(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 103279351 B
(45)授权公告日 2016.06.29

(21)申请号 201310213482.4

(22)申请日 2013.05.31

(73)专利权人 北京高森明晨信息科技有限公司
地址 100020 北京市朝阳区朝外大街甲六号万通中心C座22A01室

(72)发明人 张鹏 金晨

(74)专利代理机构 北京三高永信知识产权代理有限责任公司 11138
代理人 刘映东

(51)Int.Cl.
G06F 9/44(2006.01)

(56)对比文件
CN 103019853 A,2013.04.03,
CN 102096602 A,2011.06.15,
CN 102739785 A,2012.10.17,
US 2007/0294695 A1,2007.12.20,

审查员 李翠霞

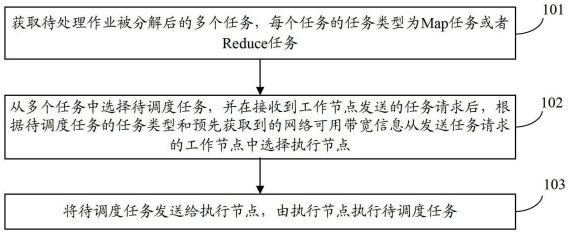
权利要求书3页 说明书13页 附图4页

(54)发明名称

一种任务调度的方法及装置

(57)摘要

本发明公开了一种任务调度的方法及装置，属于计算机领域。所述方法包括：获取待处理作业被分解后的多个任务；从多个任务中选择待调度任务，并在接收到工作节点发送的任务请求后，根据待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点；将待调度任务发送给执行节点，由执行节点执行待调度任务。本发明通过根据待调度任务的任务类型和网络可用带宽信息从发送任务请求的工作节点中选择执行节点，将待调度任务发送给执行节点执行，由于在执行任务时需要在网络中传输大量的数据，根据网络可用带宽信息而选择执行节点能够提高网络中数据传输的速率，从而减少任务执行所需的时间，提高了系统的整体性能。



CN 103279351 B

1.一种任务调度的方法,其特征在于,所述方法包括:

获取待处理作业被分解后的多个任务,每个任务的任务类型为映射Map任务或者化简Reduce任务;

从所述多个任务中选择待调度任务,并在接收到工作节点发送的任务请求后,根据所述待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点;

将所述待调度任务发送给所述执行节点,由所述执行节点执行所述待调度任务;

所述根据所述待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点之前,还包括:

预先获取网络拓扑信息,所述网络拓扑信息至少包括交换机和工作节点之间的连接关系,每个交换机的端口速率和每个工作节点的端口速率;

按照预设周期获取每个工作节点的本地可用带宽和每个交换机的可用端口速率,所述每个工作节点的本地可用带宽包括每个工作节点的本地上行可用带宽和本地下行可用带宽;

根据所述网络拓扑信息、所述每个工作节点的本地可用带宽和所述每个交换机的可用端口速率确定每两个工作节点之间的路径可用带宽;

将所述每个工作节点的本地可用带宽和所述每两个工作节点之间的路径可用带宽确定为网络可用带宽信息。

2.根据权利要求1所述的方法,其特征在于,所述根据所述待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点,包括:

如果所述待调度任务的任务类型为Map任务,根据网络可用带宽信息中每个发送任务请求的工作节点的本地上行可用带宽和所述待调度任务的输入数据的存储位置选择执行节点;

如果所述待调度任务的任务类型为Reduce任务,根据网络可用带宽信息中每两个工作节点之间的路径可用带宽选择执行节点;

其中,所述每两个工作节点之间的路径可用带宽为每个发送任务请求的工作节点与上传所述待调度任务的输入数据的工作节点之间的路径可用带宽。

3.根据权利要求2所述的方法,其特征在于,所述根据网络可用带宽信息中每个发送任务请求的工作节点的本地上行可用带宽和待调度任务的输入数据的存储位置选择执行节点,包括:

