



(12)发明专利申请

(10)申请公布号 CN 108399102 A

(43)申请公布日 2018.08.14

(21)申请号 201710069369.1

(22)申请日 2017.02.08

(71)申请人 阿里巴巴集团控股有限公司

地址 英属开曼群岛大开曼资本大厦一座四
层847号邮箱

(72)发明人 张广舟 范孝剑

(74)专利代理机构 北京安信方达知识产权代理
有限公司 11262

代理人 龙洪 栗若木

(51) Int. Cl.

G06F 9/50(2006.01)

权利要求书2页 说明书5页 附图2页

(54)发明名称

一种资源分配方法和装置

(57)摘要

一种资源分配方法和装置,用于为同一资源分配单位内共享资源的多个处理单元分配资源,一次资源分配过程包括:资源分配装置确定所述多个处理单元存储的数据量;根据所述多个处理单元存储的数据量为所述多个处理单元分配资源,存储的数据量越大的处理单元分配的资源越多。本申请可以使资源利用更加充分。

110. 确定所述多个处理单元存储的数据量

120. 根据所述多个处理单元存储的数据量为所述多个处理单元分配资源,存储的数据量越大的处理单元分配的资源越多

1. 一种资源分配方法,为同一资源分配单位内共享资源的多个处理单元进行资源的动态分配,其中一次资源分配过程包括:

确定所述多个处理单元存储的数据量;

根据所述多个处理单元存储的数据量为所述多个处理单元分配资源,存储的数据量越大的处理单元分配的资源越多。

2. 如权利要求1所述的方法,其特征在于:

所述分配的资源包括以下一种或多种:

内存;

处理器;

网络带宽;

磁盘传输带宽;

临时磁盘空间。

3. 如权利要求1所述的方法,其特征在于:

根据所述多个处理单元存储的数据量为所述多个处理单元分配资源,包括:为所述多个处理节点分配的资源与所述多个处理单元存储的数据量成正比。

4. 如权利要求1或2或3所述的方法,其特征在于:

所述为同一资源分配单位内共享资源的多个处理单元进行资源的动态分配,包括:

按照设定时间间隔,每隔一段时间启动一次资源分配过程;或者

监控到所述多个处理单元存储的数据量发生变化时,启动一次资源分配过程;或者

监控到所述多个处理单元存储的数据量的变化超过设定的阈值时,启动一次资源分配过程。

5. 如权利要求1或2或3所述的方法,其特征在于:

所述同一资源分配单位内的多个处理单元是分布式数据库系统位于同一物理机或虚拟机中的多个子节点。

6. 一种资源分配装置,包括资源分配模块,设置为:为同一资源分配单位内共享资源的多个处理单元进行资源的动态分配,所述资源分配模块包括用于完成一次资源分配过程的确定单元和分配单元,其中:

所述确定单元设置为:确定所述多个处理单元存储的数据量;

所述分配单元设置为:根据所述多个处理单元存储的数据量为所述多个处理单元分配资源,存储的数据量越大的处理单元分配的资源越多。

7. 如权利要求6所述的装置,其特征在于:

所述分配单元为所述多个处理单元分配的资源包括以下一种或多种:

内存;

处理器;

网络带宽;

磁盘传输带宽;

临时磁盘空间。

8. 如权利要求6所述的装置,其特征在于:

所述分配单元根据所述多个处理单元存储的数据量为所述多个处理单元分配资源,包

括;为所述多个处理节点分配的资源与所述多个处理单元存储的数据量成正比。

9. 如权利要求6或7或8所述的装置,其特征在于:

所述资源分配模块还包括:触发单元,设置为:

按照设定时间间隔,每隔一段时间触发所述确定单元和分配单元启动一次资源分配过程;或者

监控到所述多个处理单元存储的数据量发生变化时,触发所述确定单元和分配单元启动一次资源分配过程;或者

监控到所述多个处理单元存储的数据量的变化超过设定的阈值时,触发所述确定单元和分配单元启动一次资源分配过程。

10. 如权利要求6或7或8所述的装置,其特征在于:

所述同一资源分配单位内的多个处理单元是分布式数据库系统位于同一物理机或虚拟机中的多个子节点。

11. 一种资源分配装置,包括存储器和处理器,其特征在于:

所述存储器设置为:保存程序代码;

所述处理器设置为:读取所述程序代码,执行以下处理:

为同一资源分配单位内共享资源的多个处理单元进行资源的动态分配,其中,一次资源分配过程包括:确定所述多个处理单元存储的数据量;及,根据所述多个处理单元存储的数据量为所述多个处理单元分配资源,存储的数据量越大的处理单元分配的资源越多。

12. 如权利要求11所述的装置,其特征在于:

所述处理器分配的资源包括以下一种或多种:

内存;

处理器;

网络带宽;

磁盘传输带宽;

临时磁盘空间。

一种资源分配方法和装置

技术领域

[0001] 本发明涉及计算机技术,更具体地,涉及一种资源分配方法和装置。

背景技术

[0002] 在包括多个处理节点的系统中,资源分配是非常常见的一种操作。相关技术的资源分配多是对这些处理节点分配相同的资源,或者根据这些处理节点要承担的业务分配不同的资源。但这些资源分配方法在一定程度上并不能够充分有效地利用资源。

[0003] 以分布式数据库系统为例,其架构如图1所示,其处理节点包括主节点与子节点组成。主节点只存放元数据。子节点负责实际数据存储和计算。一个子节点一般是独立启动的一个或多个操作系统进程,分配拥有的磁盘存储空间、磁盘传输带宽、内存、处理器、网络带宽等资源。

[0004] 多个子节点可能一起运行在同一机器上,该多个子节点共享该机器的资源。如果该机器上还运行有其他应用程序,则需要该多个节点的共享资源进行限制即做资源隔离。在为该多个子节点分配共享资源时,一般的做法是,对每个子节点单独进行资源分配且分配固定的资源。例如,每个子节点分配8G内存、2Mbps网络带宽等。

[0005] 本发明的发明人研究发现,子节点在运行时所需的资源实际上和其存储的数据量是有关系的,数据量越大时对其他资源的需求也越大。而采用现有方法进行资源分配,由于资源固定,会造成数据在各个子节点分布不均时,发生数据量较大的子节点资源不够,而其他子节点的资源未充分利用的情况。以内存为例,某个节点的数据量是60G,而其他节点是40G,有60G数据的节点所需内存实际上更多,但却和其他节点分配了同样的内存;而有40G数据的节点所需内存较少,分配的内存可能未能充分使用。

[0006] 这里处理节点所需资源与存储数据量之间的关系,在其他的系统如查询系统、业务系统等中也是存在的。在这些系统中,相关技术的资源分配方法同样也不能做到对资源的充分利用。

发明内容

[0007] 本发明实施例提供了一种资源分配方法,用于为同一资源分配单位内共享资源的多个处理单元分配资源,一次资源分配过程包括:

[0008] 确定所述多个处理单元存储的数据量;

[0009] 根据所述多个处理单元存储的数据量为所述多个处理单元分配资源,存储的数据量越大的处理单元分配的资源越多。

[0010] 本发明实施例还提供了一种资源分配装置,包括为同一资源分配单位内共享资源的多个处理单元分配资源的资源分配模块,所述资源分配模块包括:

[0011] 确定单元,设置为:确定所述多个处理单元存储的数据量;

[0012] 分配单元,设置为:根据所述多个处理单元存储的数据量为所述多个处理单元分配资源,存储的数据量越大的处理单元分配的资源越多。

- [0013] 本发明实施例还提供了一种资源分配装置,包括存储器和处理器,其中:
- [0014] 所述存储器设置为:保存程序代码;
- [0015] 所述处理器设置为:读取所述程序代码,执行以下处理:
- [0016] 为同一资源分配单位内共享资源的多个处理单元分配资源,一次资源分配过程包括:确定所述多个处理单元存储的数据量;及,根据所述多个处理单元存储的数据量为所述多个处理单元分配资源,存储的数据量越大的处理单元分配的资源越多。
- [0017] 上述资源分配方案可以使资源利用更加充分。

