处理系统的运用管理方法及运用管理装置

(57) 摘要

本发明提供信息处理系统的运用管理方法及运用管理装置,用于削减信息处理装置群和供电/冷却设备的总消耗功率,实现信息处理系统的节电运用。信息处理系统(10)由信息处理装置(20₁~20_n)、供电设备(40₁~40_u)、冷却设备(50₁~50_v)和运用管理装置(60)构成。运用管理装置与装置群(20₁~20_n)和设备(40₁~40_u、50₁~50_v)相连,具有由装置群的位置和运行信息(61)以及设备的位置和环境信息(62)构成的配置信息。并且,运用管理装置使用该配置信息,求出装置群(20₁~20_n)的消耗功率、供电设备的供电损失和冷却设备的冷却功率(63、64),为了减小消耗功率、供电损失以及冷却功率的总和,向装置群(20₁~20_n)分配作业负荷(30₁~30_m)(65)。



1. 一种通过计算机系统实现的、由信息处理装置群和对该装置群进行供电或冷却的设备构成的信息处理系统的运用管理方法,其特征在于,

所述计算机系统执行以下步骤:

第一步骤,存储配置信息,该配置信息包含所述装置群的位置及运行信息和所述设备的位置及环境信息;

第二步骤,根据所述运行信息,求出与所述装置群的作业负荷对应的消耗功率;

第三步骤,根据所述配置信息和所述消耗功率,求出所述设备的供电损失或冷却功率;以及

第四步骤,为了减小所述装置群的消耗功率和所述设备的供电损失或冷却功率的总和,向所述装置群分配作业负荷。

2. 根据权利要求1所述的信息处理系统的运用管理方法,其特征在于,

在所述第二步骤中,根据所述装置群的规格信息、结构信息、运行信息,设定以所述装置群的作业负荷作为变量来计算消耗功率的装置功率函数,通过使用所述装置功率函数来计算消耗功率。

3. 根据权利要求1所述的信息处理系统的运用管理方法,其特征在于,

在所述第三步骤中,设定以所述装置群的消耗功率和所述配置信息作为变量来计算所述设备的供电损失或冷却功率的设备功率函数,通过使用所述设备功率函数来计算供电损失或冷却功率。

4. 根据权利要求1所述的信息处理系统的运用管理方法,其特征在于,

在所述第二步骤中,设定以所述装置群的作业负荷作为变量来计算消耗功率的装置功率函数,使用所述装置功率函数来进行消耗功率的计算,

在所述第三步骤中,设定以根据所述装置功率函数计算出的所述装置群的消耗功率和所述配置信息作为变量来计算所述设备的供电损失或冷却功率的设备功率函数,使用所述设备功率函数来计算供电损失或冷却功率,

在所述第四步骤中,以所述装置功率函数和所述设备功率函数的总和作为目标函数,关于向所述装置群分配作业负荷的组合最佳化问题,求出使所述目标函数最小化的最佳解或可执行的近似解。

5. 根据权利要求1所述的信息处理系统的运用管理方法,其特征在于,

在所述第二步骤中,设定以所述装置群的作业负荷作为变量来计算消耗功率的装置功率函数,使用所述装置功率函数来进行所述装置群的各个装置的消耗功率的计算,

在所述第三步骤中,针对根据所述装置功率函数计算出的所述装置群的各个装置的消耗功率,设定对所述装置群的每一个装置所需要的所述设备的供电损失或冷却功率进行计算的装置要素函数,使用所述装置要素函数来计算所述装置群的各个装置的供电损失或冷却功率,

在所述第四步骤中,通过将所述装置功率函数和所述装置要素函数合并,针对所述装置群的各个装置设定以作业负荷作为变量的装置函数,以与所述装置群的全部装置对应的所述装置函数的总和作为目标函数,关于向所述装置群分配作业负荷的组合最佳化问题,求出使所述目标函数最小化的最佳解或可执行的近似解。

6. 根据权利要求1所述的信息处理系统的运用管理方法,其特征在于,

在所述第四步骤中,以所述装置群的允许负荷条件和所述设备的允许功率条件作为制约条件,以所述装置群的消耗功率和所述设备的供电损失或冷却功率的总和作为目标函数,关于向所述装置群分配作业负荷的组合最佳化问题,求出使所述目标函数最小化的最佳解或可执行的近似解。

7. 根据权利要求 1 所述的信息处理系统的运用管理方法,其特征在于,
所述计算机系统还执行第五步骤,

所述第五步骤,根据所述装置群的位置和消耗功率,控制进行所述供电或冷却的设备,所述装置群的位置和消耗功率是根据分配给所述装置群的作业负荷和所述配置信息而得到的。

8. 根据权利要求 1 所述的信息处理系统的运用管理方法,其特征在于,
所述装置群构成虚拟化环境,

通过向物理装置群配置多个虚拟装置,来进行所述第四步骤中的所述装置群的作业负荷的分配。

9. 根据权利要求 1 所述的信息处理系统的运用管理方法,其特征在于,
具有与所述装置群对应的作业调度器,

通过所述作业调度器向所述装置群配置作业,来进行所述第四步骤中的所述装置群的作业负荷的分配。

10. 根据权利要求 1 所述的信息处理系统的运用管理方法,其特征在于,
所述装置群构成 SAN 启动环境,

通过选择所述 SAN 启动环境启动的装置,来进行所述第四步骤中的所述装置群的作业负荷的分配。

11. 根据权利要求 1 所述的信息处理系统的运用管理方法,其特征在于,

通过配置所述装置群中运行、休止或停止的装置,来进行所述第四步骤中的所述装置群的作业负荷的分配。

12. 根据权利要求 1 所述的信息处理系统的运用管理方法,其特征在于,
所述装置群由服务器装置构成,

所述计算机系统还执行第五步骤,

所述第五步骤,伴随所述装置群的作业负荷的分配的变更,在所述服务器装置间移动软件。

13. 根据权利要求 1 所述的信息处理系统的运用管理方法,其特征在于,
所述装置群由存储装置构成,

所述计算机系统还执行第五步骤,

所述第五步骤,伴随所述装置群的作业负荷的分配的变更,在所述存储装置间移动或拷贝数据。

14. 根据权利要求 1 所述的信息处理系统的运用管理方法,其特征在于,
所述装置群由网络装置构成,

所述计算机系统还执行第五步骤,

所述第五步骤,伴随所述装置群的作业负荷的分配的变更,变更所述网络装置的连接结构。

15. 一种通过计算机系统实现的、由信息处理装置群和对该装置群进行供电或冷却的设备构成的信息处理系统的运用管理方法,其特征在于,

所述计算机系统执行以下步骤:

第一步骤,监视所述装置群的运行信息以及与所述设备的供电或冷却相关的环境信息,将所述装置群的运行信息及位置和所述设备的环境信息及位置作为配置信息来存储;

第二步骤,根据当前或预定的作业负荷,向所述装置群分配虚拟作业负荷;

第三步骤,根据所述运行信息,计算与所述装置群的所述虚拟作业负荷对应的虚拟消耗功率;

第四步骤,根据所述配置信息和所述虚拟消耗功率,计算所述设备的虚拟供电损失或虚拟冷却功率;

第五步骤,为了减小所述装置群的虚拟消耗功率和所述设备的虚拟供电损失或虚拟冷却功率的总和,求出所述装置群的虚拟作业负荷的分配方法;以及

第六步骤,根据所述虚拟作业负荷的分配方法,向所述装置群分配实际作业负荷。

16. 根据权利要求 15 所述的信息处理系统的运用管理方法,其特征在于,

所述计算机系统还执行第七步骤,

所述第七步骤,根据所述第一步骤中存储的所述配置信息,对设置所述装置群和所述设备的空间中的所述装置群的位置坐标及运行信息、和该空间中的所述设备的位置坐标及与供电或冷却相关的环境信息进行映射来进行显示。

17. 一种通过计算机系统实现的、由能够以任意组合分配多个虚拟装置的多个物理装置和与所述多个物理装置进行供电或冷却的设备构成的信息处理系统的运用管理方法,其特征在于,

所述计算机系统执行以下步骤:

第一步骤,将所述多个物理装置的位置、与所述多个物理装置对应的所述多个虚拟装置的分配信息、所述多个物理装置及所述多个虚拟装置的运行信息、所述设备的位置及环境信息作为配置信息来进行存储;

第二步骤,根据所述运行信息,计算与所述多个虚拟装置的作业负荷对应的所述多个物理装置的消耗功率;

第三步骤,根据所述配置信息和所述消耗功率,计算所述设备的供电损失或冷却功率;以及

第四步骤,为了减小所述消耗功率和所述设备的供电损失或冷却功率的总和,求出向所述多个物理装置分配所述多个虚拟装置的组合。

18. 一种通过计算机系统实现的、由信息处理装置群和对该装置群进行供电或冷却的设备构成的信息处理系统的运用管理装置,其特征在于,

所述计算机系统具有以下功能:

第一功能,存储配置信息,该配置信息包含所述装置群的位置及运行信息和所述设备的位置及环境信息;

第二功能,根据所述运行信息,求出与所述装置群的作业负荷对应的消耗功率;

第三功能,根据所述配置信息和所述消耗功率,求出所述设备的供电损失或冷却功率;以及

第四功能,为了减小所述装置群的消耗功率和所述设备的供电损失或冷却功率的总和,向所述装置群分配作业负荷。

19. 根据权利要求 18 所述的信息处理系统的运用管理装置,其特征在于,

所述第二功能,根据所述装置群的规格信息、结构信息、运行信息,设定以所述装置群的作业负荷作为变量来计算消耗功率的装置功率函数,通过使用所述装置功率函数来计算消耗功率。

20. 根据权利要求 18 所述的信息处理系统的运用管理装置,其特征在于,

所述第三功能,设定以所述装置群的消耗功率和所述配置信息作为变量来计算所述设备的供电损失或冷却功率的设备功率函数,通过使用所述设备功率函数来计算供电损失或冷却功率。

信息处理系统的运用管理方法及运用管理装置

技术领域

[0001] 本发明涉及由服务器、存储系统、网络等信息处理装置群和对其进行供电或冷却的设备构成的信息处理系统的运用管理方法以及运用管理装置,尤其涉及适合于进行信息处理装置群和设备的综合的节电运用管理的技术。

背景技术

[0002] 在特开 2004-126968 号公报(专利文献 1)中,在对于并列计算机进行作业调度(job scheduling)的管理服务器中,根据计算机的温度传感器信息向温度低的计算机投入新作业,使作业从温度高的计算机向温度低的计算机移动,防止由于高温而导致的并列计算机的故障或性能下降。另外,在作业的移动前后,也根据温度信息估计各个计算机和每个计算机所具备的冷却装置的消耗功率,由此判断可否移动。

[0003] 在特开 2007-179437 号公报(专利文献 2)中,在多个计算机的管理系统中,根据计算机的温度分布和运行信息提取出过热计算机和非过热计算机,使软件从前者向后者移动,实现了节电化。另外,在提取出对象计算机后,比较软件移动前后的计算机功率和空调功率的变化,判断可否移动。根据运行信息求出计算机功率,根据温度分布求出空调功率,根据温度传感器或温度历史或运行信息求出温度分布。

[0004] 在美国专利申请公开第 2006/0259621 号说明书(专利文献 3)中,在对位于数据中心的多个服务器分配作业负荷的方法中,比较被请求的作业负荷的资料和历史资料,按照服务器和空调的功率达到最低的历史,向服务器分配所请求的作业负荷,若没有适合的历史则随机地分配。在历史资料中包含:服务器的位置、等级、运行信息、进气温度、排气温度、作业负荷的类型、服务器和空调的功率。服务器和空调的功率,根据服务器的进排气温度和比热和风量来求得,或者通过功率计来测定。

[0005] 在美国专利申请公开 2006/0259793 号说明书(专利文献 4)中,在对位于数据中心的多个服务器分配功率的方法中,为了接近理想的模拟的温度分布、即功率分布,在地理位置接近的服务器间或机架间进行功率预算的平衡,根据该预算分配,指定服务器的离散化的功率状态,防止了由于热点或冷点导致的服务器的故障。表示各服务器对应的理想功率的热系数,根据各服务器的排气温度和平均的服务器的基准排气温度和空调的供气温度来求出。

[0006] 【专利文献 1】特开 2004-126968 号公报

[0007] 【专利文献 2】特开 2007-179437 号公报

[0008] 【专利文献 3】美国专利申请公开第 2006/0259621 号说明书

[0009] 【专利文献 4】美国专利申请公开第 2006/0259793 号说明书

发明内容

[0010] 预想到近年的信息通信的飞速发展,而且面向广播和通信的融合时代,信息处理装置急剧增加。一方面,为了削减二氧化碳以防止地球变暖,迫切需要大幅度减小信息处理

装置的消耗功率。特别是在后端担当信息处理基础的数据中心中,除了信息处理装置群消耗的功率以外,进行其供电或冷却的设备所消耗的功率也占很大的比例,削减装置群和设备相加的总功率成为重要课题。

[0011] 为了数据中心的节电化,在服务器、存储系统、网络等的信息处理装置、供电设备、冷却设备、系统运用管理中开始了各种努力。在信息处理装置中,通过低功率器件/电路来提高单位消耗功率的性能、通过根据作业负荷的工作/待机状态切换来采用节电功能取得了进展。在供电设备中,开始了从变压器、不间断电源装置、配电盘、分电器等到装置群电源的供电系统的损失降低和变换效率提高;在冷却设备中,开始了空调机自身的运转效率提高或给排气口的气流设计的最佳化;而且开始引入局部冷却或液冷。在运用管理中,在主要的对策中列举了:运行信息监视、作业调度、以及基于虚拟化的装置群的运用效率改善或联合(consolidation)。

[0012] 供电设备或冷却设备,一般根据装置群的最大额定功率来设计,但运用时的设备的效率或功率很大程度上取决于设备所对应的装置群的功率的分布或变化。例如,电源变换效率取决于功率负荷,因此根据装置的工作功率和供电系统,供电损失不同。另外,空调机的效率取决于装置的功率即发热、空调机和装置间的位置关系和距离、温度、风量、风向等,因此根据装置的功率和配置,冷却功率大受影响。

[0013] 预想到今后数据中心的基于虚拟化的运用管理会扩展,为了节电化而利用联合或动态迁移(live migration),因此充分考虑设备所对应的装置群的功率分布的不均衡化或时间变化,需要削减装置群和设备的总功率的对策。以往,作为与服务器或计算机等装置群和冷却设备或空调设备相关的温度或功率的运用管理方法,已知若干个公知例,但在上述总体节电化方面,停留在局部或逐步的对策。

[0014] 其中,例如在专利文献1中,通过向并列计算机中温度低的计算机分配作业,抑制了计算机的温度上升,但在数据中心那样多种信息处理装置共存的情况下,不一定是温度低的装置节省电力。另外,在作业的移动前后,估计了计算机的消耗功率,因此即使与移动源和移动目的地相关的功率下降,但对于并列计算机全体来说,也只不过是局部的节电化。在消耗功率的估计中也包含对每个计算机附加的冷却装置的功率,但未考虑空调设备或计算机等的配置,有时也会出现即使通过作业的分配使计算机和冷却装置的功率下降,但空调功率上升,从而二者的总功率增加的情况。

[0015] 在专利文献2中,使软件从过热计算机向非过热计算机移动并分散,但温度低的计算机不一定功率低,并且,有时使软件向特定的计算机集中、使其它计算机待机或休止,更能在计算机群全体中节省电力。另外,在软件的移动前后,比较了计算机功率和空调功率的变化,但比较对象限于提取出的过热和非过热的计算机,并不考虑将计算机群和空调设备合在一起的总功率。

[0016] 在专利文献3中,根据服务器和空调的功率达到最低的历史资料,向服务器分配所请求的作业负荷,因此,对于历史的范围内的新请求,可以降低将服务器群和空调设备合在一起的功率,但当没有适合的历史时、或已分配的作业服务变化较大时无法应对或者仅根据历史积累来逐步地改善。另外,尽管在历史资料中包含了服务器的位置,但根据服务器进排气温度或功率计来求出空调功率,并未考虑服务器和空调设备的位置关系,向位于历史中出现的范围内的位置的服务器分配作业负荷,不一定使服务器群和空调设备的总功率

