



(12)发明专利

(10)授权公告号 CN 103716381 B

(45)授权公告日 2017.04.12

(21)申请号 201310681454.5

(22)申请日 2013.12.12

(65)同一申请的已公布的文献号

申请公布号 CN 103716381 A

(43)申请公布日 2014.04.09

(73)专利权人 华为技术有限公司

地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72)发明人 朱轲 周伟

(74)专利代理机构 深圳市深佳知识产权代理事务所(普通合伙) 44285

代理人 王仲凯

(51)Int.Cl.

H04L 29/08(2006.01)

H04L 12/24(2006.01)

(56)对比文件

CN 103188345 A, 2013.07.03,

CN 102004670 A, 2011.04.06,

CN 103369042 A, 2013.10.23,

CN 101576994 A, 2009.11.11,

审查员 胡浩

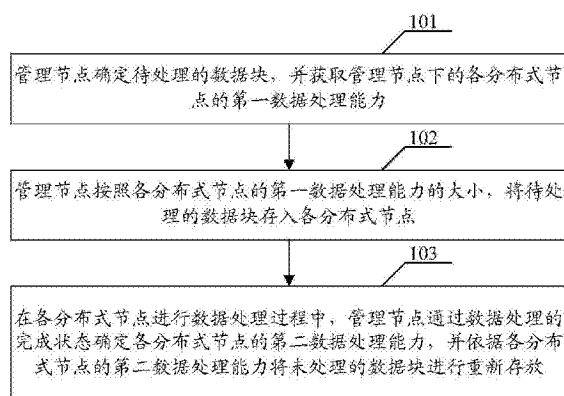
权利要求书3页 说明书17页 附图4页

(54)发明名称

一种分布式系统的控制方法,及管理节点

(57)摘要

本发明实施例公开一种分布式系统的控制方法,及管理节点,其中方法的实现包括:管理节点确定待处理的数据块,并获取管理节点下的各分布式节点的第一数据处理能力;管理节点按照各分布式节点的第一数据处理能力的大小,将待处理的数据块存入各分布式节点;在各分布式节点进行数据处理过程中,管理节点通过数据处理的完成状态确定各分布式节点的第二数据处理能力,并依据各分布式节点的第二数据处理能力将未处理的数据块进行重新存放。以上方案可以减少数据块的迁移,这样不仅可以充分利用分布式系统的计算能力,还可以通过提高数据处理的本地化率,来提升分布式系统的数据处理性能。



1. 一种分布式系统的控制方法,其特征在于,包括:

管理节点确定待处理的数据块,并获取管理节点下的各分布式节点的第一数据处理能力;

管理节点按照各分布式节点的第一数据处理能力的大小,将待处理的数据块存入各分布式节点;

在各分布式节点进行数据处理过程中,管理节点通过数据处理的完成状态确定各分布式节点的第二数据处理能力,并依据各分布式节点的第二数据处理能力将未处理的数据块进行重新存放;

其中,还包括:

管理节点获取各分布式节点的网络速率;

若按照第一数据处理能力大小或者第二数据处理能力大小,会将数据块存入第一分布式节点,并且第一分布式节点的网络速率小于网络速度预定阈值,则在分布式系统中重新选择节点进行存放。

2. 根据权利要求1所述方法,其特征在于,所述在各分布式节点进行数据处理过程中,管理节点通过数据处理的完成状态确定各分布式节点的第二数据处理能力包括:

在各分布式节点进行数据处理过程中,管理节点通过各分布式节点的空闲计算资源、待处理数据块的数量、网络状态以及加速比确定各分布式节点的第二数据处理能力。

3. 根据权利要求1所述方法,其特征在于,还包括:

在待处理的数据块全部被处理完毕后,依据各分布式节点处理的数据块调整各分布式节点的数据处理能力;

在接收到与所述待处理的数据块相同的待处理的数据块后,依据调整后的各分布式节点的数据处理能力的大小,将待处理的数据块存入各分布式节点。

4. 根据权利要求3所述方法,其特征在于,所述依据各分布式节点处理的数据块调整各分布式节点的数据处理能力包括:

依据第一次存入各分布式节点的数据块的数量以及各分布式节点实际处理数据块的数量,确定各分布式节点的数据处理本地化率;或者,依据存放在各分布式节点本地的数据块所需的处理时间以及各分布式节点的实际数据处理时间,确定各分布式节点的数据处理本地化率;

依据各分布式节点的数据处理本地化率调整各分布式节点的数据处理能力。

5. 根据权利要求1所述方法,其特征在于,所述将待处理的数据块存入各分布式节点包括:

将待处理的数据块存入各分布式节点,并由各分布式节点的中央处理器分配到图形处理器进行数据处理。

6. 根据权利要求1所述方法,其特征在于,还包括:

在待处理的数据块全部被处理完毕后,依据各数据块的处理时间以及处理各数据块的分布式节点的第一数据处理能力,确定各数据块的数据处理量;

管理节点在接收到与所述待处理的数据块相同的待处理的数据块后,依据各数据块的数据处理量以及各分布式节点的第一数据处理能力,将待处理的数据块存入各分布式节点。

7. 根据权利要求6所述方法,其特征在于,还包括:

管理节点获取各分布式节点的网络速率;

所述依据各数据块的数据处理量以及各分布式节点的第一数据处理能力,将待处理的数据块存入各分布式节点包括:

依据各数据块的数据处理量、各分布式节点的网络速率以及第一数据处理能力,将待处理的数据块存入各分布式节点。

8. 一种分布式系统的管理节点,其特征在于,包括:

数据确定单元,用于确定待处理的数据块;

能力确定单元,用于获取管理节点下的各分布式节点的第一数据处理能力;在各分布式节点进行数据处理过程中,通过数据处理的完成状态确定各分布式节点的第二数据处理能力;

分配单元,用于按照所述能力确定单元确定的各分布式节点的第一数据处理能力的大小,将待处理的数据块存入各分布式节点;依据所述能力确定单元确定的各分布式节点的第二数据处理能力将未处理的数据块进行重新存放;

其中,还包括:

速率获取单元,用于获取各分布式节点的网络速率;

所述分配单元,用于若按照第一数据处理能力大小或者第二数据处理能力大小,会将数据块存入第一分布式节点,并且第一分布式节点的网络速率小于网络速度预定阈值,则在分布式系统中重新选择节点进行存放。

9. 根据权利要求8所述管理节点,其特征在于,

所述能力确定单元,用于在各分布式节点进行数据处理过程中,通过各分布式节点的空闲计算资源、待处理数据块的数量、网络状态以及加速比确定各分布式节点的第二数据处理能力。

10. 根据权利要求8所述管理节点,其特征在于,

所述能力确定单元,还用于在待处理的数据块全部被处理完毕后,依据各分布式节点处理的数据块调整各分布式节点的数据处理能力;

所述分配单元,用于在接收到与所述待处理的数据块相同的待处理的数据块后,依据所述能力确定单元调整后的各分布式节点的数据处理能力的大小,将待处理的数据块存入各分布式节点。

11. 根据权利要求10所述管理节点,其特征在于,

所述能力确定单元,用于依据第一次存入各分布式节点的数据块的数量以及各分布式节点实际处理数据块的数量,确定各分布式节点的数据处理本地化率,依据各分布式节点的数据处理本地化率调整各分布式节点的数据处理能力;

或者,依据存放在各分布式节点本地的数据块所需的处理时间以及各分布式节点的实际数据处理时间,确定各分布式节点的数据处理本地化率,依据各分布式节点的数据处理本地化率调整各分布式节点的数据处理能力。

12. 根据权利要求8所述管理节点,其特征在于,还包括:

处理量确定单元,用于在待处理的数据块全部被处理完毕后,依据各数据块的处理时间以及处理各数据块的分布式节点的第一数据处理能力,确定各数据块的数据处理量;

所述分配单元,用于在管理节点在接收到与所述待处理的数据块相同的待处理的数据块后,依据各数据块的数据处理量以及各分布式节点的第一数据处理能力,将待处理的数据块存入各分布式节点。

13.根据权利要求12所述管理节点,其特征在于,还包括:

获取速率单元,用于获取各分布式节点的网络速率;

所述分配单元,用于依据各数据块的数据处理量、各分布式节点的网络速率以及第一数据处理能力,将待处理的数据块存入各分布式节点。

一种分布式系统的控制方法,及管理节点

技术领域

[0001] 本发明涉及通信技术领域,特别涉及一种分布式系统的控制方法,及管理节点。

背景技术

[0002] 分布式系统(distributed system)是建立在网络之上的软件系统。正是因为软件的特性,所以分布式系统具有高度的内聚性和透明性。因此,网络和分布式系统之间的区别更多的在于高层软件(特别是操作系统),而不是硬件。内聚性是指每一个数据库分布节点高度自治,有本地的数据库管理系统。透明性是指每一个数据库分布节点对用户的应用来说都是透明的,看不出是本地还是远程。在分布式数据库系统中,用户感觉不到数据是分布的,即用户不须知道关系是否分割、有无副本、数据存于哪个站点以及事务在哪个站点上执行等。