确定所述待调度任务的输入数据是否存储在每个发送任务请求的工作节点的本地;

如果所述待调度任务的输入数据存储在每个发送任务请求的工作节点的本地,则根据所述网络可用带宽信息选择将本地上行可用带宽达到第一预设阈值的工作节点作为执行节点;或者,

如果所述待调度任务的输入数据未存储在每个发送任务请求的工作节点的本地,则确定与存储所述待调度任务的输入数据的存储节点之间的路径可用带宽达到第二预设阈值的工作节点,并根据所述网络可用带宽信息从确定的工作节点中选择本地上行可用带宽达到所述第一预设阈值的工作节点作为执行节点。

4.根据权利要求2所述的方法,其特征在于,所述根据网络可用带宽信息中每两个工作

节点之间的路径可用带宽选择执行节点，包括：

根据网络可用带宽信息选择与上传所述待调度任务的输入数据的工作节点之间的路径可用带宽达到第三预设阈值的发送任务请求的工作节点作为执行节点。

5.一种任务调度的装置，其特征在于，所述装置包括：

第一获取模块，用于获取待处理作业被分解后的多个任务，每个任务的任务类型为映射Map任务或者化简Reduce任务；

第一选择模块，用于从所述第一获取模块获取到的多个任务中选择待调度任务；

第二选择模块，用于在接收到工作节点发送的任务请求后，根据所述第一选择模块选择的待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点；

发送模块，用于将所述第一选择模块选择的待调度任务发送给所述第二选择模块选择的执行节点，由所述执行节点执行所述待调度任务；

所述装置，还包括：

第二获取模块，用于预先获取网络拓扑信息，所述网络拓扑信息至少包括交换机和工作节点之间的连接关系，每个交换机的端口速率和每个工作节点的端口速率；

第三获取模块，用于按照预设周期获取每个工作节点的本地可用带宽和每个交换机的可用端口速率，所述每个工作节点的本地可用带宽包括每个工作节点的本地上行可用带宽和本地下行可用带宽；

第一确定模块，用于根据所述第二获取模块获取到的网络拓扑信息和所述第三获取模块获取到的每个工作节点的本地可用带宽和每个交换机的可用端口速率确定每两个工作节点之间的路径可用带宽；

第二确定模块，用于将所述第二获取模块获取到的每个工作节点的本地可用带宽和所述第一确定模块确定的每两个工作节点之间的路径可用带宽确定为网络可用带宽信息。

6.根据权利要求5所述的装置，其特征在于，所述第二选择模块，包括：

第一选择单元，用于在所述待调度任务的任务类型为Map任务时，根据网络可用带宽信息中每个发送任务请求的工作节点的本地上行可用带宽和所述待调度任务的输入数据的存储位置选择执行节点；

第二选择单元，用于在所述待调度任务的任务类型为Reduce任务时，根据网络可用带宽信息中每两个工作节点之间的路径可用带宽选择执行节点；

其中，所述每两个工作节点之间的路径可用带宽为每个发送任务请求的工作节点与上传所述待调度任务的输入数据的工作节点之间的路径可用带宽。

7.根据权利要求6所述的装置，其特征在于，所述第一选择单元，包括：

第一确定子单元，用于确定所述待调度任务的输入数据是否存储在每个发送任务请求的工作节点的本地；

第一选择子单元，用于在所述第一确定子单元确定所述待调度任务的输入数据存储在每个发送任务请求的工作节点的本地时，根据所述网络可用带宽信息选择将本地上行可用带宽达到第一预设阈值的工作节点作为执行节点；