成为最小。

[0017] 在专利文献 4 中,为了接近理想的温度分布而在相邻的服务器间平衡功率预算,因此服务器群全体的合计功率不降低。另外,以温度的均衡化为目的,使功率宏观地分散给了服务器群,因此,与基于服务器的联合的节电化不相容。为了进行功率分配而参照了服务器的地理位置或排气温度、空调的供气温度,但未考虑空调设备的功率或配置,无助于空调功率的降低。对服务器指定的功率状态是待机模式(待机功率)和最大工作模式(最大功率),未考虑与作业负荷对应的功率变化,未利用服务器自身的节电功能。

[0018] 如上所述,现有技术存在只不过局部或逐步地降低信息处理装置群和空调设备的功率的问题。因此,本发明的目的之一在于,通过综合地考虑装置群的功率分布或变化以及与设备的相互配置,来削减将装置群和设备合在一起的总功率,实现以数据中心为代表的信息处理系统的节电运用。从概念上来讲,对于仅装置群的节电化来说,尽可能地将作业负荷集中分配给某个装置,使其它装置休止或停止的方式较好;对于仅设备的节电化来说,通过运转效率增高的配置,使装置群的功率尽量分散的方式较好。因此,本发明提供,为了满足装置群和设备的相反的要求,使作业负荷恰当地集中并且分散地向装置群分配,为使装置群和设备的总功率达到最小而进行最佳化的运用管理方法以及运用管理装置。本发明的上述以及其它目的和新特征,通过本说明书的记载以及附图可以明了。

[0019] 本发明的代表的实施方式的特征在于,具有由信息处理装置群的位置以及运行信息和设备的位置以及环境信息构成的配置信息,具备:根据运行信息求出与装置群的作业负荷对应的消耗功率的单元;以及根据配置信息,求出与装置群的消耗功率对应的设备的供电损失或冷却功率的单元,由此,以使装置群和设备的总功率达到最小的方式分配作业负荷。

[0020] 如专利文献 2 所示,不提取出过热/非过热装置,而为了在装置全体范围内掌握装置的位置和作业负荷而收集装置群的位置和运行信息。运行信息,作为对于从某装置向其它装置移动作业负荷后计算消耗功率来说充分的信息,根据需要,也包含装置的规格信息或结构信息。若是供电设备,则为了掌握从变压器到装置电源的供电系统的供电损失而收集设备的位置和环境信息,若是冷却设备,则为了掌握从外部空气经由空气或冷媒到达装置即热源的热交换循环和冷却功率而收集设备的位置和环境信息。除了设备的内置传感器或外部传感器以外,根据需要也从装置的运行信息中取得环境信息。

[0021] 在求出装置功率的单元中,不是根据专利文献 1 那样的温度信息或专利文献 3 那样的历史对照,与专利文献 4 那样最大工作/待机状态的固定功率不同,不限于专利文献 2 那样将装置对象限于过热/非过热装置,而根据运行信息,在装置群全体范围内求出根据作业负荷的多少而变化的功率。在求出设备功率的单元中,不根据专利文献 2 那样的温度分布或专利文献 3 那样的历史对照,不像专利文献 1 那样求出装置中包含的冷却装置的功率,而根据装置群和设备的相互配置信息求出与装置群的功率分布或变化对应的设备全体的功率。在向装置群分配作业负荷的单元中,对象不限于专利文献 1 那样的温度低的装置、专利文献 2 那样的非过热装置、专利文献 3 那样的留在历史中的装置、专利文献 4 那样的地理位置接近的装置,而是根据以装置群全体作为对象而求出的最佳解来分配作业负荷。

[0022] 本发明的代表实施方式的另一特征在于,具备:监视由装置群的位置及运行信息和设备的位置及环境信息构成的配置信息,根据当前或预定的作业负荷,向装置群分配虚

拟作业负荷的单元；根据运行信息，求出与装置群的虚拟作业负荷对应的虚拟消耗功率的单元；根据配置信息和虚拟消耗功率，求出设备的虚拟供电损失或虚拟冷却功率的单元，由此，求出使装置群的虚拟消耗功率和设备的虚拟供电损失还有虚拟冷却功率的总和最小的虚拟作业负荷分配的最佳解。

[0023] 向装置群的作业负荷分配中存在大量的组合，但通过分配虚拟作业负荷来搜索使将装置群和设备合在一起的总功率为最小的解，高效地找出最佳解，根据该最佳解向装置群分配实际作业负荷。不仅包含当前的作业负荷，也包含预定或预测的作业负荷来进行虚拟作业负荷的分配。在解的搜索中，可以考虑装置的工作温度条件或允许负荷条件、供电设备的配电盘或机架电源的允许功率条件、冷却设备的冷却能力或用于避免蓄热的区段允许功率条件等制约条件。

[0024] 而且，另一特征在于，具有对设置了装置群和设备的空间中的、装置群的位置坐标以及运行信息和设备的位置坐标以及与供电或冷却相关的环境信息进行了映射的配置信息。通过在虚拟空间上对实际空间中的装置群或设备的运行状况或配置状况进行再现，对共同的虚拟空间统一地进行装置群的作业负荷分配或设备的运转模拟，实施装置群和设备紧密协同工作的运用管理。另外，通过使虚拟空间可视化，可以一并高效率地进行装置群和设备的运行监视、业务运用、故障检测、资产管理、功率管理等。

[0025] 而且，另一特征在于，具有根据装置的规格信息、结构信息、测定信息、运行信息或运行历史，对应于装置的负荷计算消耗功率的装置功率函数，由此可以根据作业负荷的分配、变更、移动等简便地求出装置功率。例如，以表示作业负荷 j ($j = 1, 2, \dots$) 的变量集合作为 l_{jk} ($k = 1, 2, \dots$)，装置 i 的装置功率函数 p_{Di} ($i = 1, 2, \dots$) 如数学式 1 那样表示。

[0026] 【数学式 1】

[0027]
$$p_{Di} = p_{Di}(l_{j1}, l_{j2}, \dots) = p_{Di}(l_j)$$

[0028] 在向装置群的作业负荷分配变更时，在装置间置换变量集合，对一个装置重复分配作业负荷时，通过在变量集合间进行运算来求出与重复的作业负荷对应的功率。在具有作业负荷的情况下（装置工作），使装置功率函数 p_{Di} 作为多变量的线性函数或非线性函数，在没有作业负荷的情况下（装置休止或停止），成为提供功率值的不连续函数。装置功率函数 p_{Di} 也可以是与变量对应的数表的返回值或插补。

[0029] 例如，作为装置功率函数 p_{Di} ，若根据装置的规格信息或结构信息而得知最大额定功率，则适宜选择与装置的 ON/OFF 对应的阶跃函数；若得知最大功率和待机功率，则适宜选择以斜率和截距表示的一次函数；若作为运行信息而得知多个工作状态和待机状态、休止状态、以及停止状态，则适宜选择带有条件的函数集合；若可以监视装置的消耗功率，则适宜选择拟合了运行历史数据的二次函数等，即根据取得的信息和需要的计算精度来适宜的选择。为了简便，可以将变量集合 l_j 集中为影响功率的主要因素。例如，主要因素是：在服务器装置情况下是处理器的最大性能、使用率、工作状态（频率、电压）等；在存储装置的情况下是启动磁盘数、访问方式、访问时间等；在网络装置的情况下是传输吞吐量、切换频率等。

[0030] 而且，另一特征在于，通过具有对应于装置的消耗功率和位置以及设备的环境信息和位置，计算设备的供电损失或冷却功率的设备功率函数，可以根据作业负荷分配所引起的装置的功率分布，来简便地求出设备功率。例如，设装置 i 的功率为 p_{Di} ，设位置为 x_i (x_i ,

$y_i, z_i)$, 设设备 i ($i = 1, 2, \dots$) 的环境信息为 ε_i , 设位置为 x_i (x_i, y_i, z_i), 供电损失函数 p_s 和冷却功率函数 p_c 如数学式 2、数学式 3 那样表示。

【0031】 【数学式 2】

$$[0032] \quad p_s = p_s((p_{D1}, x_1), (p_{D2}, x_2), \dots, (\varepsilon_1, x_1), (\varepsilon_2, x_2), \dots)$$

【0033】 【数学式 3】

$$[0034] \quad p_c = p_c((p_{D1}, x_1), (p_{D2}, x_2), \dots, (\varepsilon_1, x_1), (\varepsilon_2, x_2), \dots)$$

【0035】 供电损失函数 p_s , 根据装置的位置和供电设备的环境信息和位置, 求出从变压器、不间断电源、配电盘、分电器等到装置电源的供电系统, 按照供电系统计算与装置功率即功率负荷对应的电源效率或配电损失等, 计算出供电设备全体的供电损失。冷却功率函数 p_c , 根据装置的位置以及功率(热源分布)和冷却设备的环境信息和位置, 进行放置了装置群的室内的热流体模拟; 和从室内冷却机到制冷机、冷却塔等室外机的冷却系统的热交换模拟, 计算出冷却设备全体的冷却功率。

【0036】 在装置的位置 x_i 中, 除了三维位置坐标以外, 可以代替使用能够确定搭载了装置的机架的二维位置和机架高度方向的号码、放置装置的地砖的二维位置和从地砖起的高度等三维位置的表达。在供电设备的环境信息 ε_i 中有: 电源或配电器或配线的功率负荷特性、在各设备或机架中监视的输入输出功率等。在冷却设备的环境信息 ε_i 中有: 空气或冷媒的温湿度、流量、流速、流向、装置的工作温度、机架或给排气口的进排气温度、室内机或室外机的给排温度或流量、外部空气的温湿度、各设备的热负荷特性等。在设备的位置信息 x_i 中, 除了三维位置以外, 也可以包含各设备的连接目的地和相互距离等拓扑连接信息。

【0037】 而且, 另一特征在于, 通过以装置功率函数和设备功率函数的总和作为目标函数, 定义向装置群分配作业负荷的组合最佳化问题, 由此求出使目标函数最小化的最佳解、或可以在最佳解附近执行的近似解。例如, 表示装置群和设备的总功率的目标函数 P , 使用数学式 1 的装置功率函数 p_{Di} , 数学式 2 以及数学式 3 的设备功率函数 p_s 、 p_c , 如数学式 4 那样表示。

【0038】 【数学式 4】

$$[0039] \quad P = \sum_i p_{Di} + p_s + p_c = \sum_i p_{Di}(l_j) + p_s(p_{Di}(l_j)) + p_c(p_{Di}(l_j))$$

【0040】 装置功率 p_{Di} 是取决于对装置 i 分配的作业负荷 j 的函数 $p_{Di}(l_j)$, 取决于装置功率 p_{Di} 的设备功率 p_s 、 p_c 以 $p_{Di}(l_j)$ ($i, j = 1, 2, \dots$) 为变量, 如函数 $p_s(p_{Di}(l_j))$ 、 $p_c(p_{Di}(l_j))$ 那样来表示 (p_s 和 p_c 是装置的位置 x_i 、设备的环境信息 ε_i 、位置 x_i 的函数, 但作了简略表示)。因此, 目标函数 P 取决于对装置 i 分配的作业负荷 j , 由此可以将削减装置群和设备的总功率的问题, 返回到从作业负荷分配的组合中搜索使目标函数 P 最小化的解的最佳化问题。对 n 台的装置群允许重复地分配 m 个作业负荷的组合数 N 如数学式 5 那样表示。

【0041】 【数学式 5】

$$[0042] \quad N = {}_n H_m = {}_{n+m-1} C_m = \frac{(n+m-1)!}{m!(n-1)!}$$

【0043】 当解决该组合最佳化问题时, 当 n 或 m 较小时、或进行作业负荷分配的时间间隔足够长时, 通过解空间的完全搜索来求出最佳解, 但在除此以外的情况下, 使用高效率地搜索解的近似解法。作为近似解法, 除了局部搜索法、登山法、模拟退火法 (Simulated Annealing)、禁忌搜索法 (Tabu Search) 等一般熟知的元启发式算法 (MetaHeuristic

Algorithms) 以外,也可以应用特别针对此次的问题而利用了装置群的性能 / 功率指标或设备特性的算法。

[0044] 而且,另一特征在于,通过将装置群的功率和位置以及设备的环境信息和位置所对应的设备功率函数近似地分解为表示每个装置所需要的设备功率的要素函数,简便地解决组合最佳化问题。将装置功率函数和设备要素函数的和作为装置函数,针对每个装置进行定义,因此,针对作业负荷分配,可以在短时间内计算出目标函数的值。例如,与装置 i 对应的供电损失和冷却功率的要素函数 p_{Si} 和 p_{Ci} 如数学式 6、数学式 7 那样表示 (p_{Si} 和 p_{Ci} 是设备的环境信息 ε_i 、位置 x_i 的函数,但作了简略表示)。

[0045] 【数学式 6】

$$p_{Si} = p_{Si}(p_{Di}, x_i)$$

[0047] 【数学式 7】

$$p_{Ci} = p_{Ci}(p_{Di}, x_i)$$

[0049] 设备功率函数 p_S 、 p_C 如数学式 2、3 所示,是与装置群和设备全体对应的函数,但在装置群的功率分布或设备的环境信息和位置未大幅度变化的情况下或者没有极端偏移等情况下,可以对相互作用或非线性作用进行近似,分为每个装置的设备要素函数 p_{Si} 、 p_{Ci} 来考虑。另外,为了加入比较接近的装置间的相互作用,可以修正设备要素函数。装置函数 p_i 是设备要素函数 p_{Si} 、 p_{Ci} 和装置功率函数 p_{Di} 的和,如数学式 8 那样表示。目标函数 P 是装置群全体的装置函数 p_i 的总和,如数学式 9 那样表示。

[0050] 【数学式 8】

$$p_i = p_{Di} + p_{Si} + p_{Ci} = p_{Di}(l_j) + p_{Si}(p_{Di}(l_j)) + p_{Ci}(p_{Di}(l_j))$$

[0052] 【数学式 9】

$$P = \sum_i p_i(l_j)$$

[0054] 当比较数学式 4 和数学式 9 的目标函数 P 时,在数学式 4 中,需要针对装置群的功率分布全体来求出设备功率 p_S 、 p_C ,但在数学式 9 中分为按每个装置求出装置函数 p_i 即可,因此,与作业负荷分配对应的目标函数 P 的计算变得容易,可以高效率地进行解的搜索。另外,若得到了装置函数 p_i 的具体的表达式,则作为应该优先分配作业负荷的装置的指标,利用变量的系数或式子的微分等,在最佳解的附近缩小解空间。

[0055] 而且,另一特征在于,通过以装置群的允许条件和所述设备的允许功率条件作为制约条件,解决向装置群分配作业负荷的组合最佳化问题,由此从组合中提取出可以执行的解,高效率地求出解。同时,通过遵守制约条件,高效地发挥装置群或设备的性能,防止功率超过或温度上升所导致的故障,使可靠性提高。

[0056] 若分配了超过装置的最大性能或最大资源的作业负荷,则作业负荷对应的处理时间延长,电能(功率 × 时间)增加。因此,考虑与作业负荷的变化对应的余量来设置装置的允许负荷条件。另外,在装置功率函数中,在表示作业负荷的变量集合 l_j 中选择主要影响功率的因素,但即使是不太影响功率的因素,有时对装置的性能或资源来说也是重要的。例如包括装置的存储器容量、磁盘容量、I/O 吞吐量等。这种因素,与装置功率函数不同地,作为用于判定可否进行作业负荷分配的允许负荷条件。

[0057] 在供电设备中,通过供电系统,具有电源、配电盘、配电器等的额定功率,需要遵守

以该额定功率作为上限的允许功率条件。在不满足该条件的情况下,需要减轻与该供电系统相关的装置群的作业负荷,或者限制装置群的功率。在冷却设备中,具有空调机、局部冷却装置、液冷装置、制冷机、冷却塔等的额定处理能力,需要将装置功率抑制在该能力以下。另外,在冷却设备中,为了将装置保持在工作温度范围内、避免过度的发热集中并提高冷却效率,可以对配置装置群的空间区域、区段、机架列等设置允许功率条件。

[0058] 而且,另一特征在于,根据解决向装置群的作业负荷分配问题而求出的最佳解或近似解和配置信息,得到与装置群的位置对应的消耗功率即功率分布,控制供电设备或冷却设备。由此,使设备的运转效率提高,可以进一步抑制供电损失或冷却功率。例如,在供电设备中,根据输出功率来改变并列连接的设备的运行台数,由此提高供电效率,在冷却设备中,根据发热分布来调节空气或冷媒的供给温度、给排气口或风扇的风量或风向、制冷机或冷却塔的温度或流量等,由此提高冷却效率。