附图说明

- [0018] 图1是分布式数据库系统的架构图;
- [0019] 图2是本发明实施例一资源分配方法的流程图;
- [0020] 图3是本发明实施例一中作为示例的分布式数据库系统子节点分组的示意图;
- [0021] 图4是本发明实施例二资源分配装置的模块图。

具体实施方式

[0022] 为使本发明的目的、技术方案和优点更加清楚明白,下文中将结合附图对本发明的实施例进行详细说明。需要说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互任意组合。

[0023] 实施例一

[0024] 本实施例提供一种资源分配方法,用于为同一资源分配单位内共享资源的多个处理单元进行资源的动态分配。这里的资源分配单位是可以进行资源分配的硬件或软件实体,如一台物理机或虚拟机,一台机器内的一个资源组,也可以是统一分配多台机器资源的系统。这里的处理单元是可以存储数据和处理数据的任何单元。本实施例是以分布式数据库系统位于同一物理机或虚拟机中的多个子节点为示例进行说明。

[0025] 如图2所示,本实施例资源分配方法的一次资源分配过程包括:

[0026] 步骤110,确定所述多个处理单元存储的数据量;

[0027] 对于处理单元存储的数据量(指实际存储的数据量而不是分配的磁盘空间),可以通过操作系统监控为处理单元分配的磁盘空间中已存储的数据量获取。也可以由处理单元所属的系统对处理单元的写入、删除等带来的数据量的变化进行统计而得到。

[0028] 步骤120,根据所述多个处理单元存储的数据量为所述多个处理单元分配资源,存储的数据量越大的处理单元分配的资源越多。

[0029] 本实施例中,所述分配的资源包括以下一种或多种:内存、处理器、网络带宽、磁盘传输带宽和临时磁盘空间。这些资源中,可以只有部分资源根据存储的数据量来分配,其他资源按其他方式如按预定值分配的方式分配。

[0030] 本实施例中,为存储的数据量越大的处理单元分配越多资源可以有多种具体的方式,例如,为所述多个处理节点分配的资源与所述多个处理单元存储的数据量成正比,这里说的成正比并不要求精确到非常小的资源单位,可以按照设定的最小的资源分配单位调整,例如可分配的内存空间为10M时,如果两个处理节点即处理节点A和处理节点B存储的数据量的比例是1:2,内存以1M为最小单位进行分配,则可以给处理节点A分配3M内存,给处理

节点B分配7M内存。除了按正比分配外,还可以按照预设的存储数据量与内存容量的对应关系来分配,例如,可分配的磁盘空间为100G,内存容量为1G,如果处理节点存储的数据量为0~10G,则对应分配的内存容量为100M,如果存储的数据量为10~20G,则对应分配的内存容量为200M,依此类推。这里可以采用的算法是很多的,需要说明的是,这里说的存储的数据量越大的含义与采用的具体算法有关的,例如上面第2个示例中,如果一个处理节点存储的数据量为11G,另一个处理节点存储的数据量为13G,按照预设的存储数据量与内存容量的对应关系来分配时,两个处理节点存储的数据量在同一取值区间即10~20G,将视为相等而均分配200M的内存。但在总体上,需要为存储的数据量越大的处理单元分配越多资源,以取得充分利用资源的效果。

[0031] 本实施例中,为多个处理单元分配的资源可以是所述多个处理节点所在资源分配单位所有的资源,也可以是部分资源。例如,为分布式数据库系统位于同一物理机或虚拟机中的多个子节点分配资源时,可以将物理机或虚拟机拥有的资源留出一部分给其他应用使用,而只给多个子节点分配其中的部分资源,即做资源隔离。

[0032] 本实施例中,可以采用但不局限于以下方式实现动态分配:

[0033] 按照设定时间间隔,每隔一段时间启动一次资源分配过程;或者

[0034] 监控到所述多个处理单元存储的数据量发生变化时,启动一次资源分配过程;或者

[0035] 监控到所述多个处理单元存储的数据量的变化超过设定的阈值时,启动一次资源分配过程。

[0036] 资源分配的具体方式可以采用与相关技术相同的方式,这里不再赘述。

[0037] 下面用一个示例进行说明。

[0038] 如图3所示,本示例涉及的是分布式数据库系统,该分布式数据库系统包括主节点和多个子节点。先对所有子节点进行分组,将位于同一物理机或虚拟机中的多个子节点分为一组,这里假定均是每两个子节点一组。