[0003] 在一个分布式系统中,一组独立的计算机展现给用户的是一个统一的整体,就好像是一个系统似的。系统拥有多种通用的物理和逻辑资源,可以动态的分配任务,分散的物理和逻辑资源通过计算机网络实现信息交换。系统中存在一个以全局的方式管理计算机资源的分布式操作系统。通常,对用户来说,分布式系统只有一个模型或范型。在操作系统之上有一层软件中间件(middleware)负责实现这个模型。一个著名的分布式系统的例子是万维网(World Wide Web),在万维网中,所有的一切看起来就好像是一个文档(Web页面)一样。

[0004] 在计算机网络中,这种统一性、模型以及其中的软件都不存在。用户看到的是实际的机器,计算机网络并没有使这些机器看起来是统一的。如果这些机器有不同的硬件或者不同的操作系统,那么,这些差异对于用户来说都是完全可见的。如果一个用户希望在一台远程机器上运行一个程序,那么,他必须登陆到远程机器上,然后在那台机器上运行该程序。

[0005] 分布式系统和计算机网络系统的共同点是:多数分布式系统是建立在计算机网络之上的,所以分布式系统与计算机网络在物理结构上是基本相同的。

[0006] 他们的区别在于:分布式操作系统的设计思想和网络操作系统是不同的,这决定了他们在结构、工作方式和功能上也不同。网络操作系统要求网络用户在使用网络资源时首先必须了解网络资源,网络用户必须知道网络中各个计算机的功能与配置、软件资源、网络文件结构等情况,在网络中如果用户要读一个共享文件时,用户必须知道这个文件放在哪一台计算机的哪一个目录下;分布式操作系统是以全局方式管理系统资源的,它可以为用户任意调度网络资源,并且调度过程是“透明”的。当用户提交一个作业时,分布式操作系统能够根据需要在系统中选择最合适的处理器,将用户的作业提交到该处理程序,在处理器完成作业后,将结果传给用户。在这个过程中,用户并不会意识到有多个处理器的存在,这个系统就像是一个处理器一样。

[0007] 随着计算机、互联网和通讯技术的快速发展,网络对海量级数据存储能力和计算能力的需求日益提升。例如:云计算通过协同调度网络中已有的软硬件资源,实现了存储与

计算服务模式的虚拟化和透明化,并以其高效、灵活、拓展性强等诸多优势而成为解决网络中海量数据存储与计算的最新方案。目前在分布式系统中实现计算分配控制的方案一般可以如下:

[0008] 1、管理节点首先根据配置文件确定待处理的数据块的数量,并获取管理节点下的各分布式节点的数据处理能力;

[0009] 2、管理节点按照各分布式节点的数据处理能力的大小确定待处理的数据块的分配方案;

[0010] 3、管理节点按照上述分配方案将待处理的数据块存入各分布式节点;

[0011] 4、分布式节点对存入本地的待处理的数据块进行处理,并输出数据处理结果。

[0012] 以上方案,待处理的数据块是按照数据块的数据量或者个数来分配的,分配的参考值是以分布式节点的数据处理能力,即分布式节点的硬件处理能力。然而,实际应用中,由于数据块内容的不同等原因,会存在巨大的数据处理量的差异,数据处理量小的数据块会很快被计算完毕。为了充分利用分布式系统的计算能力,被分配处理量小的数据块的分布式节点在计算完毕以后,会向计算量大的分布式节点进行数据请求,请求将其未处理的数据块发送给自己处理。以上方案为了充分利用分布式系统的计算能力,导致数据处理的本地化率较低,分布式系统的数据处理性能也较低。

发明内容

[0013] 本发明实施例提供了一种分布式系统的控制方法,及管理节点,用于在充分利用分布式系统的计算能力的前提下,提高数据处理的本地化率,并提升分布式系统的数据处理性能。

[0014] 本发明实施例一方面提供了一种分布式系统的控制方法,包括:

[0015] 管理节点确定待处理的数据块,并获取管理节点下的各分布式节点的第一数据处理能力;

[0016] 管理节点按照各分布式节点的第一数据处理能力的大小,将待处理的数据块存入各分布式节点;

[0017] 在各分布式节点进行数据处理过程中,管理节点通过数据处理的完成状态确定各分布式节点的第二数据处理能力,并依据各分布式节点的第二数据处理能力将未处理的数据块进行重新存放。

[0018] 结合一方面的实现方式,在第一种可能的实现方式中,所述在各分布式节点进行数据处理过程中,管理节点通过数据处理的完成状态确定各分布式节点的第二数据处理能力包括:

[0019] 在各分布式节点进行数据处理过程中,管理节点通过各分布式节点的空闲计算资源、待处理数据块的数量、网络状态以及加速比确定各分布式节点的第二数据处理能力。

[0020] 结合一方面的实现方式,在第二种可能的实现方式中,所述方法,还包括:

[0021] 在待处理的数据块全部被处理完毕后,依据各分布式节点处理的数据块调整各分布式节点的数据处理能力;

[0022] 在接收到与所述待处理的数据块相同的待处理的数据块后,依据调整后的各分布式节点的数据处理能力的大小,将待处理的数据块存入各分布式节点。

[0023] 结合一方面的第二种可能的实现方式,在第三种可能的实现方式中,所述依据各分布式节点处理的数据块调整各分布式节点的数据处理能力包括:

[0024] 依据第一次存入各分布式节点的数据块的数量以及各分布式节点实际处理数据块的数量,确定各分布式节点的数据处理本地化率;或者,依据存放在各分布式节点本地的数据块所需的处理时间以及各分布式节点的实际数据处理时间,确定各分布式节点的数据处理本地化率依据各分布式节点的数据处理本地化率调整各分布式节点的数据处理能力。

[0025] 结合一方面的实现方式,在第四种可能的实现方式中,所述将待处理的数据块存入各分布式节点包括:

[0026] 将待处理的数据块存入各分布式节点,并由各分布式节点的中央处理器分配到图形处理器进行数据处理。

[0027] 结合一方面的实现方式、一方面的第一种、第二种、第三种或者第四种可能的实现方式,在第五种可能的实现方式中,所述方法,还包括:

[0028] 管理节点获取各分布式节点的网络速率;

[0029] 若按照第一数据处理能力大小或者第二数据处理能力大小,会将数据块存入第一分布式节点,并且第一分布式节点的网络速率小于网络速度预定阈值,则在分布式系统中重新选择节点进行存放。

[0030] 结合一方面的实现方式,在第六种可能的实现方式中,所述方法,还包括:

[0031] 在待处理的数据块全部被处理完毕后,依据各数据块的处理时间以及处理各数据块的分布式节点的第一数据处理能力,确定各数据块的数据处理量;

[0032] 管理节点在接收到与所述待处理的数据块相同的待处理的数据块后,依据各数据块的数据处理量以及各分布式节点的第一数据处理能力,将待处理的数据块存入各分布式节点。

[0033] 结合一方面的第六种可能的实现方式,在第七种可能的实现方式中,所述方法,其特征在于,还包括:

[0034] 管理节点获取各分布式节点的网络速率;

[0035] 所述依据各数据块的数据处理量以及各分布式节点的第一数据处理能力,将待处理的数据块存入各分布式节点包括:

[0036] 依据各数据块的数据处理量、各分布式节点的网络速率以及第一数据处理能力,将待处理的数据块存入各分布式节点。

[0037] 本发明实施例二方面提供了一种分布式系统的管理节点,包括:

[0038] 数据确定单元,用于确定待处理的数据块;

[0039] 能力确定单元,用于获取管理节点下的各分布式节点的第一数据处理能力;在各分布式节点进行数据处理过程中,通过数据处理的完成状态确定各分布式节点的第二数据处理能力;

[0040] 分配单元,用于按照所述能力确定单元确定的各分布式节点的第一数据处理能力的大小,将待处理的数据块存入各分布式节点;依据所述能力确定单元确定的各分布式节点的第二数据处理能力将未处理的数据块进行重新存放。

[0041] 结合二方面的实现方式,在第一种可能的实现方式中,所述能力确定单元,用于在各分布式节点进行数据处理过程中,管理节点通过各分布式节点的空闲计算资源、待处理

数据块的数量、网络状态以及加速比确定各分布式节点的第二数据处理能力。

[0042] 结合一方面的实现方式,在第二种可能的实现方式中,所述能力确定单元,还用于在待处理的数据块全部被处理完毕后,依据各分布式节点处理的数据块调整各分布式节点的数据处理能力;

[0043] 所述分配单元,用于在接收到与所述待处理的数据块相同的待处理的数据块后,依据所述能力确定单元调整后的各分布式节点的数据处理能力的大小,将待处理的数据块存入各分布式节点。