[0059] 而且,另一特征在于,作为向装置群分配作业负荷的手段,利用虚拟化环境、作业调度器、SAN 启动环境、运行管理。考虑作业负荷分配的契机或周期、伴随作业负荷的移动的开销 (overhead) 或损失时间 (time loss)、作业负荷的变化或产生消失、装置的作业负荷的联系 (link)、可以取得的装置运行信息、向已设或新设的系统的导入容易度、系统的安全性或可靠性等来选择这些手段,可以并列或分层地共存。在虚拟化环境中通过向物理装置群配置虚拟装置,在作业调度器中通过向装置群配置作业,在 SAN 环境中通过配置启动的装置群,在运行管理中通过配置运行、休止或停止的装置群,向装置群分配作业负荷。

[0060] 而且,另一特征在于,反映向装置群的作业负荷分配的变更,为降低装置群和设备的总功率而使装置群运行。在服务器装置中,移动应用、虚拟机、作业等提供作业负荷的软件。在存储装置中,通过移动或拷贝访问频率高的数据,作业负荷移动到移动目的地。在网络装置中,通过变更网络连接结构,运用节点变迁。

[0061] 对通过本申请中公开的发明中的代表性发明所得到的效果进行说明,例如,在数据中心等信息处理系统中,通过使信息处理装置群和设备协同工作的运用管理,可以削减将装置群和设备合在一起的总消耗功率。

附图说明

[0062] 图 1 是表示本发明的实施方式 1 的信息处理系统的运用管理装置以及运用管理方法的结构图。

[0063] 图 2 是表示本发明的实施方式 2 的信息处理系统的运用管理装置以及运用管理方法的结构图。

[0064] 图 3 是表示本发明的实施方式 3 的信息处理系统的运用管理方法中的配置信息的布局图。

[0065] 图 4 是表示本发明的实施方式 4 的信息处理系统的运用管理方法中的信息处理装置群的运行信息的数表。

[0066] 图 5 是表示本发明的实施方式 4 的信息处理系统的运用管理方法中的装置功率函数的曲线图。

[0067] 图 6 是表示本发明的实施方式 4 的信息处理系统的运用管理方法中的装置功率函数的曲线图。

[0068] 图 7 是本发明的实施方式 5 的信息处理系统的运用管理方法中的设备功率函数的说明图。

[0069] 图 8 是表示本发明的实施方式 6 的信息处理系统的运用管理装置以及运用管理方法的结构图。

[0070] 图 9 是表示本发明的实施方式 7 的信息处理系统的运用管理装置以及运用管理方法的结构图。

[0071] 图 10 是表示本发明的实施方式 8 的信息处理系统的运用管理装置以及运用管理方法的结构图。

[0072] 图 11 是表示本发明的实施方式 8 的信息处理系统的运用管理方法的说明图。

[0073] 图 12 是表示本发明的实施方式 8 的信息处理系统的运用管理方法的说明图。

[0074] 符号说明

[0075] 10、110、210、410、610 : 信息处理系统 ; 20、120 : 信息处理装置 ; 30、130 : 作业负荷 ; 40、140、440、540、640 : 供电设备 ; 50、150、450、550、650 : 冷却设备 ; 60、160、460、560、660 : 运用管理装置 ; 61、161 : 装置群的位置和运行信息 ; 62、162 : 设备的位置和环境信息 ; 63 : 装置群的消耗功率计算单元 ; 64 : 设备的消耗功率计算单元 ; 65 : 向装置群的作业负荷分配单元 ; 163 : 装置群的虚拟消耗功率计算单元 ; 164 : 设备的虚拟消耗功率计算单元 ; 165 : 向装置群的虚拟作业负荷分配单元 ; 168 : 向装置群的实际作业负荷分配单元 ; 169 : 向设备的运转指示单元 ; 220 : 机架列 ; 221 : 装置 ; 240 : 变压器 ; 241 : 不间断电源装置 ; 242 : 配电盘 ; 250 : 冷却塔 ; 251 : 制冷机 ; 252 : 空调机 ; 253 : 格栅 ; 260 ~ 263、270 ~ 275、280、281 : 按钮 ; 300 : 规格信息 ; 310 : 结构信息 ; 320 : 运行信息 ; 420 : 物理服务器 ; 421 : 系统管理程序 (hypervisor) ; 430 : 虚拟机 ; 431、521、621 : OS ; 432、530、630 : 应用 (application) ; 461、561 : 运用管理管理器 ; 462 : 虚拟化管理器 ; 463、563、663 : 设备管理器 ; 520、620 : 服务器 ; 562 : SAN 管理器 ; 570、670 : 存储系统 ; 622 : 代理 ; 623 : 服务器装置 ; 661 : 功率管理器 ; 662 : 综合管理管理器 ; 671 : 存储装置 ; 672 : 磁盘阵列 ; 673 : 数据 ; 680 : 网络 ; 681 : 网络装置。

具体实施方式

[0076] 以下, 根据附图详细说明本发明的实施方式。在用于说明实施方式的所有附图中, 原则上对相同部件赋予相同符号, 省略其重复说明。另外, 在以下的实施方式中, 为方便起见, 在必要时分割为多个部分或实施方式来进行说明, 但除了特别明确表示的情况以外, 它们并非相互无关, 存在一方是另一方的一部分或全部的变形例、细节、补充说明等关系。

[0077] 另外, 在以下的实施方式中, 当说到要素的数等 (包含个数、数值、量、范围等) 时, 除了特别明确表示的情况、以及在原理上明显地限定为特定数等情况外, 不限定为该特定数, 也可以在特定数以上或以下。而且, 在以下的实施方式中, 其构成要素 (也包含要素步骤等), 除了特别明确表示的情况以及在原理上认为明显必须的情况以外, 当然不一定是必须的。同样地, 在以下的实施方式中, 当说到构成要素等的形状、位置关系等时, 除了特别明确表示的情况以及在原理上认为明显并非如此等情况以外, 假定实质上包含与其形状等近似或者类似的情况。关于上述数值以及范围也同样如此。

[0078] (实施方式 1)

[0079] 图 1 是表示本发明的实施方式 1 的信息处理系统的运用管理装置以及运用管理方

法的结构图。信息处理系统 10 由以下各部构成：信息处理装置 $20_1 \sim 20_n$ ；向装置群 $20_1 \sim 20_n$ 供电的供电设备 $40_1 \sim 40_u$ ；冷却装置群 $20_1 \sim 20_n$ 的冷却设备 $50_1 \sim 50_v$ ；管理装置群 $20_1 \sim 20_n$ 和设备 $40_1 \sim 40_u$ 、 $50_1 \sim 50_v$ 的运用管理装置 60。

[0080] 对装置群 $20_1 \sim 20_n$ 分配作业负荷 $30_1 \sim 30_m$ 。作业负荷 $30_1 \sim 30_m$ 如图 1 所示的一例那样，有以下情况：对装置分配单独的作业负荷的情况（装置 20_{n-1} 和负荷 30_{m-2} ）、分配多个作业负荷的情况（装置 20_1 和负荷 $30_1 \sim 30_2$ 、装置 20_3 和负荷 $30_3 \sim 30_5$ 、装置 20_n 和负荷 $30_{m-1} \sim 30_m$ ）、未分配作业负荷的情况（装置 20_2 ）。

[0081] 运用管理装置 60 与装置群 $20_1 \sim 20_n$ 和设备 $40_1 \sim 40_u$ 、 $50_1 \sim 50_v$ 连接，具有由装置群 $20_1 \sim 20_n$ 的位置和运行信息 61 以及设备 $40_1 \sim 40_u$ 、 $50_1 \sim 50_v$ 的位置和环境信息 62 组成的配置信息，具备：求出装置群 $20_1 \sim 20_n$ 的消耗功率的单元 63；求出供电设备 $40_1 \sim 40_u$ 的供电损失和冷却设备 $50_1 \sim 50_v$ 的冷却功率的单元 64；求出向装置群 $20_1 \sim 20_n$ 分配作业负荷 $30_1 \sim 30_m$ 的单元 65。运用管理装置 60 例如通过使用计算机系统的程序处理来实现。

[0082] 在实施方式 1 的运用管理方法中，运用管理装置 60 通过与装置群 $20_1 \sim 20_n$ 和设备 $40_1 \sim 40_u$ 、 $50_1 \sim 50_v$ 通信，收集装置群 $20_1 \sim 20_n$ 的运行信息和设备 $40_1 \sim 40_u$ 、 $50_1 \sim 50_v$ 的环境信息，取得配置信息 61 和 62。运用管理装置 60 通过单元 63，根据装置群 $20_1 \sim 20_n$ 的运行信息 61 求出与作业负荷 $30_1 \sim 30_m$ 对应的装置群 $20_1 \sim 20_n$ 的消耗功率，通过单元 64，根据装置群 $20_1 \sim 20_n$ 的位置和由单元 63 求出的消耗功率以及设备 $40_1 \sim 40_u$ 、 $50_1 \sim 50_v$ 的位置和环境信息 62，求出供电设备 $40_1 \sim 40_u$ 的供电损失和冷却设备 $50_1 \sim 50_v$ 的冷却功率，根据单元 63 和 64，为了减小装置群 $20_1 \sim 20_n$ 和设备 $40_1 \sim 40_u$ 、 $50_1 \sim 50_v$ 的总消耗功率，求出向装置群 $20_1 \sim 20_n$ 分配的作业负荷 $30_1 \sim 30_m$ 。

[0083] 在配置信息 61 中包含的装置群 $20_1 \sim 20_n$ 的位置信息，由各装置的位置坐标或确定位置的识别数据构成，根据需要，包含装置间的连接结构数据。装置群 $20_1 \sim 20_n$ 的运行信息由表示作业负荷量的数据和表示消耗功率的数据等构成，例如由性能、资源使用率、工作 / 休止状态、额定功率或测定功率、工作温度等构成。作业负荷量数据是在装置间可以置换的通用的计量，不依赖于特定的装置、作业负荷、装置和作业负荷的关系。消耗功率数据是直接或间接地表示与作业负荷量对应的装置的消耗功率特性的数值、数学式或数表，取决于各装置。

[0084] 在配置信息 62 中包含的设备 $40_1 \sim 40_u$ 、 $50_1 \sim 50_v$ 的位置信息，由各设备的位置坐标或确定位置的识别数据、设备间的连接结构数据构成。设备 $40_1 \sim 40_u$ 、 $50_1 \sim 50_v$ 的环境信息，由设备的运行数据和工作特性数据、周边环境的监视数据等构成，例如由功率、温度、湿度、流量、流向、额定输出或额定性能等构成。工作特性数据，在供电设备的情况下是表示与功率负荷对应的供电损失特性、在冷却设备的情况下是表示与热负荷对应的消耗功率特性的数值、数学式或数表。此外，装置群 $20_1 \sim 20_n$ 的运行信息和设备 $40_1 \sim 40_u$ 、 $50_1 \sim 50_v$ 的环境信息，例如装置的温度数据、设备的功率数据那样相互补充。

[0085] 在求出装置群 $20_1 \sim 20_n$ 的消耗功率的单元 63 中，作为一个方法，将所提供的作业负荷量每次都与已知的数值数据进行比对，通过插补或统计处理等，计算消耗功率。作为另一方法，预先以作业负荷量为变量，将相对于作业负荷量的装置电力的关系公式化，或者以作业负荷量为自变量，做成输出来自数表的返回值的例程 (routine)。前者的方法，伴随装

置群 $20_1 \sim 20_n$ 的运行信息的积累,在比对和计算中需要时间,但计算精度提高。后者的方法,通过伴随运行信息的积累来修正关系式或数表,在提高精度的同时压缩了时间,另外还可以清楚地掌握作业负荷量和装置功率的关系。

[0086] 在求出设备 $40_1 \sim 40_u$ 、 $50_1 \sim 50_v$ 的消耗功率的单元 64 中,作为一种方法,根据配置信息 61 中的装置群 $20_1 \sim 20_n$ 的位置和通过单元 63 取得的装置群 $20_1 \sim 20_n$ 的消耗功率、配置信息 62 中的设备 $40_1 \sim 40_u$ 、 $50_1 \sim 50_v$ 的位置和环境信息,每次都进行模拟来计算供电损失或冷却功率。作为另一方法,预先以设想条件进行详细模拟,通过线性近似,将与各装置的位置和功率相对应的设备的要素功率的关系作为数学式、数表或简易模拟器而求出。前者的方法,若计算时间允许,则得到较高精度。后者的方法,通过预先进行了详细模拟,可以在设想条件的范围内保证精度,同时缩短计算时间。

[0087] 在求出向装置 $20_1 \sim 20_n$ 分配作业负荷 $30_1 \sim 30_m$ 的单元 65 中有若干方法。一个方法,针对向装置群的作业负荷分配的组合,通过单元 63 和 64 每次都计算装置群的功率和设备的功率,比较各组合的计算结果,选择总功率达到最小的分配解。另一方法,根据通过单元 63 预先得到的与作业负荷量对应的装置功率的关系、和通过手段 64 预先得到的与装置群的功率分布对应的设备要素功率的关系,导出作业负荷分配中的优先顺序指标,按照该指标向装置群分配作业负荷。前者的方法适合于在组合较少的情况下求出总功率达到最小的最佳解。后者的方法适合于在组合较多的情况下高速地求出最佳解或近似解。在后者的方法中,可以根据解的精度和计算时间,针对临时得到的分配解计算装置群的功率分布,重新求出与之对应的设备要素功率的关系,使用该关系导出指标,再次求出作业负荷分配,重复操作直到前后的分配解收敛于规定条件内。

[0088] 通过实施方式 1 中表示的信息处理系统 10 的运用管理方法,具有由装置群 $20_1 \sim 20_n$ 的位置和运行信息 61 以及设备 $40_1 \sim 40_u$ 、 $50_1 \sim 50_v$ 的位置和环境信息 62 构成的配置信息,由此得到对装置群 $20_1 \sim 20_n$ 分配作业负荷 $30_1 \sim 30_m$ 的最佳解,具有可以削减将装置群 $20_1 \sim 20_n$ 的消耗功率和设备 $40_1 \sim 40_u$ 、 $50_1 \sim 50_v$ 的供电损失或冷却功率相加的总消耗功率的效果。另外,根据希望的契机或计划、新作业负荷的投入、已有作业负荷的结束等,更新向装置群 $20_1 \sim 20_n$ 的作业负荷分配,由此可以实现信息处理系统 10 的节电运用。

[0089] 作为信息处理系统 10,以数据中心 (data center) 或计算机室 (computer room) 为代表,但也可以应用于工业设备、商业设备、通信设备、交通机关等的业务运用系统。另外,不仅已设系统运用管理,作为系统的新设、增设、移设、修改等的设计 / 诊断工具也是有用的。即使是跨越多个设施或场所的系统,若进行远程管理则可以应用本实施方式。

[0090] 在信息处理装置 $20_1 \sim 20_n$ 中存在服务器装置、存储装置、网络装置等,在服务器装置中包含通用服务器、专用服务器、大型计算机、并列计算机、超级计算机、嵌入式计算机、个人计算机等,在存储装置中包含磁盘、固体盘、光盘、磁带等,在网络装置中包含路由器、交换机、集线器等,另外作为周边装置还包含打印机、复印机、调制解调器、显示器等。另外,本实施方式也可以应用于具有工作 / 待机 / 休止状态的按需切换 (Demand Base Switching) 功能、所谓的节电模式的装置。例如,在服务器中相当于 DBS,在存储系统 (storage) 中相当于 MAID,在网络中相当于端口电源控制等。硬件形态不限于刀锋型 (blade)、机架 (Rack Mount) 型、塔型、专用机箱型,可以多样化地应用。

[0091] 所谓作业负荷 $30_1 \sim 30_m$,在服务器装置的情况下是指应用 (application),在存储

装置的情况下是指访问数据,在网络装置的情况下是指转发数据等。在作业负荷量数据中,在服务器装置的情况下包括性能、资源使用率、工作 / 休止状态等;在存储装置的情况下包括启动磁盘数、访问方式、访问时间等;在网络装置的情况下包括转发数据包数、切换频率等。另外,作为作业负荷的类型,也可以区分固定 / 浮动小数点运算、事务处理、数据库处理、技术计算、数据的形式或粒度、保证带宽等来考虑装置的消耗功率特性。