[0039] 每个组分配固定的共享资源,本示例中,为组内每一子节点分配一种或多种资源时,根据该子节点存储的数据量占组内所有子节点存储的总数据量的比例进行动态分配(如每5分钟调整一次)。例如,某一组内包括两个节点即节点1与节点2分,则某一种资源的分配公式为:

[0040] 为节点1分配的资源=组内可分配的资源*(节点1存储的数据量/(节点1存储的数据量+节点2存储的数据量))

[0041] 以内存为例,假定组内可分配的内存为16G,节点1存储的数据量为40G,节点2存储的数据量为60G,则:

[0042] 为节点1分配的内存=16G*(40/(40+60))=6.4G

[0043] 为节点2分配的内存=16G*(60/(40+60))=9.6G

[0044] 有些资源分配时,按上式计算得到的数值需要根据资源的最小分配单位调整。例如处理器资源用内核数表示时,假定可分配的内核数为8,则节点1可分配3个内核,节点2可分配5个内核。

[0045] 其他资源也可以参考此公式,但并不是所有资源都要采取这种方式。

[0046] 下面再对“临时磁盘空间”这一资源的分配说明如下:

[0047] 子节点运行查询时,如果涉及大数据量排序等操作,会需要大量的临时磁盘空间来存放排序的中间结果。特别地,本示例把“临时磁盘空间”作为一种资源进行动态分配(相关技术中没有对“临时磁盘空间”进行动态分配的算法,对“临时磁盘空间”的动态分配可以对子节点存储量变化时对存储空间的需求做出快速反应,使得子节点有足够资源执行相应处理)。

[0048] 假如每个组有两个节点,总的存储空间为160G,且节点1有40G数据量,节点2有60G数据量,则此时组内可分配的临时磁盘空间为: $160G - (40G + 60G) = 60G$,使用上述算法,可得到:

[0049] 为节点1分配的临时磁盘空间 $= 60G * (40 / (40 + 60)) = 24G$

[0050] 为节点2分配的临时磁盘空间 $= 60G * (60 / (40 + 60)) = 36G$ 。

[0051] 通过上述分配,给数据较多的节点2分配了更多的临时磁盘空间,可以使原来可能因临时空间不够而无法完成的查询完成执行,充分利用了可用资源。

[0052] 实施例二

[0053] 本实施例提供一种资源分配装置,以实现实施例一的资源分配方法。本实施例的资源分配装置包括资源分配模块,设置为:为同一资源分配单位内共享资源的多个处理单元进行资源的动态分配,如图4所示,所述资源分配模块包括用于完成一次资源分配过程的确定单元10和分配单元29,其中:

[0054] 所述确定单元10设置为:确定所述多个处理单元存储的数据量;

[0055] 所述分配单元20设置为:根据所述多个处理单元存储的数据量为所述多个处理单元分配资源,存储的数据量越大的处理单元分配的资源越多。

[0056] 本实施例中,

[0057] 所述分配单元为所述多个处理单元分配的资源包括以下一种或多种:内存、处理器、网络带宽、磁盘传输带宽和临时磁盘空间。

[0058] 本实施例中,

[0059] 所述分配单元根据所述多个处理单元存储的数据量为所述多个处理单元分配资源,包括:为所述多个处理节点分配的资源与所述多个处理单元存储的数据量成正比。

[0060] 本实施例中,

[0061] 所述资源分配模块还包括:触发单元,设置为:

[0062] 按照设定时间间隔,每隔一段时间触发所述确定单元和分配单元启动一次资源分配过程;或者

[0063] 监控到所述多个处理单元存储的数据量发生变化时,触发所述确定单元和分配单元启动一次资源分配过程;或者

[0064] 监控到所述多个处理单元存储的数据量的变化超过设定的阈值时,触发所述确定单元和分配单元启动一次资源分配过程。

[0065] 本实施例中,

[0066] 所述同一资源分配单位内的多个处理单元是分布式数据库系统位于同一物理机或虚拟机中的多个子节点。

[0067] 上述资源分配模块功能的具体实现可参见实施例一方法中的相应说明。上述资源分配模块的各单元还可以设置为实现实施例一方法中的其他处理。

[0068] 本实施例还提供了一种资源分配装置,包括存储器和处理器,其中:

[0069] 所述存储器设置为:保存程序代码;

[0070] 所述处理器设置为:读取所述程序代码,执行以下处理:

[0071] 为同一资源分配单位内共享资源的多个处理单元进行资源的动态分配,其中,一次资源分配过程包括:确定所述多个处理单元存储的数据量;及,根据所述多个处理单元存储的数据量为所述多个处理单元分配资源,存储的数据量越大的处理单元分配的资源越多。

[0072] 上述处理器读取程序代码后可执行实施例一方法中的任何处理,这里不再一一赘述。

[0073] 上述本发明实施例序号仅仅为了描述,不代表实施例的优劣。通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本发明实施例的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,计算机,服务器,或者网络设备等)执行本发明各个实施例所述的方法。

[0074] 以上所述仅为本发明的优选实施例而已,并不用于限制本发明,对于本领域的技术人员来说,本发明可以有各种更改和变化。凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