[0044] 结合一方面的第二种可能的实现方式,在第三种可能的实现方式中,所述能力确定单元,用于依据第一次存入各分布式节点的数据块的数量以及各分布式节点实际处理数据块的数量,确定各分布式节点的数据处理本地化率;或者,依据存放在各分布式节点本地的数据块所需的处理时间以及各分布式节点的实际数据处理时间,确定各分布式节点的数据处理本地化率依据各分布式节点的数据处理本地化率调整各分布式节点的数据处理能力。

[0045] 结合二方面的实现方式、二方面的第一种、第二种或者第三种可能的实现方式,在第四种可能的实现方式中,所述管理节点,还包括:

[0046] 速率获取单元,用于获取各分布式节点的网络速率;

[0047] 所述分配单元,用于若按照第一数据处理能力大小或者第二数据处理能力大小,会将数据块存入第一分布式节点,并且第一分布式节点的网络速率小于网络速度预定阈值,则在分布式系统中重新选择节点进行存放。

[0048] 结合一方面的实现方式,在第五种可能的实现方式中,所述管理节点,还包括:

[0049] 处理量确定单元,用于在待处理的数据块全部被处理完毕后,依据各数据块的处理时间以及处理各数据块的分布式节点的第一数据处理能力,确定各数据块的数据处理量;

[0050] 所述分配单元,用于在管理节点在接收到与所述待处理的数据块相同的待处理的数据块后,依据各数据块的数据处理量以及各分布式节点的第一数据处理能力,将待处理的数据块存入各分布式节点。

[0051] 结合一方面的第五种可能的实现方式,在第六种可能的实现方式中,所述管理节点,还包括:

[0052] 获取速率单元,用于获取各分布式节点的网络速率;

[0053] 所述分配单元,用于依据各数据块的数据处理量、各分布式节点的网络速率以及第一数据处理能力,将待处理的数据块存入各分布式节点。

[0054] 从以上技术方案可以看出,本发明实施例具有以下优点:在分布式系统的各分布式节点进行数据处理过程中,数据处理量小的数据块会迅速被处理完毕,这样这些被处理完毕的数据块所占用的数据处理能力将会被释放,因此在分布式节点进行数据处理过程中,各分布式节点的数据处理能力实际上是在不停变化中的。基于此,本发明实施例提出了动态确定分布式节点的数据处理能力的方案,从而对未处理的数据块进行重新存放,来减少数据块的迁移。这样不仅可以充分利用分布式系统的计算能力,还可以通过提高数据处理的本地化率,来提升分布式系统的数据处理性能。

附图说明

[0055] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简要介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域的普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0056] 图1为本发明实施例方法流程示意图;

[0057] 图2为本发明实施例分布式异构系统架构示意图;

[0058] 图3为本发明实施例方法流程示意图;

[0059] 图4为本发明实施例分布式异构系统数据存放示意图;

[0060] 图5为本发明实施例分布式异构系统数据存放示意图;

[0061] 图6为本发明实施例管理节点结构示意图;

[0062] 图7为本发明实施例管理节点结构示意图;

[0063] 图8为本发明实施例管理节点结构示意图;

[0064] 图9为本发明实施例管理节点结构示意图;

[0065] 图10为本发明实施例管理节点结构示意图。

具体实施方式

[0066] 为了使本发明的目的、技术方案和优点更加清楚,下面将结合附图对本发明作进一步地详细描述,显然,所描述的实施例仅仅是本发明一部份实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其它实施例,都属于本发明保护的范围。

[0067] 本发明实施例提供了一种分布式系统的控制方法,如图1所示,包括:

[0068] 101:管理节点确定待处理的数据块,并获取管理节点下的各分布式节点的第一数据处理能力;

[0069] 在本发明实施例中,管理节点是确定数据块如何分配的决策节点,它可以与接收业务处理任务(待处理数据块)的分发设备集成在一起作为管理节点使用,也可以单独存在于网络中,这种情况下管理节点的分配决策将会下发到分发设备指令分发设备按照分配决策进行数据块的存放。管理节点与分发设备是否集成在一起,对此本发明实施例不予限定。

[0070] 管理节点确定待处理的数据块的方式可以具体如下:在接收到处理任务以后,可以根据与这个处理任务对应的配置文件确定待处理的数据块的数量,这样处理任务可以拆分为很多的数据块,那么这些数据块就是待处理的数据块。

[0071] 102:管理节点按照各分布式节点的第一数据处理能力的大小,将待处理的数据块存入各分布式节点;

[0072] 可选地,本发明实施例可以应用于异构的分布式系统中,基于传统CPU(Central Processing Unit,中央处理器)的分布式计算体系结构具有超强的海量数据处理能力,但不具有足够的高性能计算能力,将具有超强并行计算能力的GPU(Graphic Processing Unit,图形处理器)与云计算相融合,基于CPU/GPU异构平台的高性能计算体系结构是云计算硬件平台发展的一个重要方向。本发明实施例提供的异构的分布式系统的实现方案具体

如下:上述将待处理的数据块存入各分布式节点包括:将待处理的数据块存入各分布式节点,由各分布式节点的中央处理器分配到图形处理器进行数据处理。

[0073] 进一步地,在数据块被存入的过程中,由于各分布式节点的网络速率并不相同,数据块存入各分布式节点所需的时间也不同,若仅考虑各分布式节点的数据处理能力,则仍然不能尽快的完成数据块的处理,因此本发明实施例提出了进一步参考各分布式节点的网络速率的方案,具体如下:上述方法,还包括:管理节点获取各分布式节点的网络速率;若按照第一数据处理能力大小或者第二数据处理能力大小,会将数据块存入第一分布式节点,并且第一分布式节点的网络速率小于网络速度预定阈值,则在分布式系统中重新选择节点进行存放。

[0074] 分布式节点的网络速率的获取方案可以有很多,例如:按照节点网络速率 $(N_S) = (\text{非本地处理的数据块大小}) / (\text{传输时间})$ 、节点网络状态 $(S_S) = N_S / (\text{分布式网络平均网速})$,然后按照节点网络状态来确定网络传输速率。还可以由各分布式节点上报自身的网络速率。网络速率的获得方式,本发明实施例不予限定。

[0075] 103:在各分布式节点进行数据处理过程中,管理节点通过数据处理的完成状态确定各分布式节点的第二数据处理能力,并依据各分布式节点的第二数据处理能力将未处理的数据块进行重新存放。

[0076] 在分布式系统的各分布式节点进行数据处理过程中,数据处理量小的数据块会迅速被处理完毕,这样这些被处理完毕的数据块所占用的数据处理能力将会被释放,因此在分布式节点进行数据处理过程中,各分布式节点的数据处理能力实际上是在不停变化中的。基于此,本发明实施例提出了动态确定分布式节点的数据处理能力的方案,从而对未处理的数据块进行重新存放,来减少数据块的迁移。这样不仅可以充分利用分布式系统的计算能力,还可以通过提高数据处理的本地化率,来提升分布式系统的数据处理性能。

[0077] 本发明实施例还提供了第二数据处理能力的具体计算方案,具体如下:上述在各分布式节点进行数据处理过程中,管理节点通过数据处理的完成状态确定各分布式节点的第二数据处理能力包括:在各分布式节点进行数据处理过程中,管理节点通过各分布式节点的空闲计算资源、待处理数据块的数量、网络状态以及加速比确定各分布式节点的第二数据处理能力。

[0078] 在分布式系统中,经常会有很多业务数据是需要经常处理的,例如一些需要定时周期性处理的业务报表、统计数据库的计算等等。这些数据在计算完一次以后,数据块的再次分配是可以参考前一次的数据处理的实际执行情况来进行分配的,这样可以进一步的使数据块的实际数据处理量与各分布式节点的数据处理能力相适应,来减少数据块的迁移,具体方案如下:进一步地,上述方法,还包括:在待处理的数据块全部被处理完毕后,依据各分布式节点处理的数据块调整各分布式节点的数据处理能力;在接收到与上述待处理的数据块相同的待处理的数据块后,依据调整后的各分布式节点的数据处理能力的大小,将待处理的数据块存入各分布式节点。

[0079] 由于数据块已经被处理完毕,那么数据块的实际处理节点(或者每个分布式节点实际处理的数据块)是可知的,那么据此是可以制定相关的规则来重新确定分布式节点的数据处理能力。被调整后的数据处理能力将不再单体现了数据处理的速度,还会与数据块的内容等建立内在的关联,从而更贴近实际处理时间的优化,基于此本发明实施例给出了