[0092] 在供电设备 $40_1 \sim 40_u$ 中存在变压器、不间断电源装置、配电盘、分电器、安装有装置的机架的电源等,也包含功率传感器、电流 / 电压传感器、漏电传感器等,有时供电电源的发电站、送电设备、变电设备、供电目的地的装置的电源也作为对象。在冷却设备 $50_1 \sim 50_v$ 中存在空调机、制冷机、冷却塔、给排气口、风扇、管道、冷媒配管等,也包含配置在室内外的温度传感器、湿度传感器、流量 / 流向传感器等,有时液冷装置、机架列或机架前面 / 背面的局部冷却装置、装置的冷却装置也作为对象。

[0093] 在运用管理装置 60 中,具有设置信息处理装置群 $20_1 \sim 20_n$ 和设备 $40_1 \sim 40_u$ 、 $50_1 \sim 50_v$ 的公共管理器服务器的形态、或者使两者的管理器服务器协同工作的形态,也可以使装置 $20_1 \sim 20_n$ 的一部分来担负管理器功能。本实施方式的运用管理方法,可以作为中间件、应用、嵌入式控制软件、固件等软件来安装,例如也可以将解决最佳化问题的引擎作为硬件来安装。管理器为了取得装置的运行信息或设备的环境信息,在各装置或各设备中设置了代理、服务处理器、接口等,但装置群或设备也可以构建自主分散系统。此外,本实施方式中的削减装置群和设备的总功率的效果,通过求出表示装置群的位置以及运行信息和设备的位置以及环境信息的配置信息、和装置群的消耗功率的单元;求出设备的消耗功率的单元;求出向装置群分配作业负荷的单元而发挥出来,而不由实施方式 1 或图 1 所示的装置本身或计算步骤等限定。

[0094] (实施方式 2)

[0095] 图 2 是表示本发明的实施方式 2 的信息处理系统的运用管理装置以及运用管理方法的结构图。信息处理系统 110 由信息处理装置 $120_1 \sim 120_n$;向装置群 $120_1 \sim 120_n$ 供电的供电设备 $140_1 \sim 140_u$;冷却装置群 $120_1 \sim 120_n$ 的冷却设备 $150_1 \sim 150_v$;管理装置 $120_1 \sim 120_n$ 和设备 $140_1 \sim 140_u$ 、 $150_1 \sim 150_v$ 的运用管理装置 160_1 、 160_2 构成,向装置群 $120_1 \sim 120_n$ 分配作业负荷 $130_1 \sim 130_m$ 。

[0096] 运用管理装置 160_1 与装置群 $120_1 \sim 120_n$ 相连,监视装置群 $120_1 \sim 120_n$ 的运行信息,具有由装置群 $120_1 \sim 120_n$ 的位置和运行信息构成的配置信息 161,具备:求出装置群 $120_1 \sim 120_n$ 的消耗功率的单元 163;求出向装置群 $120_1 \sim 120_n$ 分配虚拟作业负荷 $130_1 \sim 130_m$ 的单元 165;向装置群 $120_1 \sim 120_n$ 分配实际作业负荷 $130_1 \sim 130_m$ 的单元 168。运用管理装置 160_2 与设备 $140_1 \sim 140_u$ 、 $150_1 \sim 150_v$ 相连,监视设备 $140_1 \sim 140_u$ 、 $150_1 \sim 150_v$ 的环境信息,具有由设备 $140_1 \sim 140_u$ 、 $150_1 \sim 150_v$ 的位置和环境信息构成的配置信息 162,具备:求出供电设备 $140_1 \sim 140_u$ 的供电损失和冷却设备 $150_1 \sim 150_v$ 的冷却功率的单元 164;向设备 $140_1 \sim 140_u$ 、 $150_1 \sim 150_v$ 提供运转指示的单元 169。

[0097] 在实施方式 2 的运用管理方法中,通过装置侧的运用管理装置 160_1 和设备侧的运用管理装置 160_2 协同工作,在运用管理装置 $160_1 \sim 160_2$ 间交换需要的数据。在运用管理装置 160_1 中,通过单元 163,根据装置群 $120_1 \sim 120_n$ 的运行信息 161 求出与作业负荷对应的装置群 $120_1 \sim 120_n$ 的消耗功率的关系。在运用管理装置 160_2 中,通过单元 164,根据运

用管理装置 160₂ 得到的设备 140₁ ~ 140_u、150₁ ~ 150_v 的配置信息 162, 求出与从运用管理装置 160₁ 得到的装置群 120₁ ~ 120_n 的位置和消耗功率对应的、供电设备 140₁ ~ 140_u 的供电损失和冷却设备 150₁ ~ 150_v 的冷却功率的关系。

[0098] 而且, 在运用管理装置 160₁ 中, 根据单元 163 中的与作业负荷对应的装置功率的关系、和单元 164 中的与装置功率分布对应的设备功率的关系, 以适合于装置群 120₁ ~ 120_n 或设备 140₁ ~ 140_u、150₁ ~ 150_v 的制约条件 166 和对应于作业负荷 130₁ ~ 130_m 的运用策略 167, 同时通过单元 165 使装置群 120₁ ~ 120_n 和设备 140₁ ~ 140_u、150₁ ~ 150_v 的总消耗功率最小化的方式, 导出向装置群 120₁ ~ 120_n 的虚拟作业负荷的分配解。接受该结果后, 运用管理装置 160₁ 通过单元 168 在适当的契机向装置群 120₁ ~ 120_n 分配实际作业负荷, 运用管理装置 160₂ 通过单元 169 根据装置群 120₁ ~ 120_n 的功率分布和配置信息 162, 向设置 140₁ ~ 140_u、150₁ ~ 150_v 提供运转指示。

[0099] 通过实施方式 2 中表示的信息处理系统 110 的运用管理方法, 在求出向装置群 120₁ ~ 120_n 的虚拟作业负荷的分配解的单元 165 中, 作为虚拟作业负荷, 与当前的作业负荷 130₁ ~ 130_m 一起, 也可以将预定或预测的作业负荷计算在内, 对于制约条件 166 或运用策略 167, 可以包含从当前开始的时间推移来良好预测地进行判断, 避免实际作业负荷分配中的由于从最佳解背离、违反条件导致的故障、由于烦杂的修正而导致的开销等。另外, 通过划分为单元 165 的虚拟作业负荷分配和单元 168 的实际作业负荷分配, 也考虑作业负荷的预测以及分配解的导出所需要的时间, 可以适当地设定在单元 168 中执行分配变更的契机、和在单元 169 中向设备 140₁ ~ 140_u、150₁ ~ 150_v 进行运转指示的契机。

[0100] 在制约条件 166 中包括: 考虑了装置 120₁ ~ 120_n 的工作温度、最大作业负荷、相对于作业负荷的变化的余量等的允许负荷条件; 考虑了供电设备 140₁ ~ 140_u 的额定功率、电源效率等的允许功率条件; 考虑了冷却设备 150₁ ~ 150_v 的额定能力、流量 / 流向等的区段或附近装置间的允许功率条件等。在运用策略 167 中记载的内容中包括: 作业负荷的优先度、执行契机、处理时间、进度、与作业负荷对应的装置 120₁ ~ 120_n 的资源分配、服务级别、功率成本等。在单元 165 中, 在向装置群 120₁ ~ 120_n 分配虚拟作业负荷的阶段, 通过在先提供制约条件 166 和运用策略 167, 与在临时求出分配解后验证制约条件 166 和运用策略 167 相比没有退步, 并且从虚拟作业负荷分配的组合中搜索候补, 因此高效地求出解。另外, 通过制约条件 166 可以避免装置群 120₁ ~ 120_n 或设备 140₁ ~ 140_u、150₁ ~ 150_v 的不适当或故障, 通过运用策略 167 可以在节电化的同时, 实现使服务提高的高级运用管理。

[0101] 此外, 在实施方式 2 中, 通过设置了装置侧的运用管理装置 160₁ 和设备侧的运用管理装置 160₂, 将两者的计算处理分离开来避免复杂化, 同时经由两者间的接口紧密地进行协同工作。若将两者安装在不同的管理服务器中, 则通过分散处理来实现高速化, 但若具有足够的计算性能, 则也可以安装在公共的管理服务器中。另外, 在实施方式 2 中, 可以根据装置群 120₁ ~ 120_n 的功率分布, 在先向设备 140₁ ~ 140_u、150₁ ~ 150_v 进行运转指示, 因此, 与以往那样被动地控制设备的情况相比, 不需要具有过剩的余量, 可以不浪费地进行高效率的运转。

[0102] (实施方式 3)

[0103] 图 3 是表示本发明的实施方式 3 的信息处理系统的运用管理方法中的配置信息的布局图, 表示例如实施方式 1 或实施方式 2 中所示的运用管理装置的显示画面的一例。信息

处理系统 210 由以下部分构成:信息处理装置群;搭载了信息处理装置群的机架列 $220_1 \sim 220_{10}$;由变压器 240、不间断电源装置 241 和配电盘 $242_1 \sim 242_4$ 构成的供电设备;由冷却塔 250、制冷机 251、空调机 $252_1 \sim 252_6$ 和格栅 (grill) 253 构成的冷却设备。

[0104] 供电设备从变压器 240 经由不间断电源装置 241 和配电盘 $242_1 \sim 242_4$ 向机架列 $220_1 \sim 220_{10}$ 供电,再从机架列的配电单元向信息处理装置供电。冷却设备进行地板下空气调节,空调机 $252_1 \sim 252_6$ 从格栅 253 (虽有多个但用一个符号代表) 供给冷空气,吸收由于装置群发热所加热的温空气,空调机 $252_1 \sim 252_6$ 和制冷机 251 和冷却塔 250 相互进行热交换,冷却塔向外部空气放热。

[0105] 实施方式 3 的布局 (layout) 是映射了装置群的位置坐标和运行信息以及设备的位置坐标和环境信息的配置信息,是将设置装置群和设备的实际空间信息在运用管理装置中的虚拟空间上再现而可视化的信息。配置信息,除了室内的装置的配置 (图 3 是全体图,因此未显示)、机架列 $220_1 \sim 220_{10}$ 、配电盘 $242_1 \sim 242_4$ 、空调机 $252_1 \sim 252_6$ 、格栅 253 等的配置以外,也表示从位于室外的变压器 240 和不间断电源装置 241 到位于室内的配电盘 $242_1 \sim 242_4$ 和机架列 $220_1 \sim 220_{10}$ 的供电系统的连接结构、和由位于室外的冷却塔 250 和制冷机 251 和空调机 $252_1 \sim 252_6$ 构成的热交换系统的连接结构。

[0106] 在实施方式 3 中,通过点击显示画面上的按钮 260 ~ 263、270 ~ 275、280、281,可以进行多种显示。通过作业负荷显示按钮 260 可以显示装置群的作业负荷分布。在布局的全体显示中针对每个机架列 $220_1 \sim 220_{10}$,在扩大显示中针对每个装置,通过图像的颜色和浓淡来显示服务器、存储系统 (storage)、网络等装置的种类和作业负荷量的多少,若在装置上安装了鼠标,则通过弹出式显示来显示作业负荷的详细数据。在弹出式 (hop up) 显示中,根据鼠标或键盘的操作可以显示装置的规格信息、结构信息、资产信息等。

[0107] 同样地,通过功率显示按钮 261 显示每个机架列 $220_1 \sim 220_{10}$ 或每个装置的消耗功率分布,并且显示变压器 240 和不间断电源装置 241 和配电盘 $242_1 \sim 242_4$ 的输出功率或供电损失、冷却塔 250 和制冷机 251 和空调机 $252_1 \sim 252_6$ 的功率,相加后通过数值显示装置群和设备的总功率。通过温度显示按钮 262,显示分布在室内的传感器的温度分布,并且显示装置的工作温度、机架列 $220_1 \sim 220_{10}$ 的进排气温度、空调机 $252_1 \sim 252_6$ 的给排气温度、格栅 253 的供气温度、冷却塔 250 或制冷机 251 的冷媒的供排温度、外部气温。通过故障显示按钮 263,与装置群或设备的故障一起显示与装置的工作温度条件或允许负荷条件、装置或设备的允许功率条件等对应的警报及其内容。

[0108] 图 3 是以 xy 平面看到的布局图,但可以通过按钮 270 ~ 275 改变平面来显示。按钮 270 指定 xy 平面显示,按钮 271 选择是在 z 轴全体范围内显示 xy 平面,还是指定特定的 z 坐标来截取 xy 平面。按钮 272 ~ 275 也同样。通过切换 xy 平面的按钮 270、yz 平面的按钮 272、zx 平面的按钮 274,可以显示三维空间的信息。按钮 280 和 281 分别是扩大显示和缩小显示,通过前者可以显示装置或设备的详细状况,通过后者可以显示全体状况。另外,也可以根据俯瞰地掌握全体的需要,来进行三维显示。

[0109] 通过实施方式 3 的配置信息,可以一并可视化地掌握设置了装置群和设备的实际空间中的装置群的运行状况或设备的环境信息,根据配置信息,可以使实施方式 1 或实施方式 2 中说明的装置群的作业负荷分配或设备的运转指示统一且紧密地协同工作。不仅进行削减将装置群和设备合并的总消耗功率的节电运用,还一并处理装置群和设备的运行

监视、功率管理、业务运用、故障检测、资产管理等,因此可以提高运用管理者的业务效率。另外,可以诊断现行系统中的装置群和设备双方的效率、风险、成本,因此在系统的移设、增设、修改、更新等的计划立案或施工中可以发挥作用。

[0110] (实施方式 4)

[0111] 图 4 是表示本发明的实施方式 4 的信息处理系统的运用管理方法中的信息处理装置群的允许信息的数表,表示信息处理装置是服务器时的一例。服务器的运行信息包含规格信息 300、结构信息 310、运行信息(或测定信息、运行历史)320。例如实施方式 1 或实施方式 2 所示的运用管理装置具有每个装置的规格信息 300、结构信息 310 和运行信息 320,作为同一数据文件或相互关联的不同数据文件来管理。此外,规格信息 300、结构信息 310 和运行信息 320 的项目、顺序、组合等,根据需要可以改变。

[0112] 规格信息 300 包含:装置管理号码、装置管理者、装置名称、装置品名、装置型号、装置性能 a、装置性能 b、装置性能 c、额定功率(最大)、额定功率(待机)、额定温度(最大)、额定温度(最小)等。结构信息 310 包含:CPU 型号、CPU 最大频率、CPU 搭载数、存储器型号、存储器容量、存储器搭载数、芯片组型号、磁盘型号、磁盘容量、磁盘搭载数、I/O 传输速度、I/O 个数等。运行信息 320 包含:装置位置坐标、作业负荷类型、装置/CPU 状态、CPU 工作频率、CPU 使用率、CPU 工作温度、存储器使用率、磁盘传输速度、I/O 传输速度、装置工作功率、装置进气温度、装置排气温度等。

[0113] 在实施方式 4 中,通过规格信息 300、结构信息 310 和运行信息 320,导出数学式 1 所示的装置功率函数。例如,在简便的导出方法中,设规格信息 300 的额定功率(最大)为 $p_{\max i}$ 、额定功率(最小)为 $p_{\text{idle}i}$ 、结构信息 310 的 CPU 最大频率为 $f_{\max}$ 、运行信息 320 的 CPU 工作频率为 f_i 、CPU 使用率为 α_i ,以 $f_i/f_{\max} \cdot \alpha_i$ 作为作业负荷量,装置 i 的装置功率函数 p_{Di} ($i = 1, 2, \dots$) 如数学式 10 那样表示。当具有多个 CPU 或者 CPU 中具有多个核等情况下,为了简便,取 f_i 和 α_i 的平均值即可。

[0114] 【数学式 10】

$$[0115] \quad p_{Di} \cong (p_{\max i} - p_{\text{idle}i}) \cdot \frac{f_i}{f_{\max i}} \cdot \alpha_i + p_{\text{idle}i}$$

[0116] 在没有作业负荷的情况下或未分配作业负荷的情况下,装置 i 的装置功率函数 p_{Di} ,例如设想 CPU 使用率 $\alpha_i = 0\%$,根据数学式 10 得出 $p_{Di} = p_{\text{idle}i}$ 。在作为运行信息 320 的装置/CPU 状态而具有休眠状态或休止状态时,在 CPU 使用率 $\alpha_i = 0\%$ 中,定义各状态所对应的功率 p_{Di} 即可。另外,假定当前在装置 i 中进行了作业负荷 $f_i/f_{\max} \cdot \alpha_i$,当估计将其向装置 j 移动后的装置功率时,设规格信息 300 的装置性能 a 为 γ_a 、装置性能 b 为 γ_b 、装置性能 c 为 γ_c ,运行信息 320 的作业负荷类型为 w ($w = a, b, c$),如数学式 11 那样进行替换,并将其带入装置 j 的装置功率函数 p_{Dj} 即可。