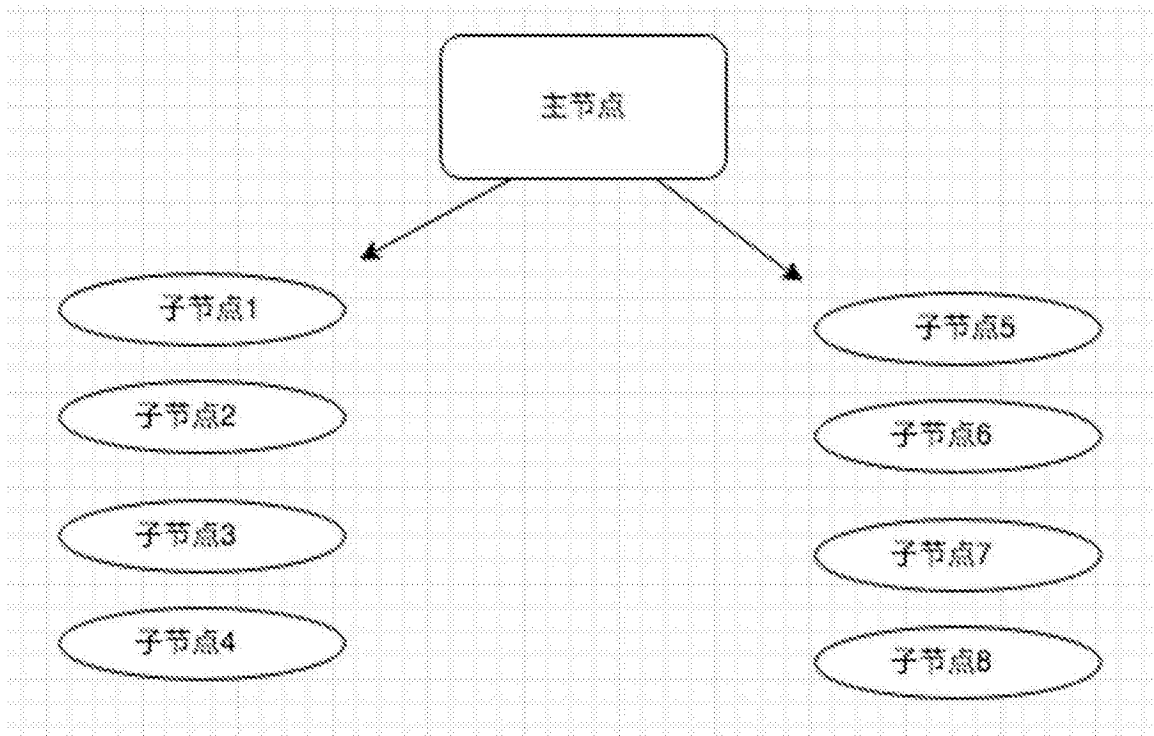


图1

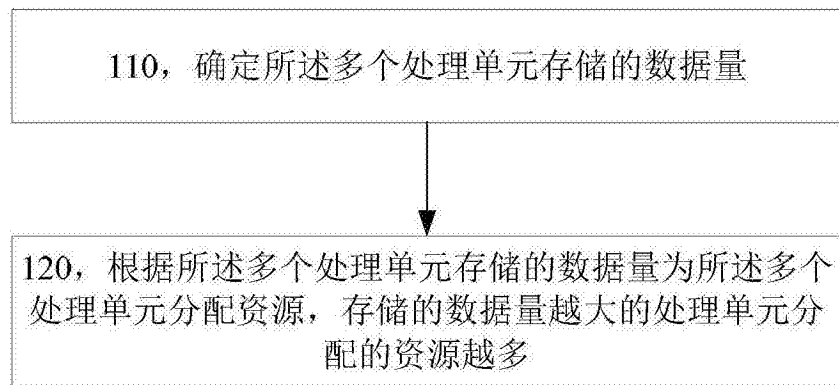


图2

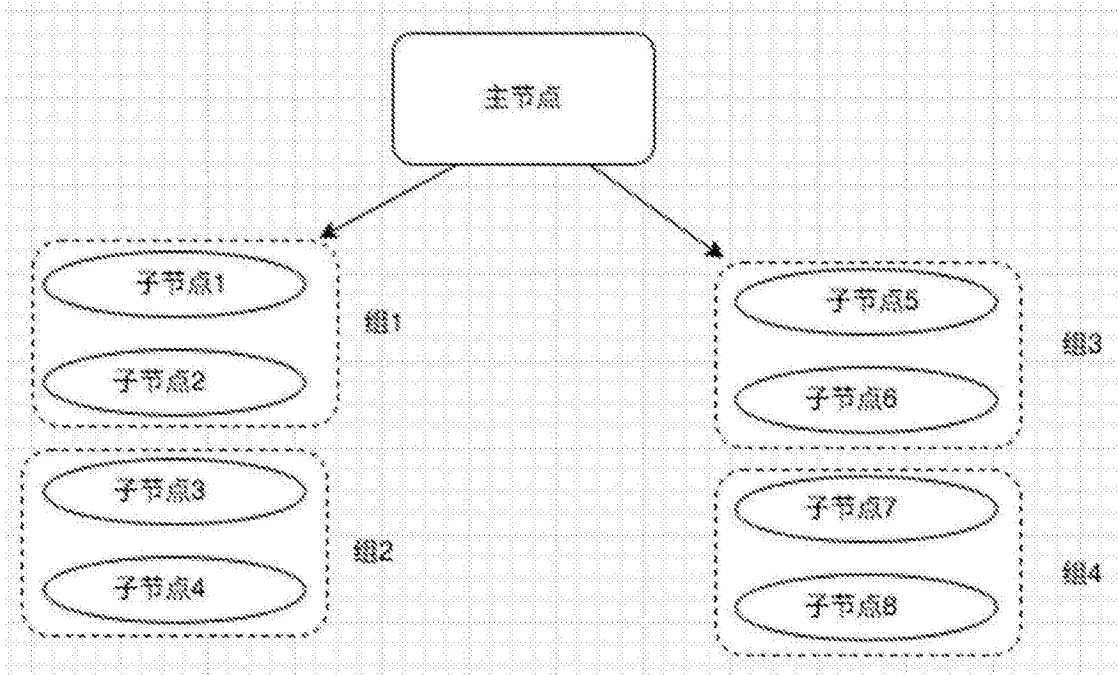


图3

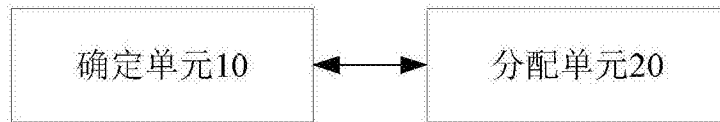


图4