依据各分布式节点处理的数据块调整各分布式节点的数据处理能力具体实现方式的举例，如下：可选地，上述依据各分布式节点处理的数据块调整各分布式节点的数据处理能力包括：依据第一次存入各分布式节点的数据块的数量以及各分布式节点实际处理数据块的数量，确定各分布式节点的数据处理本地化率；或者，依据存放在各分布式节点本地的数据块所需的处理时间以及各分布式节点的实际数据处理时间，确定各分布式节点的数据处理本地化率依据各分布式节点的数据处理本地化率调整各分布式节点的数据处理能力。

[0080] 可以理解的是，如果各分布式节点的数据处理能力被调整以后，在进行数据块分配过程中，将会有与原有的分配结果发生改变，相应存放的方式也会发生改变，进而减少数据块的迁移，提高数据处理的本地化率。

[0081] 在分布式系统中，经常会有很多业务数据是需要经常处理的，例如一些需要定时周期性处理的业务报表、统计数据库的计算等等。这些数据在计算完一次以后，数据块的再次分配是可以参考前一次的数据处理的实际执行情况来进行分配的，这样可以进一步的使数据块的实际数据处理量与各分布式节点的数据处理能力相适应，来减少数据块的迁移；与前一方案不同的是，基于第一次数据块的处理，由于已经获知的信息有：各分布式节点的数据处理能力，可以获知的有各数据块的实际处理时间，那么基于这两点是可以计算出各数据块的数据处理量的，数据处理量可以有各种参考值，例如采用实际处理的分布式节点的数据处理能力与实际处理时间的积作为参考值。基于此本实施例具体可以如下：进一步地，上述方法，还包括：

[0082] 在待处理的数据块全部被处理完毕后，依据各数据块的处理时间以及处理各数据块的分布式节点的第一数据处理能力，确定各数据块的数据处理量；

[0083] 管理节点在接收到与上述待处理的数据块相同的待处理的数据块后，依据各数据块的数据处理量以及各分布式节点的第一数据处理能力，将待处理的数据块存入各分布式节点。

[0084] 在获知各数据块的数据处理量以后，分配数据块的策略可以继续增加可能影响分布式系统的实际处理时间的参数，例如：网络速率。可以获知的是，如果一个分布式节点仅处理一个数据块，那么处理时间应该是这个数据块传输到这个节点的时间、该分布式节点处理这个数据块的时间，以及数据处理结果回传的时间之和。因此网络速率是会影响到分布式系统的实际处理时间的，应该将数据块分配给以上“时间之和”最小的节点，基于此本发明实施例提供了如下解决方案：进一步地，上述方法，还包括：管理节点获取各分布式节点的网络速率；上述依据各数据块的数据处理量以及各分布式节点的第一数据处理能力，将待处理的数据块存入各分布式节点包括：依据各数据块的数据处理量、各分布式节点的网络速率以及第一数据处理能力，将待处理的数据块存入各分布式节点。

[0085] 以下实施例将以异构的分布式系统为例对本发明实施例进行更详细的举例说明，需要说明的是，在任意的分布式系统中均可以使用本发明实施例方案并不仅限于异构的分布式系统。

[0086] 本实施例在由异构Hadoop（一种分布式系统基础架构）集群构成的云计算平台中，每个分布式节点的软硬件配置是不相同的，其计算能力也会有明显的差异，高速的分布式节点会比低速分布式节点更快的处理完存储在本地磁盘的数据。为了尽快地完成数据处理任务，快速节点在处理完本地输入数据任务时，会耗费有限的网络带宽来请求处理附近慢

速分布式节点未处理的数据。在异构的分布式系统中,各分布式节点的处理能力是根据CPU+GPU来计算硬件的数据处理能力(也可以视为是计算能力),进行文件存放和任务调度。本发明实施例的分布式异构系统架构示意图,如图2所示,包含,客户端201、管理节点202、分布式节点203;客户端201和分布式节点203可能有很多个,客户端201提出分布式处理的业务需求,即待处理数据块的来源;分布式节点203用于处理存放到本地的数据块得到数据处理结果,数据处理的结果可以输出可以通过管理节点202回传到客户端201。分布式节点203从上层到底层主要可以有Task Tracker(任务追踪)CUDA(Compute Unified Device Architecture,一种运算架构)/OPENCL(Computing Language,运算语言)、OS(Operating System,操作系统)、CPU、GPU。管理节点202的主要功能是数据块的分配决策以及数据块的存放。

[0087] 本发明实施例中,管理节点主要的功能有两个部分:一、数据存放:在分布式系统中,数据存放以分布式节点计算能力为标准,计算能力越强,存放的数据块数越多。二、任务分配:控制节点动态计算分布式节点计算能力,任务分配器(可以是集成在管理节点内的功能单元)根据分布式节点的计算能力,以数据为中心,将计算任务发给计算能力强的节点进行处理。分布式节点的功能是:任务执行:分布式节点从分布式文件系统获取输入数据后,CPU将计算任务放到GPU上进行运算,并返回结果。具体流程如图3所示。

[0088] 数据存放流程,如图3左侧流程所示,包括如下步骤:

[0089] 301、控制节点根据配置文件计算需存放的数据块数量。

[0090] 302、控制节点根据各分布式节点的硬件配置或运行标准程序等的参数,计算各分布式节点的计算能力。

[0091] 另外,还会在计算任务(数据块)被分布式节点处理完毕时,根据任务执行状况更新计算能力。此处图3右侧流程图将给出更详细的说明。

[0092] 303、控制节点根据各分布式节点的计算能力,进行数据块分配。

[0093] 304、控制节点按照分配结果将数据块存入分布式节点。

[0094] 305、控制节点计算分布式节点的网络状态,进入步骤302后,控制节点根据网络状态和分布式节点存放的数据块数量,重新计算分布式节点计算能力。其中步骤302至305是重复执行的,直到所有数据块存放完毕。

[0095] 在本实施例中,数据存放原则可以如下:

[0096] 1、以块为单位,在系统中查找一个计算能力最强的分布式节点进行数据存放。

[0097] 2、分布式节点计算能力随着存放数据块的数量增加而减弱;另外,网络状态越好,计算能力越强。

[0098] 3、在各分布式节点计算能力相同时,选择存放数据块数量较少的分布式节点存放。

[0099] 4、分布式节点计算能力和存放数据块的块数都相同时,可以随机选择。

[0100] 基于本发明实施例提供的数据块的分配/存放规则,以下给出了一个数据存放示例:如图4所示,有6个数据块B_1~B_6需要存放到6个分布式节点C_1~C_6。在本例中,需要存放数据块总数为6块,系统中有6个数据节点,其计算能力分别为1,2,3,4,5,6.对应的分布式节点编号为节点C_1,C_2,C_3,C_4,C_5,C_6。具体存放过程如表1~表6所示。

[0101] 1、根据存放规则:将B_1存入到节点6;

[0102] 调整系统节点计算能力,寻找B_2可存放的节点。

[0103] 表1

[0104]

	节 点 C_1	节 点 C_2	节 点 C_3	节点 C_4	节 点 C_5	节点 C_6
初始计算	1	2	3	4	5	6

[0105]

能力						
当前存放 块与块数	0	0	0	0	0	1 (B_1)
当前计算 能力	1	2	3	4	5	5

[0106] 2、根据存放规则:将B_2存入到节点5;

[0107] 调整系统节点计算能力,寻找B_3可存放的节点。

[0108] 表2

[0109]

	节点 C_1	节点 C_2	节点 C_3	节点 C_4	节点 C_5	节点 C_6
初始计算 能力	1	2	3	4	5	6
当前存放 块与块数	0	0	0	0	1 (B_2)	1 (B_1)
当前计算 能力	1	2	3	4	4	5

[0110] 3、根据存放规则:将B_3存入到节点6;

[0111] 调整系统节点计算能力,寻找B_4可存放的节点。

[0112] 表3

[0113]

	节点 C_1	节点 C_2	节点 C_3	节点 C_4	节点 C_5	节点 C_6
初始计算 能力	1	2	3	4	5	6
当前存放 块与块数	0	0	0	0	1 (B_2)	2 (B_1, B_3)

[0114]

当前计算 能力	1	2	3	4	4	4
------------	---	---	---	---	---	---

[0115] 4、根据存放规则：将B_4存入到节点4；

[0116] 调整系统节点计算能力，寻找B_5可存放的节点。

[0117] 表4

[0118]

	节点 C_1	节点 C_2	节点 C_3	节点 C_4	节点 C_5	节点 C_6
初始计算 能力	1	2	3	4	5	6
当前存放 块与块数	0	0	0	1 (B_4)	1 (B_2)	2 (B_1, B_3)
当前计算 能力	1	2	3	3	4	4

[0119] 5、根据存放规则：将B_5存入到节点5；

[0120] 调整系统节点计算能力，寻找B_6可存放的节点。

[0121] 表5

[0122]

	节点 C_1	节点 C_2	节点 C_3	节点 C_4	节点 C_5	节点 C_6
初始计算能力	1	2	3	4	5	6
当前存放块与块数	0	0	0	1 (B_4)	2 (B_2, B_5)	2 (B_1, B_3)
当前计算能力	1	2	3	3	3	4