[0117] 【数学式 11】

$$[0118] \quad \frac{f_i}{f_{\max i}} \cdot \alpha_i \rightarrow \frac{\gamma_{wi}}{\gamma_{wj}} \cdot \frac{f_i}{f_{\max i}} \cdot \alpha_i$$

[0119] 如果替换后的作业负荷量 $\gamma_{wi}/\gamma_{wj} \cdot f_i/f_{\max} \cdot \alpha_i$ 超过 100%,设移动目的地装置 j 的功率 $p_{Dj} = p_{\max j}$,考虑到作业负荷的处理时间相当于超过的部分来延长(功率量增加)即可。此外,虽未特别限定,但所谓作业负荷类型,例如固定小数点运算处理、浮动小数点运算

处理、存储器读写处理等,是用于区别不依存于 CPU 频率而单纯依存于硬件架构,其处理能力会大幅变化的处理内容的类型。另外,装置性能 γ_w 未特别限定,是该每个作业负荷类型的基准测试积分 (Benchmark Score) 等。

[0120] 作为已有的测定信息或运行历史而得到运行信息 320 的装置工作功率时,根据作业负荷量 $f_i/f_{\max} \cdot \alpha_i$ 和装置工作功率 p_{measi} 的数据求出装置功率函数 p_{Di} 的近似式。图 5 是将这些数据绘图的一例,是表示近似的装置功率函数 p_{Di} 的图表。函数 p_{Di} 将一次函数的斜率设为 a_{Di} ,将截距设为 b_{Di} ,如数学式 12 那样表示。在图 5 的例子中,实线表示近似函数, $a_{Di} = 1.3175$, $b_{Di} = 152.16$ 。

[0121] 【数学式 12】

$$[0122] \quad p_{Di} \cong a_{Di} \cdot \frac{f_i}{f_{\max i}} \cdot \alpha_i + b_{Di}$$

[0123] 装置 i 在作为节电功能而根据作业负荷进行多个工作状态 (CPU 工作频率、CPU 核电压) 的按需切换时,装置功率 p_{Di} 阶段状地变化。在这种情况下,对于各工作状态和作业负荷量范围分别定义数学式 10 或数学式 12 这样的一次函数,通过函数集合来表示装置功率函数 p_{Di} 。作为别的简便方法,也可以通过一个二次函数来近似阶段状的装置功率函数 p_{Di} 。图 6 是将三个状态的按需切换中的作业负荷量 $f_i/f_{\max} \cdot \alpha_i$ 和装置工作功率 p_{measi} 的数据绘图的一例,是表示二次近似的装置功率函数 p_{Di} 的图表。将二次项系数设为 a_{Di} ,将一次项的系数设为 b_{Di} ,将常数项设为 c_{Di} ,如数学式 13 那样表示函数 p_{Di} 。在图 6 的例子中,虚线表示阶段状的一次函数,实线表示近似的二次函数, $a_{Di} = 0.0134$, $b_{Di} = 0.7297$, $c_{Di} = 103.16$ 。

[0124] 【数学式 13】

$$[0125] \quad p_{Di} \cong a_{Di} \cdot \left(\frac{f_i}{f_{\max i}} \cdot \alpha_i \right)^2 + b_{Di} \cdot \left(\frac{f_i}{f_{\max i}} \cdot \alpha_i \right) + c_{Di}$$

[0126] 当无法直接取得装置 i 的装置工作功率 p_{measi} ,而得知热电阻和热时间常数时,可以根据运行信息 320 的进气温度和排气温度,间接地导出工作功率 p_{measi} 。将热电阻设为 R_{thi} ,将热时间常数设为 τ_{thi} ,将进气温度设为 T_{ini} ,将排气温度设为 T_{outi} ,如数学式 14 那样表示工作功率 p_{measi} 。与工作功率的变化相比,温度具有热时间常数而缓慢地变化,因此考虑温度变化的斜率来进行修正。

[0127] 【数学式 14】

$$[0128] \quad p_{\text{measi}} \cong \frac{1}{R_{thi}} \cdot \left\{ (T_{outi} - T_{ini}) + \tau_{thi} \cdot \frac{d(T_{outi} - T_{ini})}{dt} \right\}$$

[0129] 若间接地得知装置工作功率 p_{measi} ,则可以与作业负荷量的 $f_i/f_{\max} \cdot \alpha_i$ 的测定结合,导出数学式 12 或数学式 13 那样的装置功率函数 p_{Di} 。另外,即使热电阻 R_{thi} 和热时间常数 τ_{thi} 最初不明,若已知额定功率 (最大) $p_{\max i}$ 或额定功率 (最小) $p_{\text{idle}i}$,则根据进气温度 T_{ini} 以及排气温度 T_{outi} 的测定数据或温度变化的斜率数据,可以计算出热电阻 R_{thi} 或热时间常数 τ_{thi} 。

[0130] 通过实施方式 4 的装置功率函数,可以简便地求出与作业负荷对应的装置群的消耗功率,良好预测地掌握作业负荷量和装置功率的关系,可以高效地进行针对装置群的作业负荷分配。另外,即使没有取得全部装置的运行信息,也可以如数学式 10 或数学式 14 那样根据一部分数据导出装置功率函数,若得到测定信息或运行历史,则可以如数学式 12

或数学式 13 那样导出与实际测量符合的装置功率函数,而且,若存储了测定信息或运行历史,则可以提高近似精度。

[0131] 在实施方式 4 中,以服务器装置作为一例,但即使是存储装置或网络装置,也可以同样地根据装置的运行信息导出装置功率函数。在实施方式 4 中,作为表示作业负荷量的主要变量,使用了 CPU 最大频率、CPU 工作频率、CPU 使用率,但若是存储装置,则可以使用启动磁盘数、访问方式、访问时间等,若是网络装置则可以使用转发数据包数、切换频率等。

[0132] (实施方式 5)

[0133] 图 7 是本发明的实施方式 5 的信息处理系统的运用管理方法中的设备功率函数的说明图,以实施方式 3 的信息处理系统 210 为例。例如,当装置 $i221_i$ 位于机架列 220_7 时,供电设备如箭头 243 所示那样,从变压器 240 经由不间断电源装置 241 和配电盘 242_3 向装置 i 供电。装置 $i221_i$ 消耗的功率变化为显热,加热空气。在冷却设备中,从空调机 $252_1 \sim 252_6$ 经由地板下、格栅(在图 7 中为了易于观察,省略了地板砖和格栅)向装置 $i221_i$ 供给冷空气,装置 $i221_i$ 的温空气如箭头 $254_1 \sim 254_6$ 所示那样返回空调机 $252_1 \sim 252_6$,空调机 $252_1 \sim 252_6$ 吸取的热经由制冷机 251 和冷却塔 250 被放出到外部空气中。

[0134] 与装置 i 同样地,若在装置群全体范围内得知位置坐标和装置功率或装置功率函数,则可以计算供电设备的供电损失和冷却设备的冷却功率。例如实施方式 4 那样,如果根据运行信息 320 得到装置群的位置坐标来得到数学式 10、数学式 12 或数学式 13 那样的装置功率函数,则可以求出装置群全体的功率分布,将它们作为输入,通过数学式 2 或数学式 3 所示的设备功率函数 p_s 、 p_c 计算设备功率。在设备功率函数 p_s 、 p_c 的计算中,可以使用预先输入了设备的位置坐标、连接结构、工作特性等环境信息的模拟器。

[0135] 为了进一步简便地求出设备功率,通过在规定条件的范围内进行线性近似,将设备功率函数 p_s 、 p_c 分解为数学式 6 或数学式 7 所示的设备要素函数 p_{si} 、 p_{ci} 。设备功率函数 p_s 、 p_c 成为每个装置的要害函数 p_{si} 、 p_{ci} 的装置群全体的线性和。通过按照箭头 243 所示的供电系统预先进行模拟而求出装置 i 的供电设备的要素函数 p_{si} 。如箭头 $254_1 \sim 254_6$ 所示,考虑装置 i 给空调机 $252_1 \sim 252_6$ 的热负荷的影响,预先进行模拟来求出冷却设备的要素函数 p_{ci} 。在模拟中,在分配装置 i 的功率的同时计算供电/冷却设备功率的灵敏度系数。在简化的情况下,将设备要素函数 p_{si} 、 p_{ci} 作为一次函数,将斜率设为 a_{si} 、 a_{ci} ,将截距作为 b_{si} 、 b_{ci} ,如数学式 15、数学式 16 那样表示。

[0136] 【数学式 15】

$$[0137] \quad p_{si} \cong a_{si} \cdot p_{di} + b_{si}$$

[0138] 【数学式 16】

$$[0139] \quad p_{ci} \cong a_{ci} \cdot p_{di} + b_{ci}$$

[0140] 若求出设备要素函数 p_{si} 、 p_{ci} ,则可以求取与装置功率函数 p_{di} 的和,导出如数学式 8 那样的装置函数 p_i 。装置函数 p_i 表示每个装置消耗的装置功率和设备功率。例如,当使用实施方式 4 中说明的数学式 12 所示的装置功率函数 p_{di} 时,如数学式 17 那样表示装置函数 p_i 。重新定义装置函数 p_i 的斜率 a_i 和截距 b_i 。装置群和设备的总消耗功率 P 是装置群全体的装置函数 p_i 的总和,将数学式 9 改写成数学式 18 那样。

[0141] 【数学式 17】

$$[0142] \quad p_i = p_{Di} + p_{Si} + p_{Ci} = a_i \cdot \left(\frac{f_i}{f_{\max i}} \cdot \alpha_i \right) + b_i$$

$$[0143] \quad a_i = a_{Di} \cdot (1 + a_{Si} + a_{Ci})$$

$$[0144] \quad b_i = b_{Di} \cdot (1 + a_{Si} + a_{Ci}) + b_{Si} + b_{Ci}$$

[0145] 【数学式 18】

$$[0146] \quad P = \sum_i p_i = \sum_i \left\{ a_i \cdot \left(\frac{f_i}{f_{\max i}} \cdot \alpha_i \right) + b_i \right\}$$

[0147] 若看数学式 18 则可以明了,通过将向装置群的作业负荷(在此, $f_i/f_{\max} \cdot \alpha_i$) 的分配最佳化,可以将装置群和设备的总功率 P 最小化。即,为了使目标函数 P 最小化,解决向装置群分配作业负荷的组合最佳化问题即可。当装置较多时,组合的数量如数学式 5 所示那样增大,因此,进行组合的全部搜索是不现实的。在实施方式 5 中,通过进行适当的近似,使数学式 18 的目标函数 P 成为一次函数的线性和,因此以装置函数 p_i 的斜率 a_i 和截距 b_i 作为指标,可以高效地发现最佳解。

[0148] (实施方式 6)

[0149] 图 8 是表示本发明的实施方式 6 的信息处理系统的运用管理装置以及运用管理方法的结构图。信息处理系统 410 由物理服务器 420₁ ~ 420_n、向服务器群 420₁ ~ 420_n 供电的供电设备 440₁ ~ 440_u、冷却服务器群 420₁ ~ 420_n 的冷却设备 450₁ ~ 450_v、管理服务器群 420₁ ~ 420_n 和设备 440₁ ~ 440_u、450₁ ~ 450_v 的运用管理装置 460 构成,构成虚拟化环境。物理服务器 420₁ ~ 420_n 中,通过系统管理程序 (hypervisor) 421₁ ~ 421_n 配置虚拟机 430₁ ~ 430_m,在其上运行 OS 431₁ ~ 431_m 和作为作业负荷的应用 432₁ ~ 432_m。在运用管理装置 460 中,管理物理服务器 420₁ ~ 420_n 和虚拟机 430₁ ~ 430_m 的虚拟化管理器 462;管理设备 440₁ ~ 440_u、450₁ ~ 450_v 的设备管理器 463;削减服务器群 420₁ ~ 420_n 和设备 440₁ ~ 440_u、450₁ ~ 450_v 的总消耗功率的运用管理管理器 461 协同工作。

[0150] 在实施方式 6 中,虚拟化管理器 462 通过系统管理程序 421₁ ~ 421_n 管理针对物理服务器 420₁ ~ 420_n 的虚拟机 430₁ ~ 430_m 的分配、或针对虚拟机 430₁ ~ 430_m 的资源分配 (CPU、存储器容量等资源分配),取得虚拟机 430₁ ~ 430_m 和物理服务器 420₁ ~ 420_n 的运行信息 (例如物理服务器的 CPU 使用率、存储器使用容量、磁盘传输速度、网络传输速度等;虚拟机的 CPU 使用率、存储器使用率、磁盘传输速度、网络传输速度等)。根据该运行信息,运用管理管理器 461 估计应用、即作业负荷 432₁ ~ 432_m,导出与作业负荷对应的物理服务器 420₁ ~ 420_n 的功率的关系。例如,当对于一个物理服务器存在多个虚拟机时,若虚拟化本身的开销较小,则可以认为多个虚拟机的运行率或运行量的和成为物理服务器的运行率或运行量。例如,当设想在物理服务器中存在共计 8 核的 CPU,向虚拟机 A 分配了 2 个核,向虚拟机 B 分配了 6 个核时,虚拟机 A 的 CPU 使用率的四分之一与虚拟机 B 的 CPU 使用率的四分之三的和相当于物理服务器的 CPU 使用率,如数学式 10、数学式 12 或数学式 13 那样表示装置功率函数。

[0151] 设备管理器 463 取得设备 440₁ ~ 440_u、450₁ ~ 450_v 的位置和环境信息,导出与物理服务器 420₁ ~ 420_n 的功率分布对应的设备功率的关系。运用管理管理器 461 根据与作业负荷对应的物理服务器功率的关系、以及与物理服务器群的功率分布对应的设备功率的关系,求出与物理服务器对应的作业负荷分配的最佳解。按照该解,虚拟化管理器 462 将运

行应用 $432_1 \sim 432_m$ 的虚拟机 $430_1 \sim 430_m$ 向物理服务器 $420_1 \sim 420_n$ 分配。在图 8 中表示如箭头 433 所示那样将应用 432_2 与虚拟机 430_2 一起从物理服务器 420_2 向 420_1 移动的一例。

[0152] 根据实施方式 6, 通过利用虚拟化环境的动态迁移功能, 可以使应用 $432_1 \sim 432_m$ 在运行的状态下在物理服务器 $420_1 \sim 420_n$ 之间转移, 可以简便地执行使服务器群和设备的总功率最小化的作业负荷分配, 可以抑制伴随作业负荷的移动的开销或损失时间来实现信息处理系统的节电化。实施方式 6 是服务器装置的例子, 但在存储装置或网络装置的虚拟化环境中也可以实现同样的节电运用。此外, 在实施方式 6 中采用了运用管理管理器 461 和虚拟化管理器 462 和设备管理器 463 相互协同工作的方法, 但也可以根据接口或软件规模等将一方安装在另一方中。

[0153] (实施方式 7)

[0154] 图 9 是表示本发明的实施方式 7 的信息处理系统的运用管理装置以及运用管理方法的结构图。信息处理系统 510 由服务器 $520_1 \sim 520_n$ 、存储系统 $570_1 \sim 570_r$ 、进行它们的供电和冷却的供电设备 $540_1 \sim 540_u$ 和冷却设备 $550_1 \sim 550_v$ 、管理这些装置群和设备全部的运用管理装置 560 构成, 构成 SAN(存储区域网络)环境。在服务器 $520_1 \sim 520_n$ 中, 在 OS $521_1 \sim 521_n$ 上运行作为作业负荷的营业 $530_1 \sim 530_m$ 。在运用管理装置 560 中, 管理由服务器群 $520_1 \sim 520_n$ 和存储系统 $570_1 \sim 570_r$ 构成的 SAN 的 SAN 管理器 562; 管理设备 $540_1 \sim 540_u$ 、 $550_1 \sim 550_v$ 的设备管理器 563; 削减服务器群 $520_1 \sim 520_n$ 和存储系统 $570_1 \sim 570_r$ 和设备 $540_1 \sim 540_u$ 、 $550_1 \sim 550_v$ 的总消耗功率的运用管理管理器 561 协同工作。