[0123] 6、根据存放规则：将B_6存入到节点6；调整系统节点计算能力。

[0124] 表6

[0125]

	节点 C_1	节点 C_2	节点 C_3	节点 C_4	节点 C_5	节点 C_6
初始计算能力	1	2	3	4	5	6
当前存放块与块数	0	0	0	1 (B_4)	2 (B_2, B_5)	3 (B_1, B_3, B_6)
当前计算能力	1	2	3	3	3	3

[0126] 数据块存放完毕后的图如图5所示，请对比图4所示，在此不再赘述。

[0127] 本发明实施例的任务调度原则可以如下：

[0128] 1、任务分配器动态节点计算能力，以数据为中心进行任务调度。

[0129] 2、当有大量计算资源空闲时，则将未分配的任务分配到空闲节点（将数据进行转移）。

[0130] 任务分配过程如图3所示，包括如下步骤：

[0131] 306：控制节点获取各分布式节点的状态（分布式节点的状态包括：空闲的计算资源，待处理数据块数，节点网络状态，加速比等），依据各分布式节点的状态确定各分布式节

点的计算能力,可以选取分布式系统中计算能力最强的分布式节点。将本步骤中确定的计算能力传递给步骤302。

[0132] 以上步骤中参数计算公式可以参考如下:

[0133] 节点网络速率(N_S)=(任务执行时非本地数据块大小)/(传输时间);

[0134] 节点网络状态(S_S)= N_S /(分布式网络平均网速);

[0135] 加速比=(数据块本地执行平均耗时)/(数据块非本地执行平均耗时);

[0136] 步骤306确定的新的计算能力被传递给步骤302以后,将出现如下结果:控制节点控制下的任务分配器选取数据块,分配给上述计算能力最强的分布式节点。当数据块在该分布式节点上(计算能力强的分布式节点)的所有任务都已经计算完时,从分布式系统中选取计算能力最弱的分布式节点上的数据块对应的任务分配给上述计算能力强的分布式节点。以上步骤306是循环执行的,直到计算任务完成,进入步骤307。

[0137] 307:当计算任务完成后,计算分布式节点执行任务的本地化率,并调整下次执行数据存放时作为参考的各分布式节点的计算能力。

[0138] 任务结束后节点计算能力调整,可参考的公式如下:

[0139] 节点执行任务的本地化率(L)=(节点所执行任务的数据存放于本地的数据块总数)/(节点所执行任务的数据块总数)。

[0140] 或本地化率(L)=(节点所执行任务的数据存放于本地的数据块时间)/(节点所执行任务总的执行时间);

[0141] 节点调整后计算能力= K *(数据计算能力)/(节点执行任务的本地化率)。

[0142] 上述 K 是权值,技术人员可以进行设定,其具体取值本发明实施例对此不予限定。

[0143] 本发明实施例还提供了一种分布式系统的管理节点,如图6所示,包括:

[0144] 数据确定单元601,用于确定待处理的数据块;

[0145] 能力确定单元602,用于获取管理节点下的各分布式节点的第一数据处理能力;在各分布式节点进行数据处理过程中,通过数据处理的完成状态确定各分布式节点的第二数据处理能力;

[0146] 分配单元603,用于按照上述能力确定单元602确定的各分布式节点的第一数据处理能力的大小,将待处理的数据块存入各分布式节点;依据上述能力确定单元602确定的各分布式节点的第二数据处理能力将未处理的数据块进行重新存放。

[0147] 在本发明实施例中,管理节点是确定数据块如何分配的决策节点,它可以与接收业务处理任务(待处理数据块)的分发设备集成在一起作为管理节点使用,也可以单独存在于网络中,这种情况下管理节点的分配决策将会下发到分发设备指令分发设备按照分配决策进行数据块的存放。管理节点与分发设备是否集成在一起,对此本发明实施例不予限定。管理节点确定待处理的数据块的方式可以具体如下:在接收到处理任务以后,可以根据与这个处理任务对应的配置文件确定待处理的数据块的数量,这样处理任务可以拆分为很多的数据块,那么这些数据块就是待处理的数据块。

[0148] 在分布式系统的各分布式节点进行数据处理过程中,数据处理量小的数据块会迅速被处理完毕,这样这些被处理完毕的数据块所占用的数据处理能力将会被释放,因此在分布式节点进行数据处理过程中,各分布式节点的数据处理能力实际上是在不停变化中的。基于此,本发明实施例提出了动态确定分布式节点的数据处理能力的方案,从而对未处

理的数据块进行重新存放,来减少数据块的迁移。这样不仅可以充分利用分布式系统的计算能力,还可以还可以通过提高数据处理的本地化率,来提升分布式系统的数据处理性能。

[0149] 可选地,本发明实施例还提供了第二数据处理能力的具体计算方案,具体如下:上述能力确定单元602,用于在各分布式节点进行数据处理过程中,管理节点通过各分布式节点的空闲计算资源、待处理数据块的数量、网络状态以及加速比确定各分布式节点的第二数据处理能力。

[0150] 在分布式系统中,经常会有很多业务数据是需要经常处理的,例如一些需要定时周期性处理的业务报表、统计数据库的计算等等。这些数据在计算完一次以后,数据块的再次分配是可以参考前一次的数据处理的实际执行情况来进行分配的,这样可以进一步的使数据块的实际数据处理量与各分布式节点的数据处理能力相适应,来减少数据块的迁移,具体方案如下:进一步地,上述能力确定单元602,还用于在待处理的数据块全部被处理完毕后,依据各分布式节点处理的数据块调整各分布式节点的数据处理能力;

[0151] 上述分配单元603,用于在接收到与上述待处理的数据块相同的待处理的数据块后,依据上述能力确定单元602调整后的各分布式节点的数据处理能力的大小,将待处理的数据块存入各分布式节点。

[0152] 由于数据块已经被处理完毕,那么数据块的实际处理节点(或者每个分布式节点实际处理的数据块)是可知的,那么据此是可以制定相关的规则来重新确定分布式节点的数据处理能力。被调整后的数据处理能力将不再单体现了数据处理的速度,还会与数据块的内容等建立内在的关联,从而更贴近实际处理时间的优化,基于此本发明实施例给出了依据各分布式节点处理的数据块调整各分布式节点的数据处理能力具体实现方式的举例,如下:可选地,上述能力确定单元602,用于依据第一次存入各分布式节点的数据块的数量以及各分布式节点实际处理数据块的数量,确定各分布式节点的数据处理本地化率;或者,依据存放在各分布式节点本地的数据块所需的处理时间以及各分布式节点的实际数据处理时间,确定各分布式节点的数据处理本地化率依据各分布式节点的数据处理本地化率调整各分布式节点的数据处理能力。

[0153] 可以理解的是,如果各分布式节点的数据处理能力被调整以后,在进行数据块分配过程中,将会有与原有的分配结果发生改变,相应存放的方式也会发生改变,进而减少数据块的迁移,提高数据处理的本地化率。

[0154] 进一步地,在数据块被存入的过程中,由于各分布式节点的网络速率并不相同,数据块存入各分布式节点所需的时间也不同,若仅考虑各分布式节点的数据处理能力,则仍然不能尽快的完成数据块的处理,因此本发明实施例提出了进一步参考各分布式节点的网络速率的方案,具体如下:如图7所示,上述管理节点,还包括:

[0155] 速率获取单元701,用于获取各分布式节点的网络速率;

[0156] 上述分配单元603,用于若按照第一数据处理能力大小或者第二数据处理能力大小,会将数据块存入第一分布式节点,并且第一分布式节点的网络速率小于网络速度预定阈值,则在分布式系统中重新选择节点进行存放。

[0157] 分布式节点的网络速率的获取方案可以有很多,例如:按照节点网络速率(N_S)=(非本地处理的数据块大小)/(传输时间)、节点网络状态(S_S)= N_S /(分布式网络平均网速),然后按照节点网络状态来确定网络传输速率。还可以由各分布式节点上报自身的网络

速率。网络速率的获得方式,本发明实施例不予限定。

[0158] 在分布式系统中,经常会有很多业务数据是需要经常处理的,例如一些需要定时周期性处理的业务报表、统计数据库的计算等等。这些数据在计算完一次以后,数据块的再次分配是可以参考前一次的数据处理的实际执行情况来进行分配的,这样可以进一步的使数据块的实际数据处理量与各分布式节点的数据处理能力相适应,来减少数据块的迁移;与前一方案不同的是,基于第一次数据块的处理,由于已经获知的信息有:各分布式节点的数据处理能力,可以获知的有各数据块的实际处理时间,那么基于这两点是可以计算出各数据块的数据处理量的,数据处理量可以有各种参考值,例如采用实际处理的分布式节点的数据处理能力与实际处理时间的积作为参考值。基于此本实施例具体可以如下:进一步地,如图8所示,上述管理节点,还包括:

[0159] 处理量确定单元801,用于在待处理的数据块全部被处理完毕后,依据各数据块的处理时间以及处理各数据块的分布式节点的第一数据处理能力,确定各数据块的数据处理量;

[0160] 上述分配单元603,用于在管理节点在接收到与上述待处理的数据块相同的待处理的数据块后,依据各数据块的数据处理量以及各分布式节点的第一数据处理能力,将待处理的数据块存入各分布式节点。

[0161] 在获知各数据块的数据处理量以后,分配数据块的策略可以继续增加可能影响分布式系统的实际处理时间的参数,例如:网络速率。可以获知的是,如果一个分布式节点仅处理一个数据块,那么处理时间应该是这个数据块传输到这个节点的时间、该分布式节点处理这个数据块的时间,以及数据处理结果回传的时间之和。因此网络速率是会影响分布式系统的实际处理时间的,应该将数据块分配给以上“时间之和”最小的节点,基于此本发明实施例提供了如下解决方案:进一步地,如图9所示,上述管理节点,还包括:获取速率单元901,用于获取各分布式节点的网络速率;

[0162] 上述分配单元603,用于依据各数据块的数据处理量、各分布式节点的网络速率以及第一数据处理能力,将待处理的数据块存入各分布式节点。

[0163] 本发明实施例还提供了另一种分布式系统的管理节点,如图10所示,包括:接收器1001、发射器1002、处理器1003以及存储器1004;

[0164] 其中,上述处理器1003,用于确定待处理的数据块;获取管理节点下的各分布式节点的第一数据处理能力;按照各分布式节点的第一数据处理能力的大小,将待处理的数据块存入各分布式节点;在各分布式节点进行数据处理过程中,通过数据处理的完成状态确定各分布式节点的第二数据处理能力,并依据各分布式节点的第二数据处理能力将未处理的数据块进行重新存放。

[0165] 在本发明实施例中,管理节点是确定数据块如何分配的决策节点,它可以与接收业务处理任务(待处理数据块)的分发设备集成在一起作为管理节点使用,也可以单独存在于网络中,这种情况下管理节点的分配决策将会下发到分发设备指令分发设备按照分配决策进行数据块的存放。管理节点与分发设备是否集成在一起,对此本发明实施例不予限定。管理节点确定待处理的数据块的方式可以具体如下:在接收到处理任务以后,可以根据与这个处理任务对应的配置文件确定待处理的数据块的数量,这样处理任务可以拆分为很多的数据块,那么这些数据块就是待处理的数据块。

[0166] 在分布式系统的各分布式节点进行数据处理过程中,数据处理量小的数据块会迅速被处理完毕,这样这些被处理完毕的数据块所占用的数据处理能力将会被释放,因此在分布式节点进行数据处理过程中,各分布式节点的数据处理能力实际上是在不停变化中的。基于此,本发明实施例提出了动态确定分布式节点的数据处理能力的方案,从而对未处理的数据块进行重新存放,来减少数据块的迁移。这样不仅可以充分利用分布式系统的计算能力,还可还可以通过提高数据处理的本地化率,来提升分布式系统的数据处理性能。

[0167] 本发明实施例还提供了第二数据处理能力的具体计算方案,具体如下:上述处理器1003,用于在各分布式节点进行数据处理过程中,管理节点通过数据处理的完成状态确定各分布式节点的第二数据处理能力包括:用于在各分布式节点进行数据处理过程中,管理节点通过各分布式节点的空闲计算资源、待处理数据块的数量、网络状态以及加速比确定各分布式节点的第二数据处理能力。

[0168] 在分布式系统中,经常会有很多业务数据是需要经常处理的,例如一些需要定时周期性处理的业务报表、统计数据库的计算等等。这些数据在计算完一次以后,数据块的再次分配是可以参考前一次的数据处理的实际执行情况来进行分配的,这样可以进一步的使数据块的实际数据处理量与各分布式节点的数据处理能力相适应,来减少数据块的迁移,具体方案如下:进一步地,上述处理器1003,还用于在待处理的数据块全部被处理完毕后,依据各分布式节点处理的数据块调整各分布式节点的数据处理能力;在接收到与上述待处理的数据块相同的待处理的数据块后,依据调整后的各分布式节点的数据处理能力的大小,将待处理的数据块存入各分布式节点。

[0169] 由于数据块已经被处理完毕,那么数据块的实际处理节点(或者每个分布式节点实际处理的数据块)是可知的,那么据此是可以制定相关的规则来重新确定分布式节点的数据处理能力。被调整后的数据处理能力将不再单体现了数据处理的速度,还会与数据块的内容等建立内在的关联,从而更贴近实际处理时间的优化,基于此本发明实施例给出了依据各分布式节点处理的数据块调整各分布式节点的数据处理能力具体实现方式的举例,如下:可选地,上述处理器1003,用于依据各分布式节点处理的数据块调整各分布式节点的数据处理能力包括:用于依据第一次存入各分布式节点的数据块的数量以及各分布式节点实际处理数据块的数量,确定各分布式节点的数据处理本地化率;或者,依据存放在各分布式节点本地的数据块所需的处理时间以及各分布式节点的实际数据处理时间,确定各分布式节点的数据处理本地化率依据各分布式节点的数据处理本地化率调整各分布式节点的数据处理能力。

[0170] 可以理解的是,如果各分布式节点的数据处理能力被调整以后,在进行数据块分配过程中,将会有与原有的分配结果发生改变,相应存放的方式也会发生改变,进而减少数据块的迁移,提高数据处理的本地化率。

[0171] 可选地,本发明实施例可以应用于异构的分布式系统中,基于传统CPU(Central Processing Unit,中央处理器)的分布式计算体系结构具有超强的海量数据处理能力,但不具有足够的高性能计算能力,将具有超强并行计算能力的GPU(Graphic Processing Unit,图形处理器)与云计算相融合,基于CPU/GPU异构平台的高性能计算体系结构是云计算硬件平台发展的一个重要方向。本发明实施例提供的异构的分布式系统的实现方案具体如下:上述处理器1003,用于将待处理的数据块存入各分布式节点包括:将待处理的数据块

存入各分布式节点,并由各分布式节点的中央处理器分配到图形处理器进行数据处理。

[0172] 进一步地,在数据块被存入的过程中,由于各分布式节点的网络速率并不相同,数据块存入各分布式节点所需的时间也不同,若仅考虑各分布式节点的数据处理能力,则仍然不能尽快的完成数据块的处理,因此本发明实施例提出了进一步参考各分布式节点的网络速率的方案,具体如下:上述处理器1003,还用于获取各分布式节点的网络速率;

[0173] 若按照第一数据处理能力大小或者第二数据处理能力大小,会将数据块存入第一分布式节点,并且第一分布式节点的网络速率小于网络速度预定阈值,则在分布式系统中重新选择节点进行存放。

[0174] 分布式节点的网络速率的获取方案可以有很多,例如:按照节点网络速率(N_S)=(非本地处理的数据块大小)/(传输时间)、节点网络状态(S_S)= N_S /(分布式网络平均网速),然后按照节点网络状态来确定网络传输速率。还可以由各分布式节点上报自身的网络速率。网络速率的获得方式,本发明实施例不予限定。

[0175] 在分布式系统中,经常会有很多业务数据是需要经常处理的,例如一些需要定时周期性处理的业务报表、统计数据库的计算等等。这些数据在计算完一次以后,数据块的再次分配是可以参考前一次的数据处理的实际执行情况来进行分配的,这样可以进一步的使数据块的实际数据处理量与各分布式节点的数据处理能力相适应,来减少数据块的迁移;与前一方案不同的是,基于第一次数据块的处理,由于已经获知的信息有:各分布式节点的数据处理能力,可以获知的有各数据块的实际处理时间,那么基于这两点是可以计算出各数据块的数据处理量的,数据处理量可以有各种参考值,例如采用实际处理的分布式节点的数据处理能力与实际处理时间的积作为参考值。基于此本实施例具体可以如下:进一步地,上述处理器1003,还用于在待处理的数据块全部被处理完毕后,依据各数据块的处理时间以及处理各数据块的分布式节点的第一数据处理能力,确定各数据块的数据处理量;在接收到与上述待处理的数据块相同的待处理的数据块后,依据各数据块的数据处理量以及各分布式节点的第一数据处理能力,将待处理的数据块存入各分布式节点。