[0155] 在实施方式 7 中, SAN 管理器 562 进行服务器 $520_1 \sim 520_n$ 的运行监视和管理, 运用管理管理器 561 评价与应用即作业负荷 $530_1 \sim 530_m$ 对应的服务器 $520_1 \sim 520_n$ 的功率。设备管理器 563 根据设备 $540_1 \sim 540_u$ 、 $550_1 \sim 550_v$ 的位置和环境信息, 求出与服务器 $520_1 \sim 520_n$ 的功率分布对应的设备功率。运用管理管理器 561 根据服务器功率分布与设备功率的关系, 为使服务器群和设备的总功率最小化而求出作业负荷分配。根据该解, SAN 管理器 562 临时中断现行的应用 $530_1 \sim 530_m$, 向存储系统 $570_1 \sim 570_r$ 提交, 通过 SAN 启动来启动已分配的服务器 $520_1 \sim 520_n$, 继续进行应用 $530_1 \sim 530_m$ 。服务器 $520_1 \sim 520_n$ 中未分配的服务器停止, 切断电源。此外, 在作业负荷分配的执行中, 充分考虑应用 $530_1 \sim 530_m$ 的启动/中断/停止契机。

[0156] 根据实施方式 7, 通过利用 SAN 启动功能可以实施使服务器群和设备的总功率最小化的节电运用。具有在实施方式 6 中的虚拟化环境不适当的情况下或在已有的系统中易于引入本实施方式的优点。另外, 通过不仅结合服务器群和设备, 还结合存储系统 $570_1 \sim 570_r$ 来运用, 进一步实现了节能化。

[0157] (实施方式 8)

[0158] 图 10 是表示本发明的实施方式 8 的信息处理系统的运用管理装置以及运用管理方法的结构图。信息处理系统 610 由服务器 $620_1 \sim 620_n$ 、存储系统 $670_1 \sim 670_r$ 、网络 $680_1 \sim 680_s$ 、进行它们的供电和冷却的供电设备 $640_1 \sim 640_u$ 和冷却设备 $650_1 \sim 650_v$ 、综合管理装置群和设备的运用管理装置 660 构成, 构成综合管理环境。在服务器 $620_1 \sim 620_n$ 中, 在 OS $621_1 \sim 621_n$ 上运行作为作业负荷的应用 $630_1 \sim 630_m$ 和综合管理代理 $622_1 \sim 622_n$ 。在运用管理装置 660 中, 综合管理服务器 $620_1 \sim 620_n$ 和存储系统 $670_1 \sim 670_r$ 和网络 $680_1 \sim 680_s$

的管理器 662 ; 管理设备 $640_1 \sim 640_u$ 、 $650_1 \sim 650_v$ 的设备管理器 663 ; 削减装置群 $620_1 \sim 620_n$ 、 $670_1 \sim 670_r$ 、 $680_1 \sim 680_s$ 和设备 $640_1 \sim 640_u$ 、 $650_1 \sim 650_v$ 的总功率的功率管理器 661 相互协同工作。

[0159] 在实施方式 8 中, 综合管理管理器 662 通过代理 $662_1 \sim 662_n$ 进行服务器 $620_1 \sim 620_n$ 的运行管理和业务运用, 功率管理器 661 估计与应用 $630_1 \sim 630_m$ 对应的服务器 $620_1 \sim 620_n$ 的功率, 设备管理器 663 估计设备 $640_1 \sim 640_u$ 、 $650_1 \sim 650_v$ 的功率。功率管理器 661 与设备管理器 663 协同, 求出作业负荷分配, 根据该作业负荷分配, 综合管理管理器 662 进行服务器 $620_1 \sim 620_n$ 的运行 / 停止, 切换使应用 $630_1 \sim 630_m$ 工作的服务器。在作业调度中, 通过对服务器 $620_1 \sim 620_n$ 投入作业 (应用) $630_1 \sim 630_m$ 的配置可以进行作业负荷分配。

[0160] 关于存储系统 $670_1 \sim 670_r$ 、网络 $680_1 \sim 680_s$, 也可以与服务器 $620_1 \sim 620_n$ 的运用管理同样地进行。图 11 是表示存储系统的一例的说明图, 图 12 是表示网络的一例的说明图。图 11 中, 存储装置 671_1 、 671_2 由磁盘阵列 672 构成, 保存数据 $673_1 \sim 673_3$ 。如箭头 674 所示, 通过将作为作业负荷的访问频率高的数据 673_2 从存储装置 671_2 向 671_1 移动或拷贝, 可以进行希望的作业负荷分配。在图 12 中, 在网络装置 $681_1 \sim 681_3$ 上连接服务器装置 $623_1 \sim 623_4$ 等, 如箭头 682 所示那样, 将与服务器装置 623_3 对应的连接从网络装置 681_3 切换到 681_2 , 由此移动作业负荷。在变更作业负荷分配后, 存储装置 671_1 、 671_2 或网络装置 $681_1 \sim 681_3$ 使自身的节电功能起作用, 控制电源。此外, 在作业负荷的移动中, 需要与存储装置、网络一起充分考虑处理时间或与处理相伴的功率等的开销或风险。

[0161] 根据实施方式 8, 可以通过运行管理或作业调度来综合管理服务器 / 存储系统 / 网络装置群, 并且可以使全部装置群和设备的总消耗功率最小化。通过将综合管理管理器 662 具有的策略运用、资产管理、故障监视、电源管理、综合控制台等功能和功率管理器 661 的节电功能组合, 可以实现更高级且自由的系统运用。此外, 功率管理器 661 或设备管理器 663 也可以作为综合管理管理器 662 的一部分或子系统来进行安装。

[0162] 下面, 列举通过以上本实施方式的信息处理系统的运用管理装置以及运用管理方法而得到的主要的效果。

[0163] 通过上述各实施方式的手段, 通过使信息处理装置群和设备协同工作的运用管理, 可以削减将装置群和设备合在一起的总消耗功率, 因此, 具有可以实现数据中心等信息处理系统的节电运用、可以在降低运用成本的同时抑制过度的设备投资的效果, 也可以有助于地球变暖的防止和二氧化碳的削减。

[0164] 此外, 上述各实施方式的手段着眼于装置群和设备的功率, 但通过取得与装置群的作业负荷、运行信息、设备的环境信息相关的时间平均或时间积分等, 可以容易地进行考虑了作业负荷的处理时间的最佳化, 也可以削减装置群和设备的电能 (功率 $\times$ 时间 = 能量)。另外, 在向装置群的作业负荷分配中, 除了功率或电能以外, 通过也考虑作业负荷的处理优先度、基于计划或事件等的执行契机、与送电系统或收费制度对应的电力成本等, 可以作为进一步提高有益性和便利性的策略运用来进行应用。

[0165] 本实施方式, 作为信息处理系统自身的自主的节电运用管理方法是适合的, 但根据本实施方式而得到的作业负荷分配的解、或装置群和设备的个别功率或总功率的结果, 在已设系统中的节能诊断和改善、新设系统的设计或建设、装置或设备的增设、移设、修改、

更新中的计划或调整、将来的信息处理预测所对应的风险把握或成本计算、取得的运行管理或场景设定中,也提供了有用的运用管理方针。

[0166] 以上,根据实施方式,具体说明了由本发明人作出的发明,但本发明不限于所述实施方式,在不脱离其主旨的范围内可以进行各种变更。

[0167] 产业上的可利用性

[0168] 本实施方式可以应用于由服务器、存储系统、网络等的信息处理装置群和供电设备或冷却设备构成的信息处理系统中的综合的节电运用管理,尤其适用于作为信息处理基础的数据中心。另外,本实施方式除了用于信息处理系统的自主运用管理以外,也可以作为系统构建工具、节能诊断工具、运行监视工具、运用管理者或设备管理者的辅助工具而适用于广泛的用途。

[0169] 而且,本实施方式不仅针对于配置在一个场所的装置群和设备,对于多个场所也可以应用,例如在多层的数据中心或大范围分离的数据中心中也有用,像网格计算或云计算等信息处理装置和设备分散的情况下也可以利用。另外,本实施方式主要以信息处理装置群和设备为对象,但只要是消耗功率或能量的装置和用于其的设备,就可以应用本实施方式,例如对电气装置、机械装置、动力装置、热装置等的运用或控制也发挥作用,作为系统规模,电子部件群和电源或冷却器、信息模块群和电源单元或冷却单元、数据中心群和发电站或包含布局在内的冷却环境等也作为对象,可以对省电、节能作出贡献。

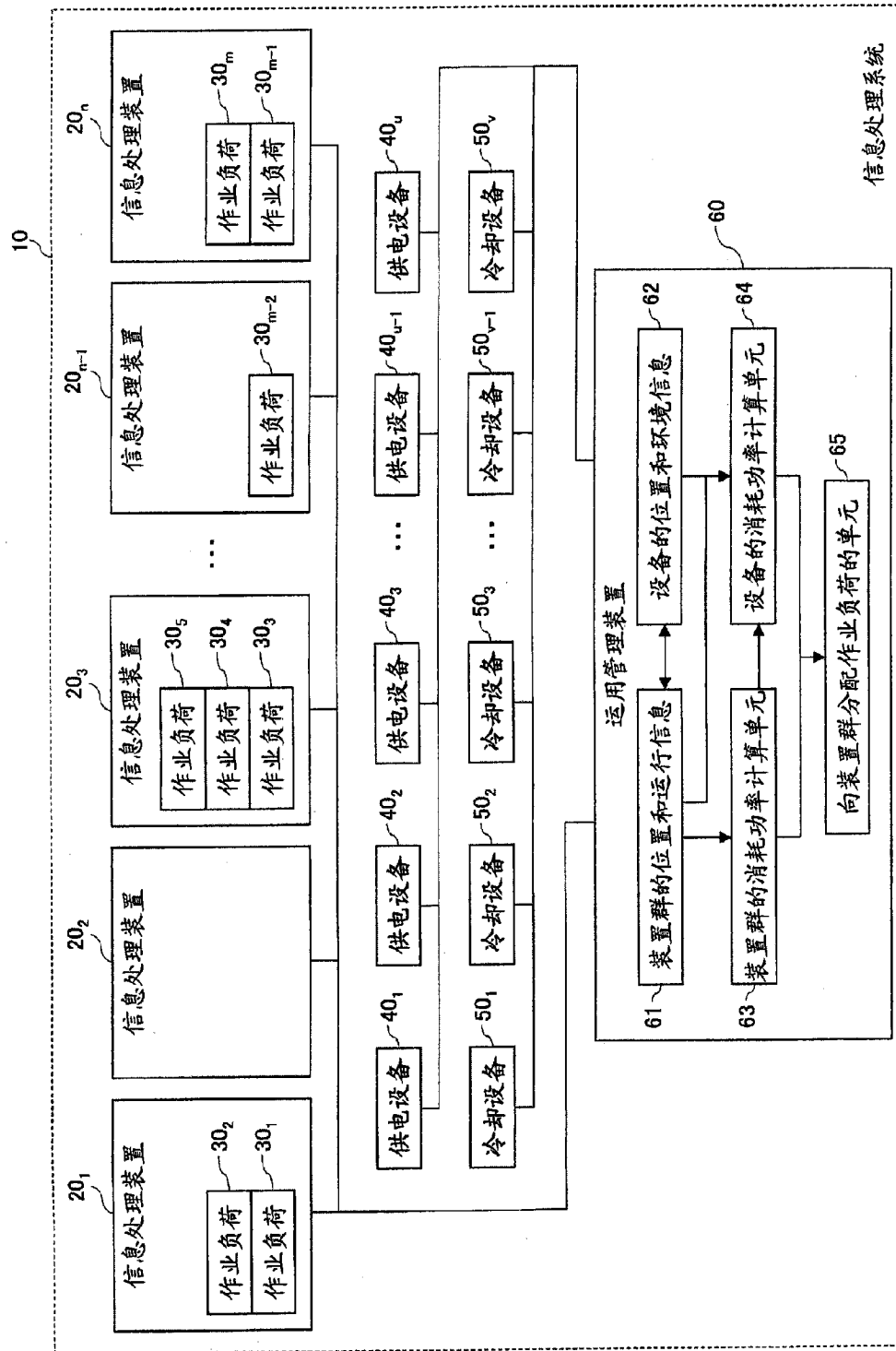


图 1

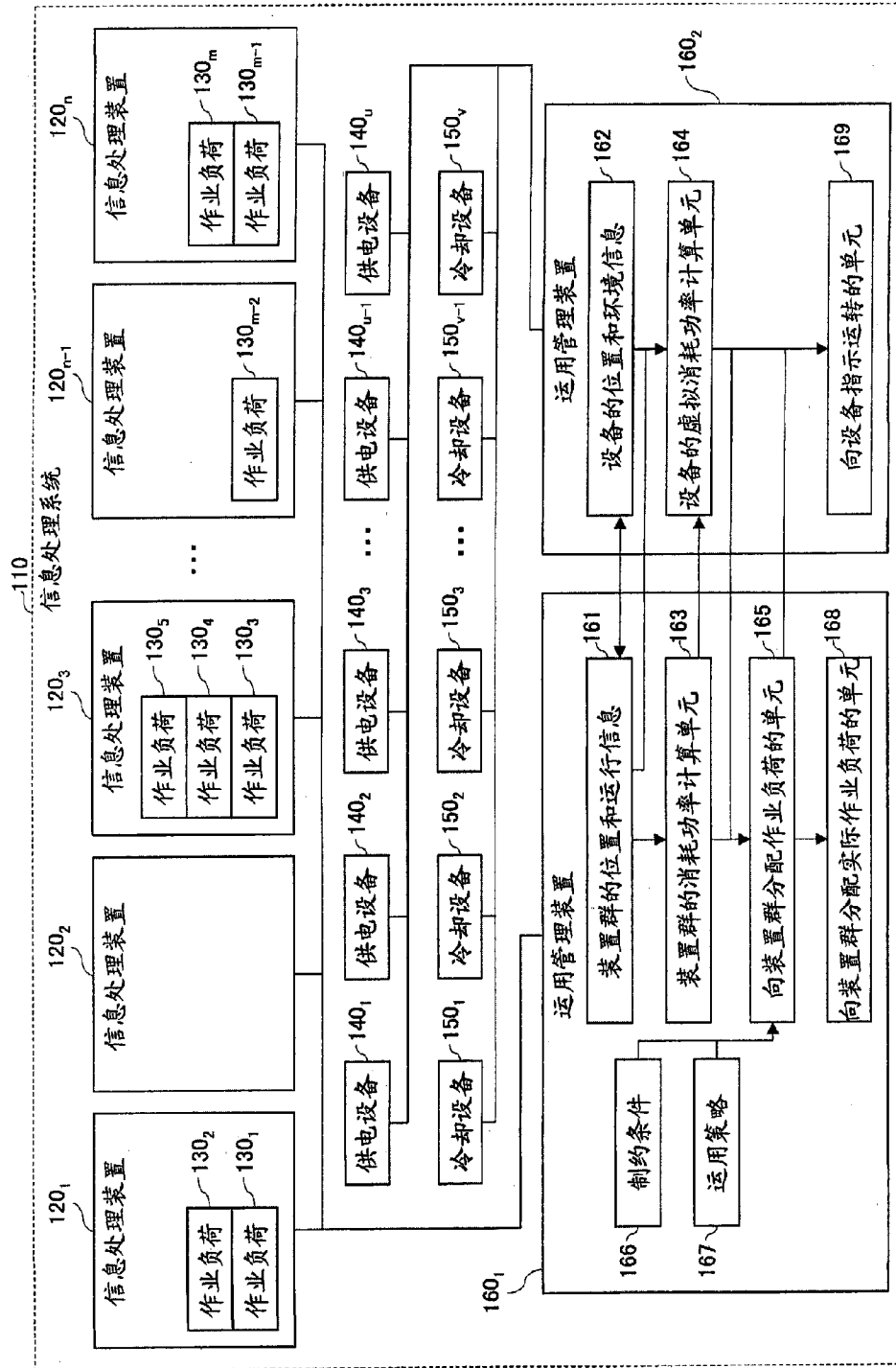
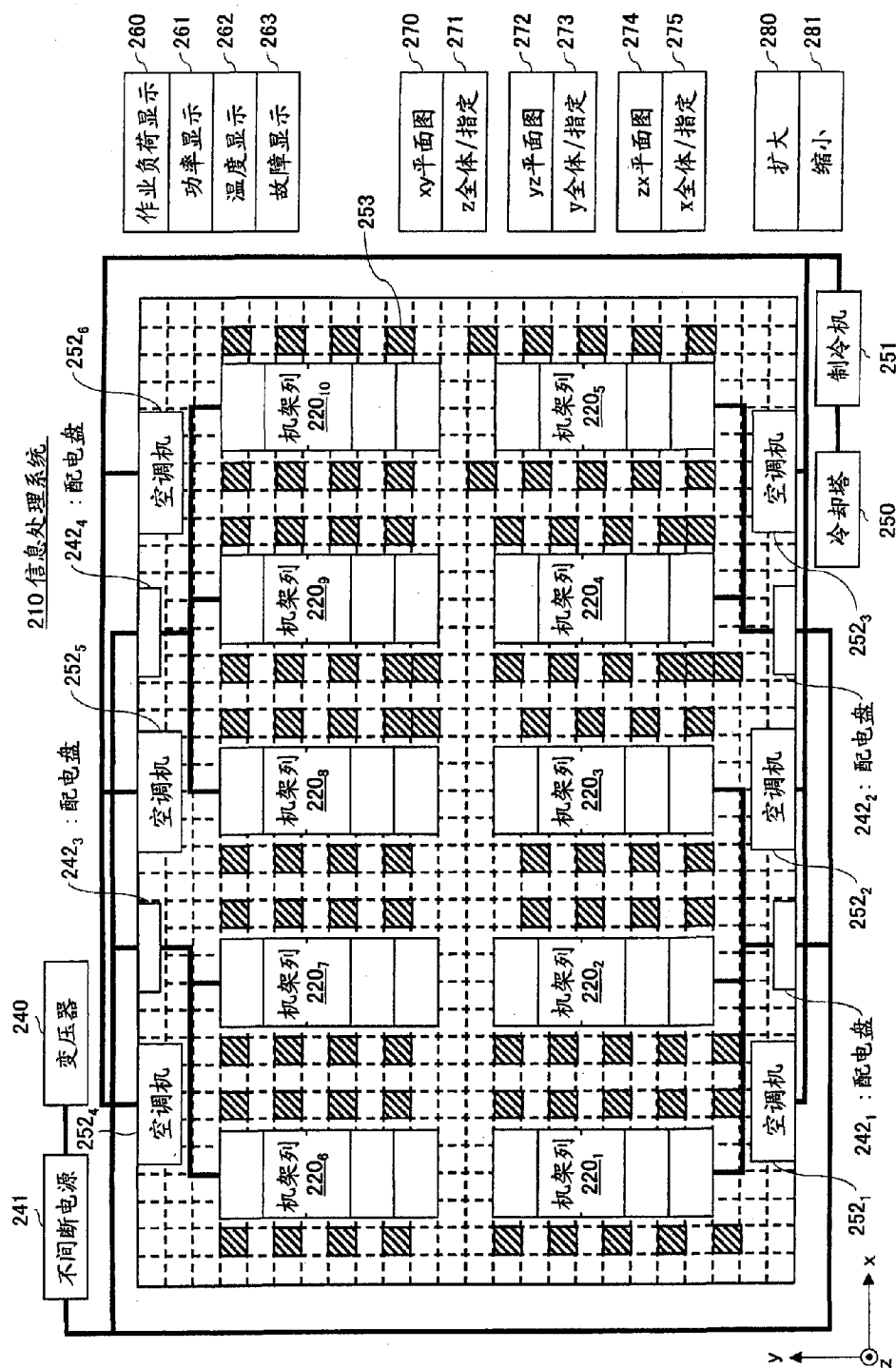