[0176] 在获知各数据块的数据处理量以后,分配数据块的策略可以继续增加可能影响分布式系统的实际处理时间的参数,例如:网络速率。可以获知的是,如果一个分布式节点仅处理一个数据块,那么处理时间应该是这个数据块传输到这个节点的时间、该分布式节点处理这个数据块的时间,以及数据处理结果回传的时间之和。因此网络速率是会影响到分布式系统的实际处理时间的,应该将数据块分配给以上“时间之和”最小的节点,基于此本发明实施例提供了如下解决方案:进一步地,上述处理器1003,还用于获取各分布式节点的网络速率;上述依据各数据块的数据处理量以及各分布式节点的第一数据处理能力,将待处理的数据块存入各分布式节点包括:依据各数据块的数据处理量、各分布式节点的网络速率以及第一数据处理能力,将待处理的数据块存入各分布式节点。

[0177] 值得注意的是,上述装置实施例中,所包括的各个单元只是按照功能逻辑进行划分的,但并不局限于上述的划分,只要能够实现相应的功能即可;另外,各功能单元的具体名称也只是为了便于相互区分,并不用于限制本发明的保护范围。

[0178] 另外,本领域普通技术人员可以理解实现上述各方法实施例中的全部或部分步骤是可以通程序来指令相关的硬件完成,相应的程序可以存储于一种计算机可读存储介质中,上述提到的存储介质可以是只读存储器,磁盘或光盘等。

[0179] 以上仅为本发明较佳的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明实施例揭露的技术范围内,可轻易想到的变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应该以权利要求的保护范围为准。

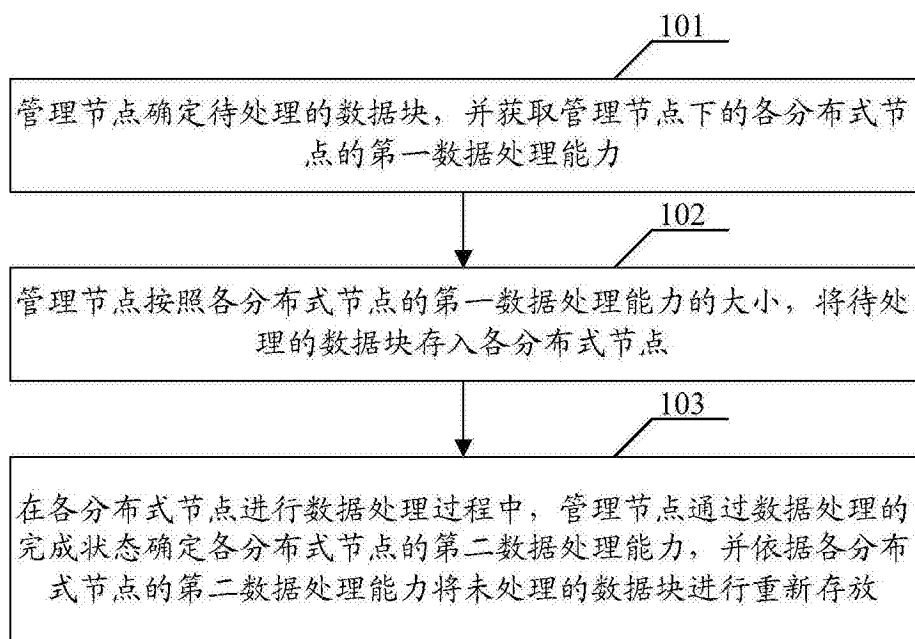


图1

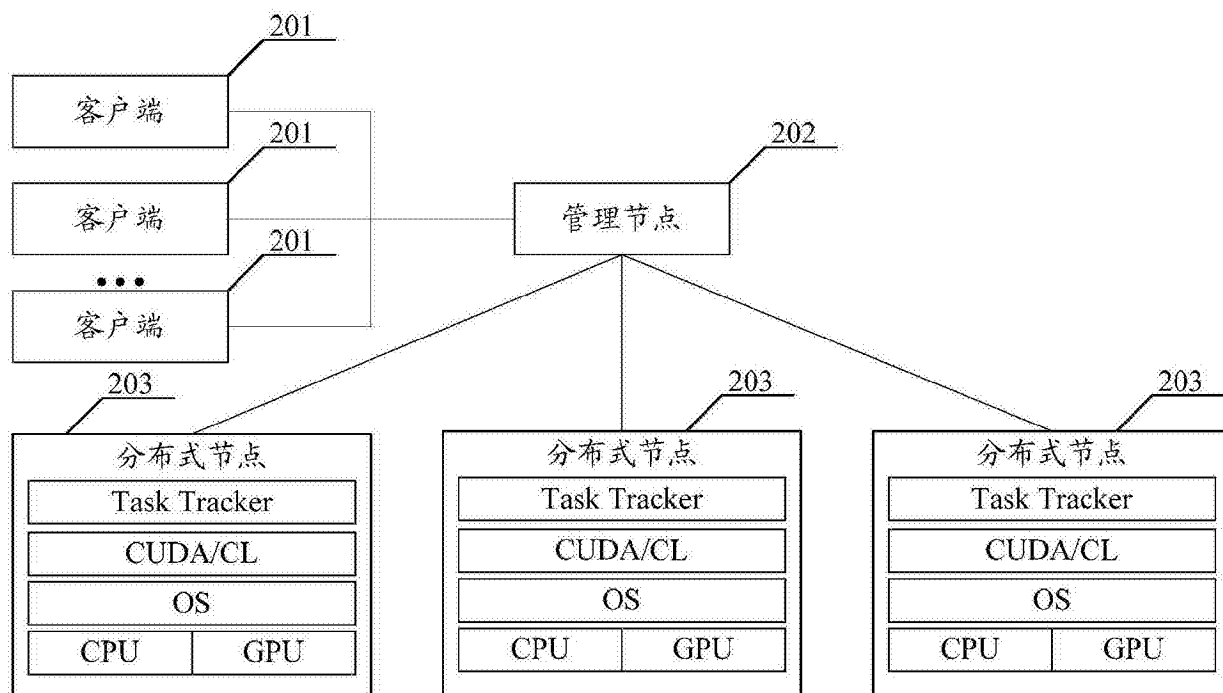


图2

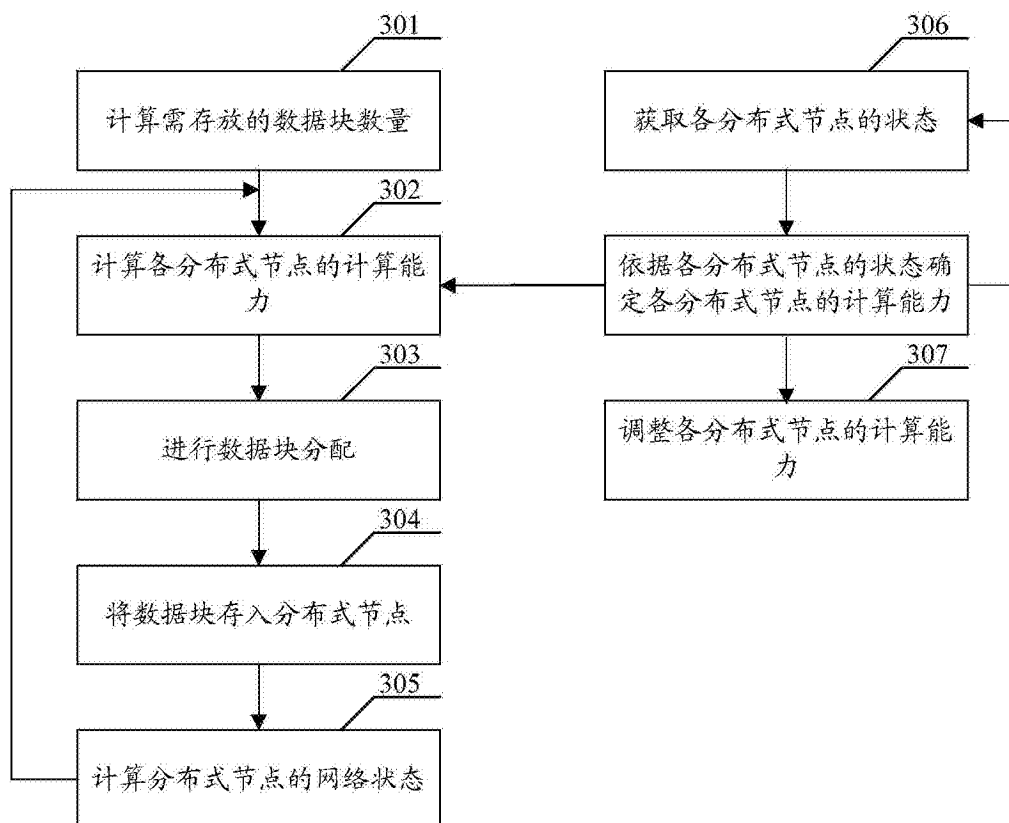


图3

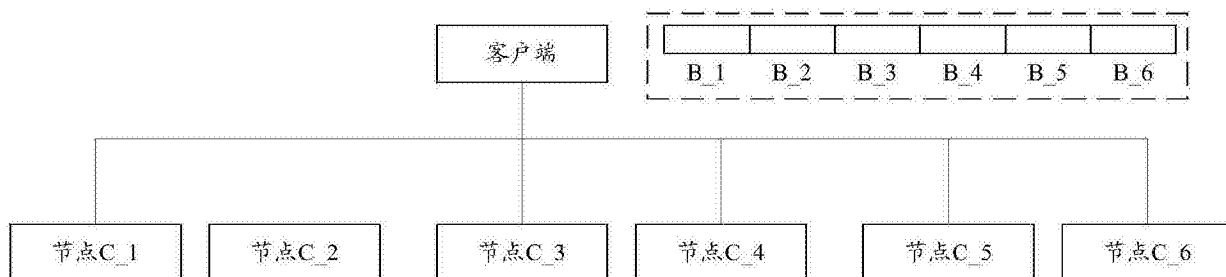


图4

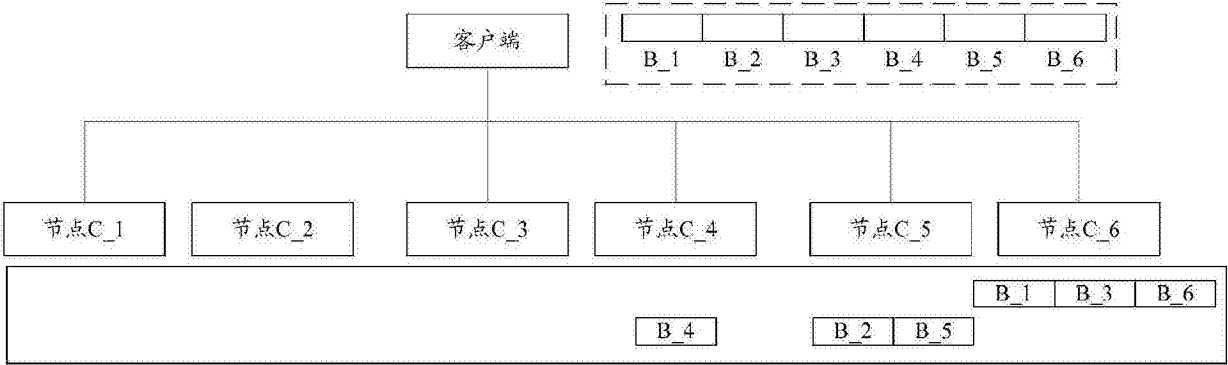


图5

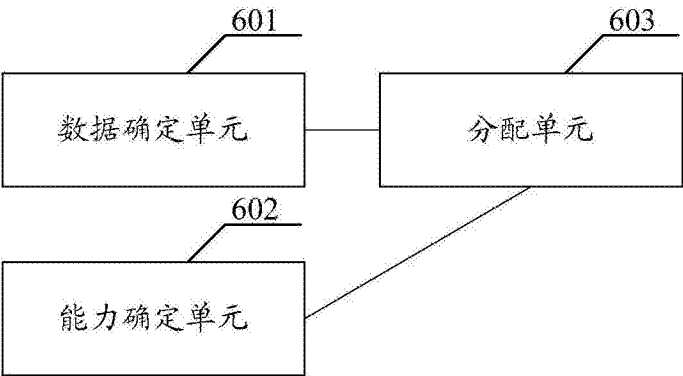


图6

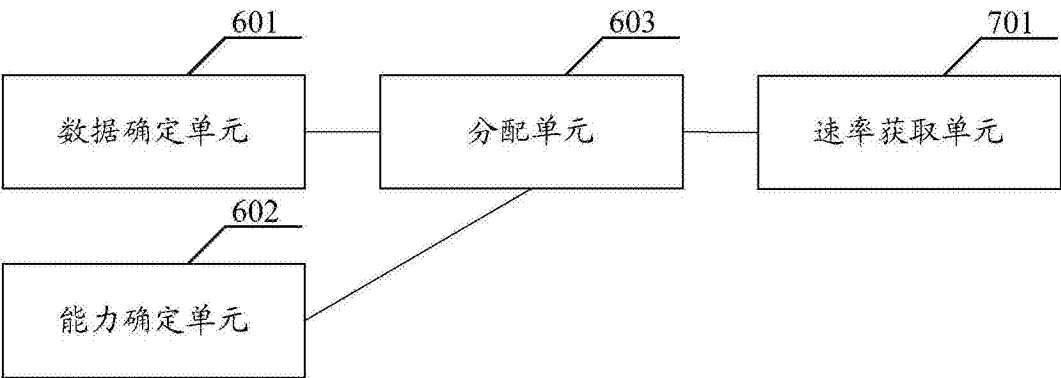


图7

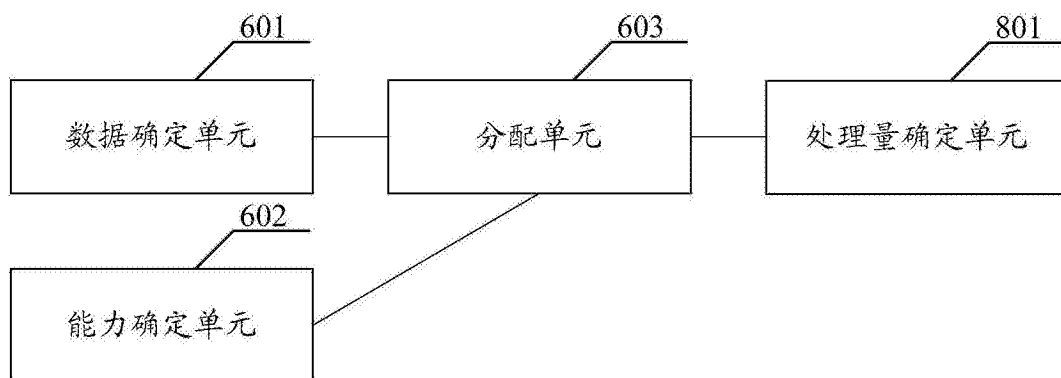


图8

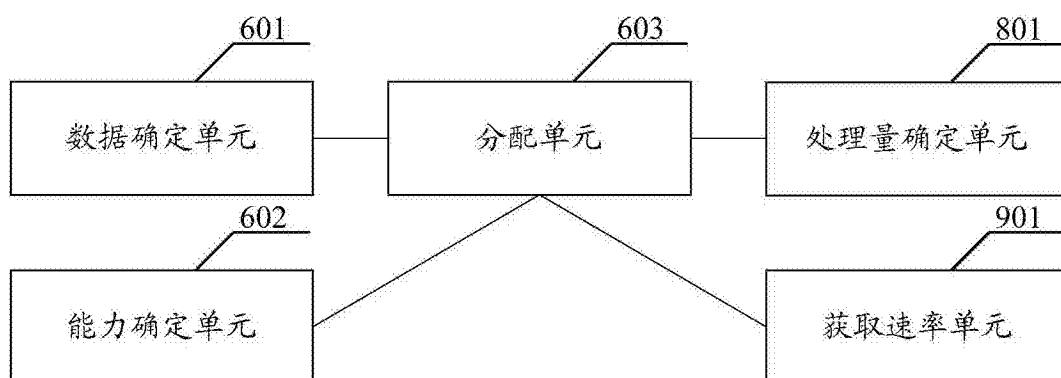


图9

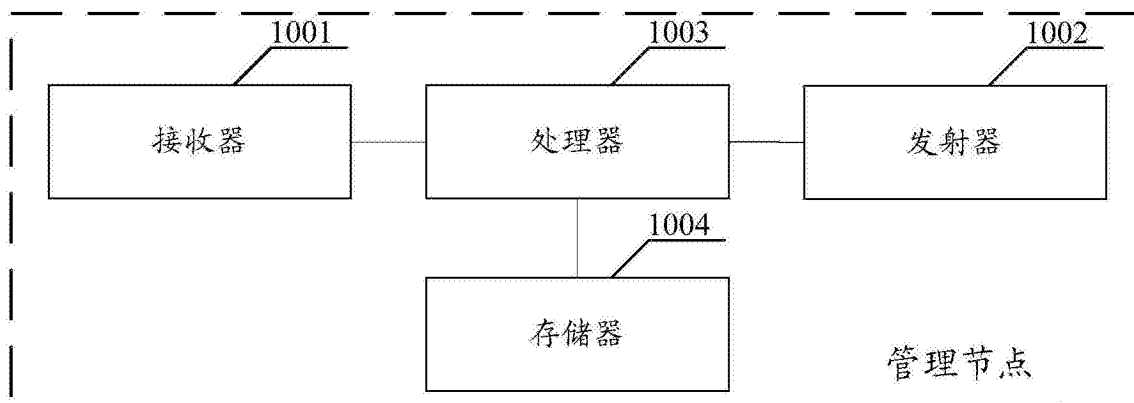


图10