图 2



3

规格信息300											
装置管理号码	装置管理者	装置名称	装置品名	装置型号	装置性能a	装置性能b	装置性能c	额定功率最大	额定功率待机	额定温度最大	额定温度最小
					γ_a	γ_b	γ_c	P_{max}	P_{idle}		
结构信息310											
CPU型号	CPU最大频率 f_{max}	CPU搭载数	存储器型号	存储器容量	存储器搭载数	芯片组型号	磁盘型号	磁盘容量	磁盘搭载数	I/O传输速度	I/O个数
运行信息(或测定信息、运行历史)320											
装置位置坐标x, y, z	作业负荷类型w	装置/CPU状态	CPU工作频率f	CPU使用率 α	CPU工作温度	存储器使用率	磁盘传输速度	I/O传输速度	装置工作功率 P_{meas}	装置进气温度 T_{in}	装置排气温度 T_{out}

图 4

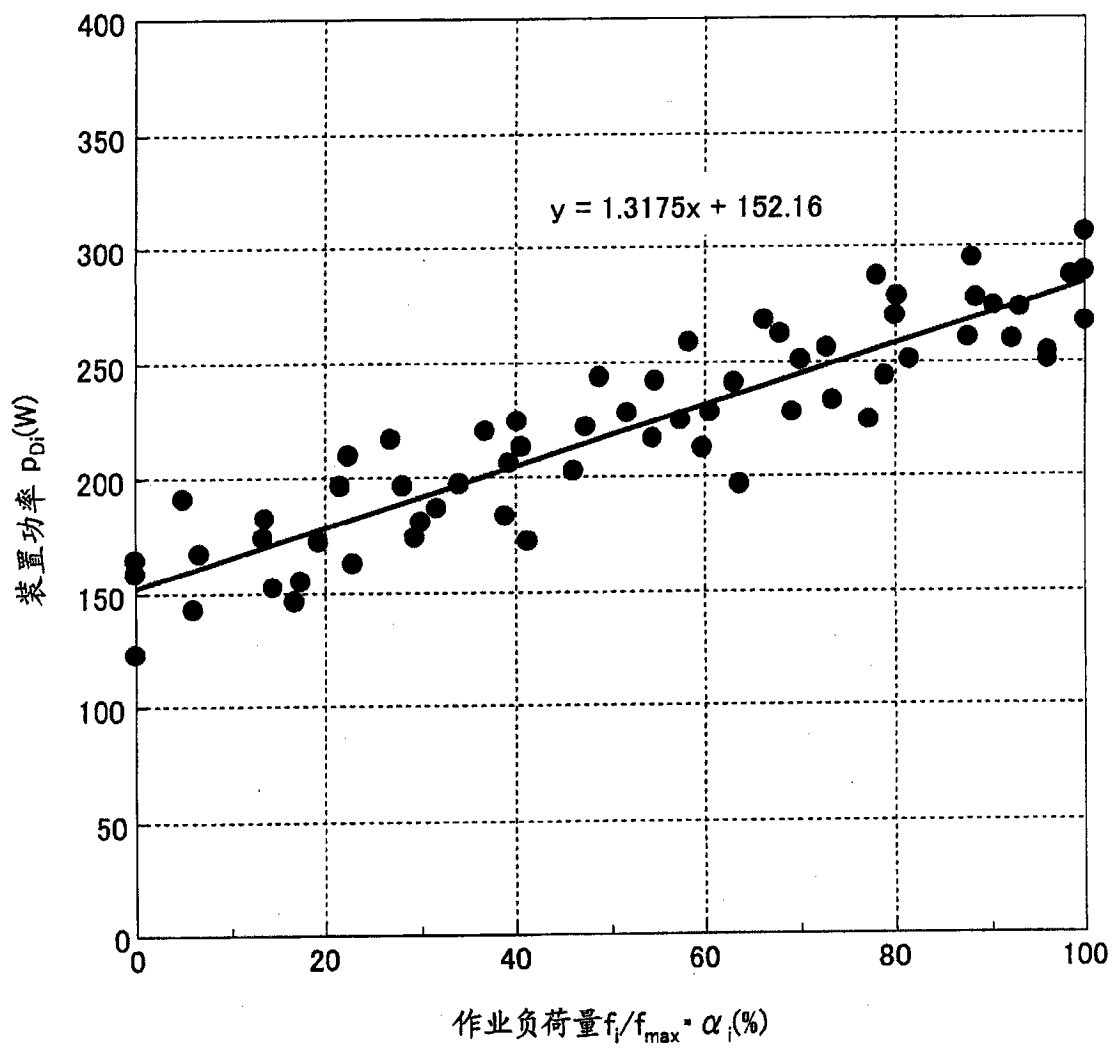


图 5

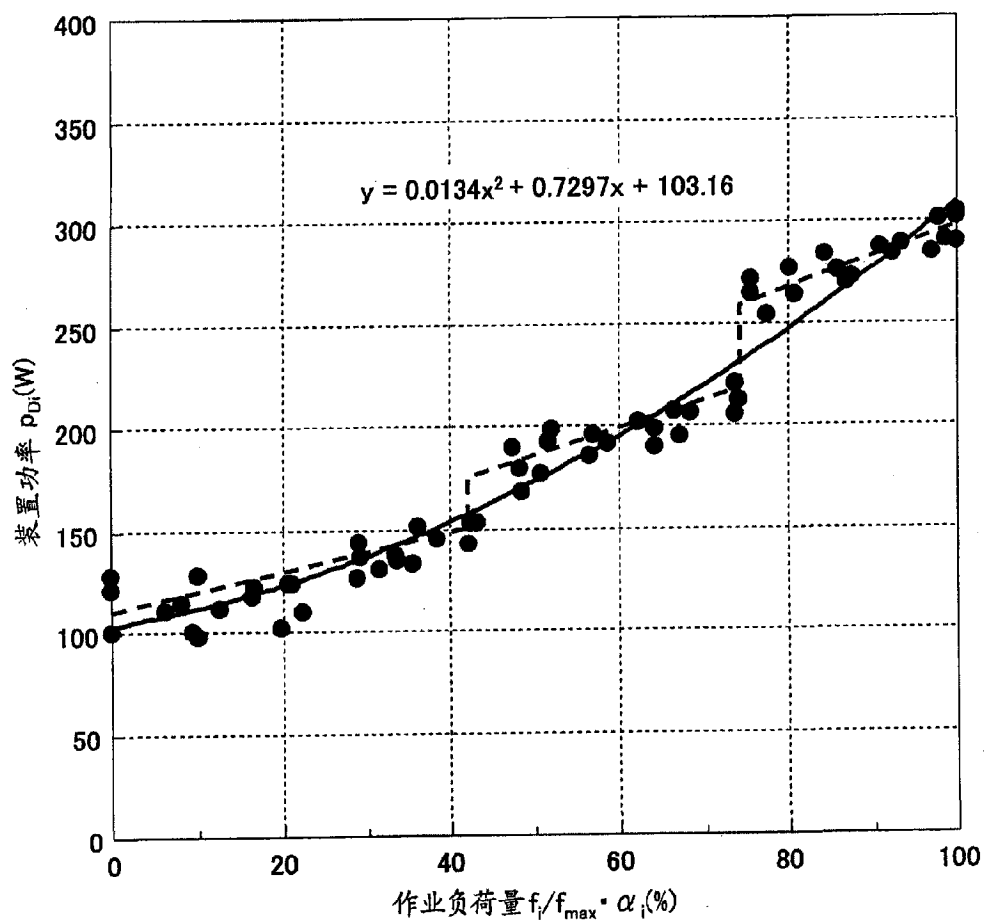


图 6

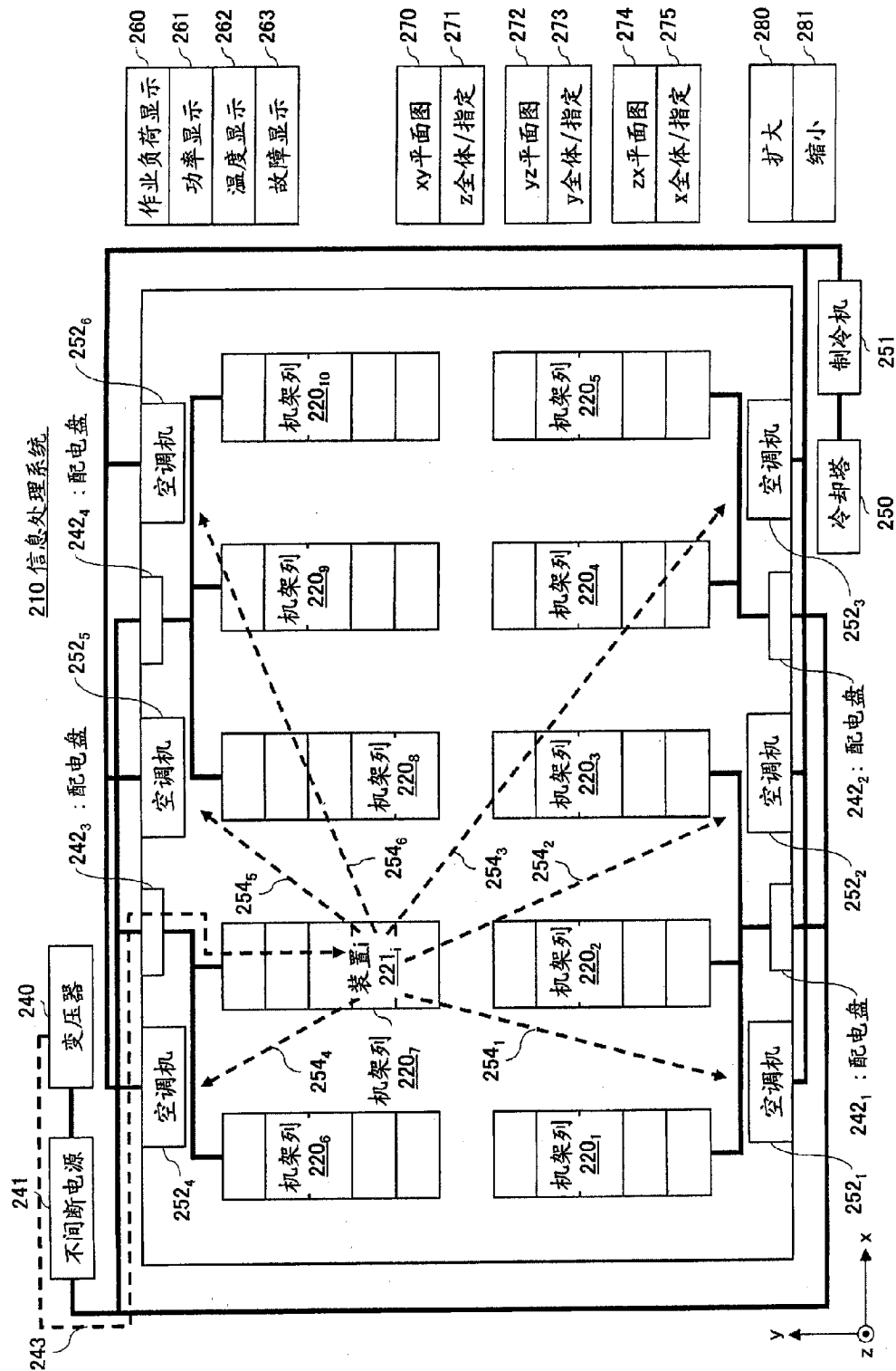


图 7

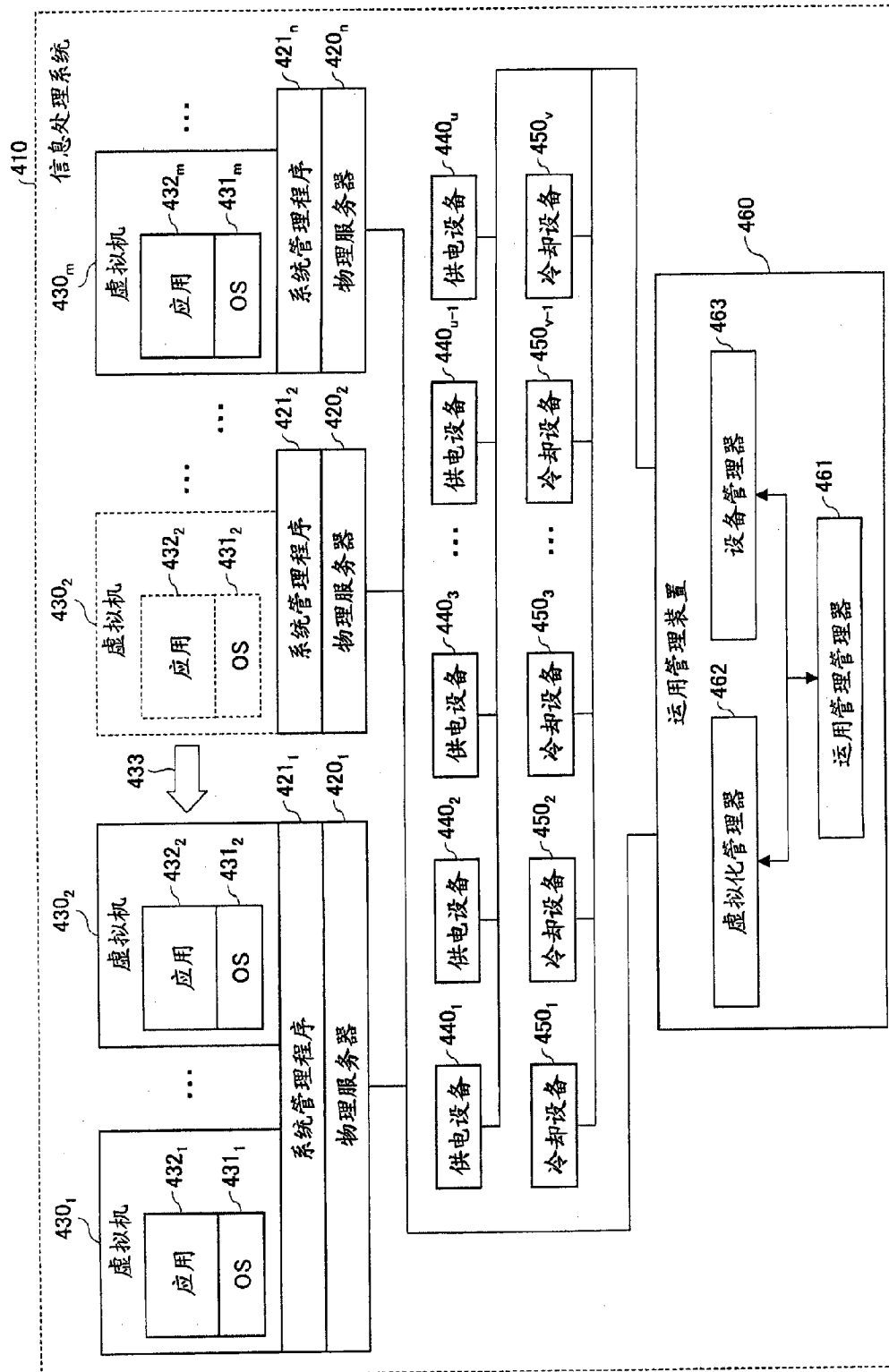
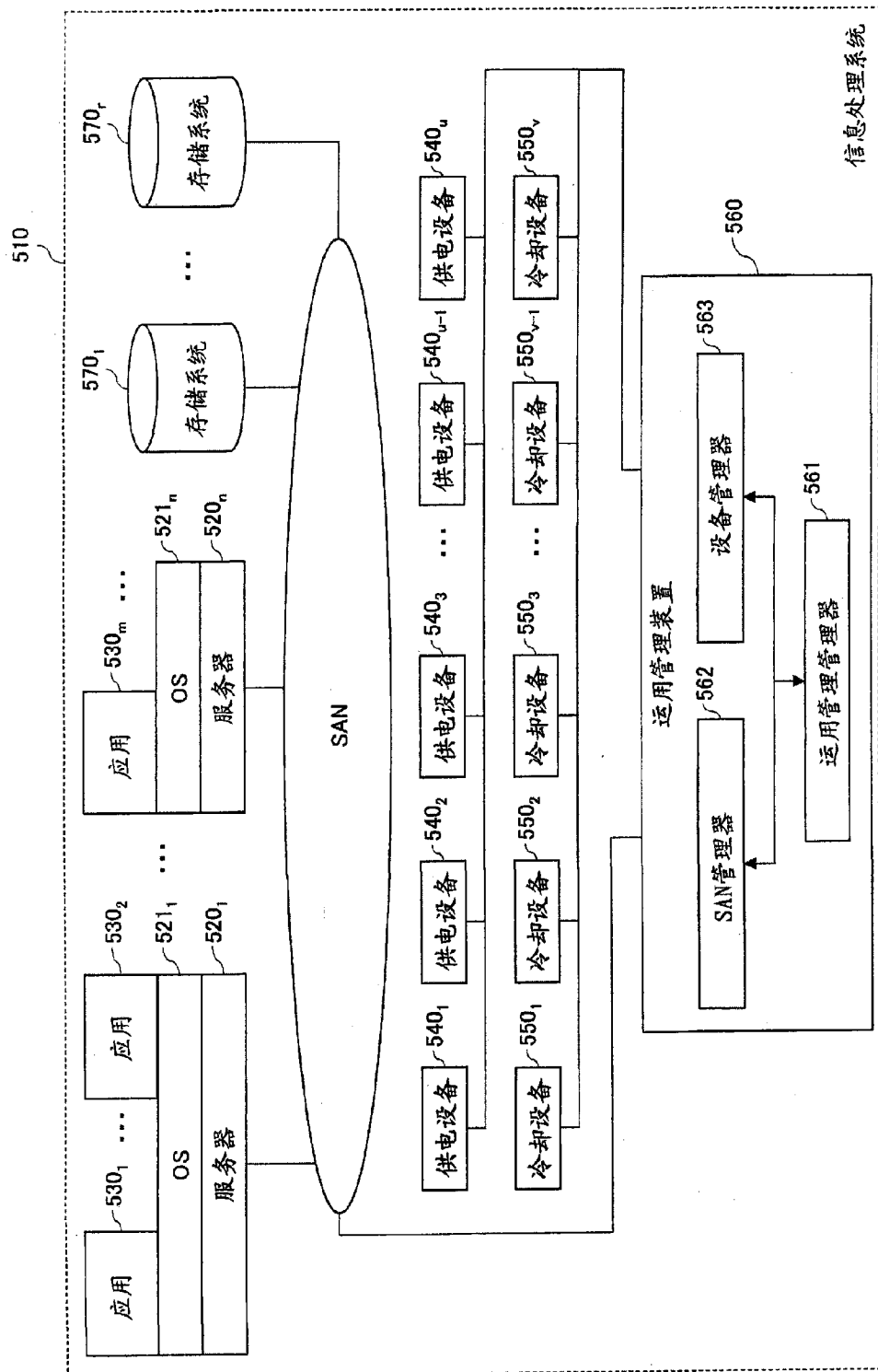
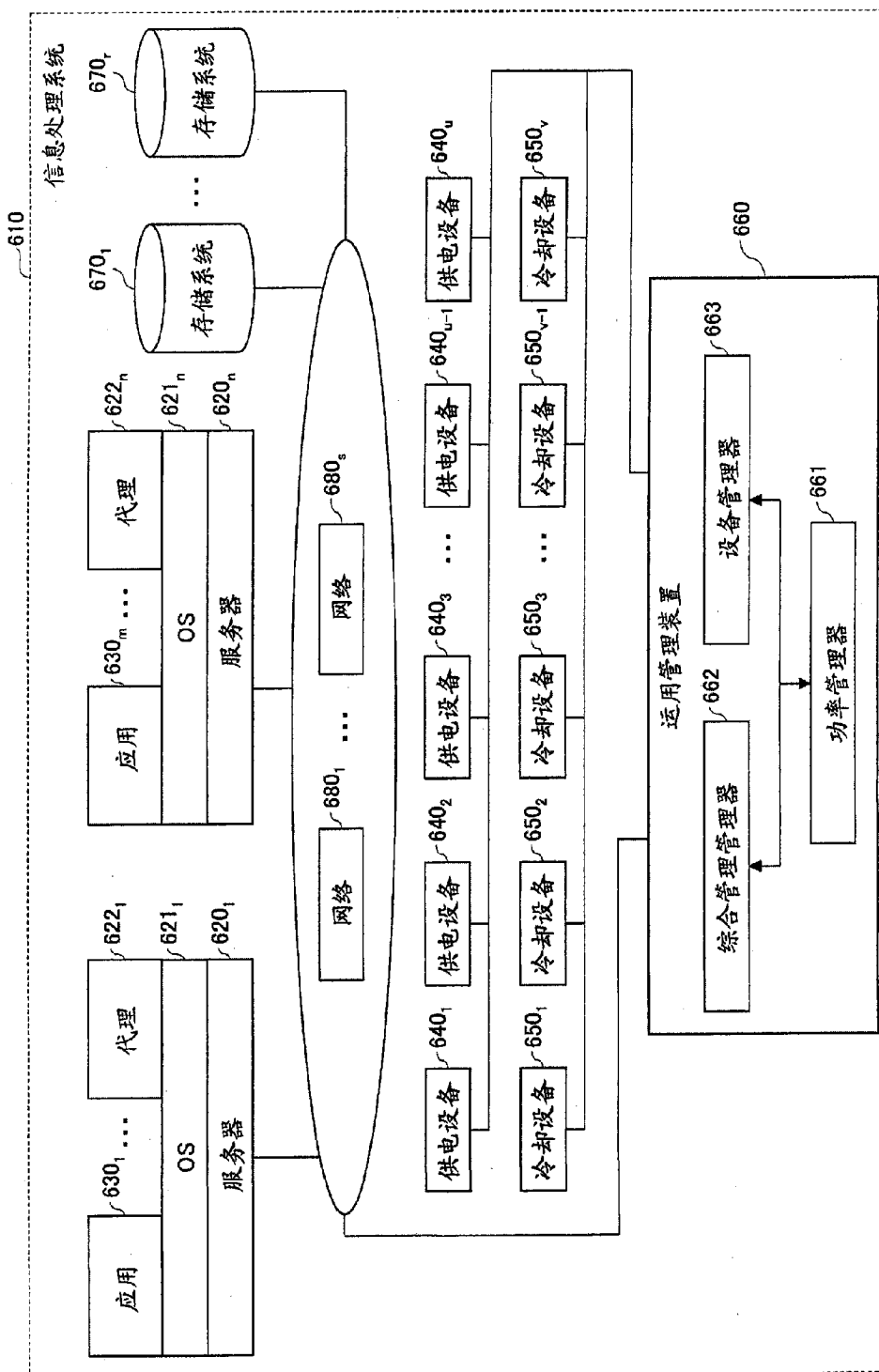


图 8



9



10

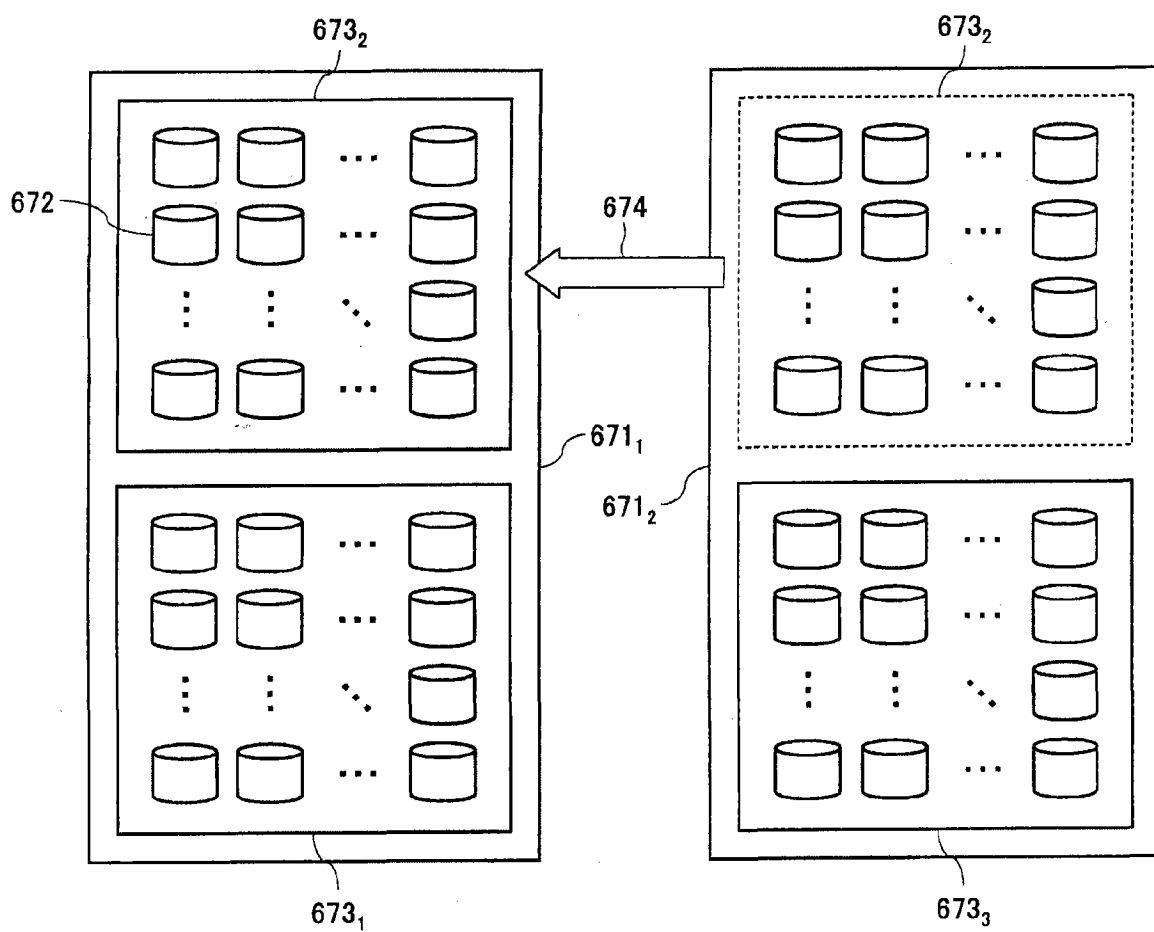


图 11

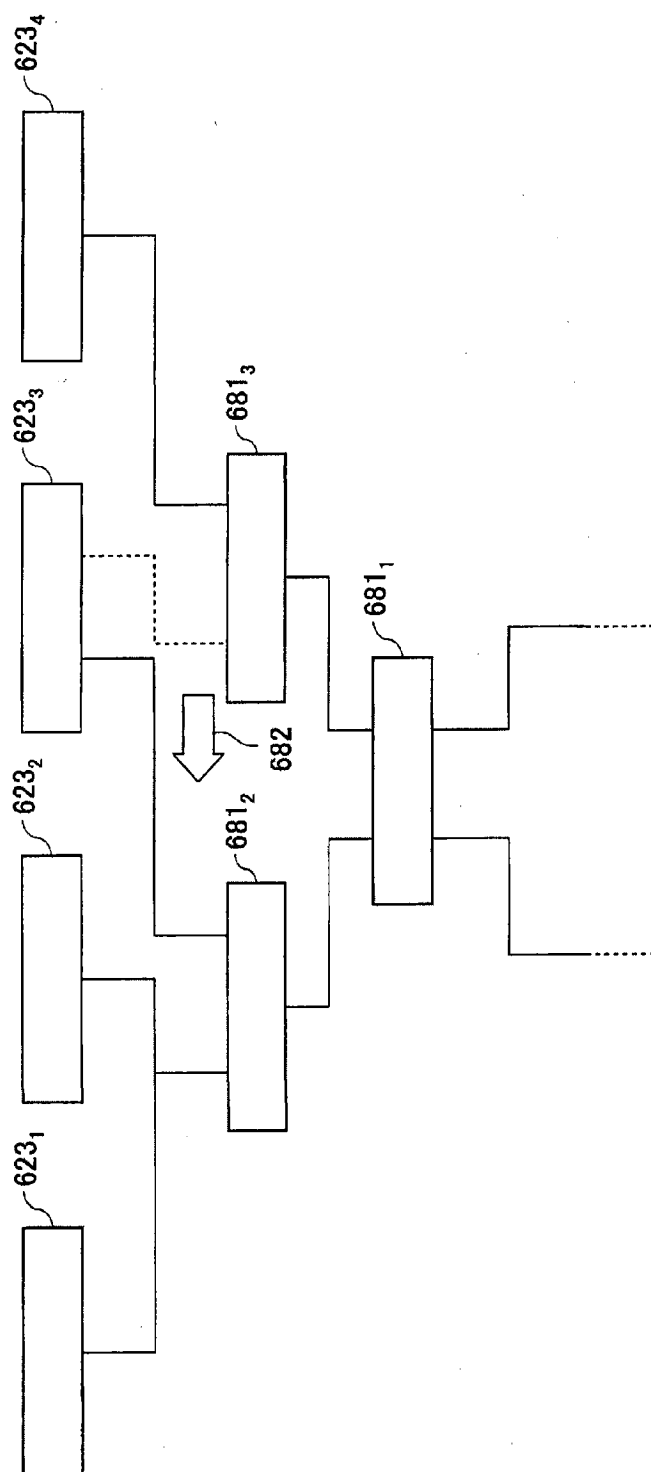


图 12