第二确定子单元，用于在所述第一确定子单元确定所述待调度任务的输入数据未存储在每个发送任务请求的工作节点的本地时，确定与存储所述待调度任务的输入数据的存储

节点之间的路径可用带宽达到第二预设阈值的工作节点；

第二选择子单元，用于根据所述网络可用带宽信息从所述第二确定子单元确定的工作节点中选择本地上行可用带宽达到所述第一预设阈值的工作节点作为执行节点。

8.根据权利要求6所述的装置，其特征在于，所述第二选择单元，用于根据网络可用带宽信息选择与上传所述待调度任务的输入数据的工作节点之间的路径可用带宽达到第三预设阈值的发送任务请求的工作节点作为执行节点。

一种任务调度的方法及装置

技术领域

[0001] 本发明涉及计算机领域，特别涉及一种任务调度的方法及装置。

背景技术

[0002] MapReduce(映射化简)系统在数据处理和分析领域得到广泛应用，该系统最大的优点是实现了并行化数据处理，可以自动将待处理作业并行化分解为多个子任务，并调度到服务器集群上并行执行。MapReduce系统包括终端(Client)、调度节点(Master)和多个工作节点(Worker)。其中，客户端用于将待处理作业发送给调度节点；调度节点用于将待处理作业分解为多个任务，每个任务的任务类型可以为Map(映射)任务或Reduce(化简)任务；调度节点还需要进行任务调度，从生成的多个任务中选择待调度任务，并从多个工作节点中选择执行待调度任务的执行节点；执行节点用于执行获取到的待调度任务。其中，Map任务的输出数据将作为Reduce任务的输入数据，Reduce任务的输出数据即为数据处理结果。在MapReduce系统中，调度节点的任务调度是数据处理的核心，调度节点在任务调度时对待调度任务的选择和对执行待调度任务的执行节点的选择将关系到MapReduce系统的整体性能。

[0003] 目前，Facebook(脸书)公司提供了一种任务调度方法：公平调度算法(Fair Scheduler)，将客户端发送的待处理作业分为小作业和大作业，任务调度的原则是保证选择的执行节点能够快速执行小作业的Map任务和Reduce任务，并保证执行节点执行大作业的Map任务和Reduce任务的服务质量。Yahoo(雅虎)公司也提供了一种任务调度方法：计算能力调度算法(Capacity Scheduler)，根据工作节点的计算能力选择执行节点，用于执行待处理作业的Map任务和Reduce任务。

[0004] 在实现本发明的过程中，发明人发现现有技术至少存在以下问题：

[0005] 现有的任务调度方法在任务调度时，仅考虑了执行节点的计算资源，然而执行节点在执行Map任务和Reduce任务时都需要通过网络进行数据传输，如果仅考虑了执行节点的计算资源，将会导致任务执行所需的时间较长，系统的整体性能较差。

发明内容

[0006] 为了解决现有技术的问题，本发明实施例提供了一种任务调度的方法及装置。所述技术方案如下：

[0007] 一方面，提供了一种任务调度的方法，所述方法包括：

[0008] 获取待处理作业被分解后的多个任务，每个任务的任务类型为Map任务或者Reduce任务；

[0009] 从所述多个任务中选择待调度任务，并在接收到工作节点发送的任务请求后，根据所述待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点；

[0010] 将所述待调度任务发送给所述执行节点，由所述执行节点执行所述待调度任务。

[0011] 进一步地，所述根据所述待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点之前，还包括：

[0012] 预先获取网络拓扑信息，所述网络拓扑信息至少包括交换机和工作节点之间的连接关系，每个交换机的端口速率和每个工作节点的端口速率；

[0013] 按照预设周期获取每个工作节点的本地可用带宽和每个交换机的可用端口速率，所述每个工作节点的本地可用带宽包括每个工作节点的本地上行可用带宽和本地下行可用带宽；

[0014] 根据所述网络拓扑信息、所述每个工作节点的本地可用带宽和所述每个交换机的可用端口速率确定每两个工作节点之间的路径可用带宽；

[0015] 将所述每个工作节点的本地可用带宽和所述每两个工作节点之间的路径可用带宽确定为网络可用带宽信息。

[0016] 具体地，所述根据所述待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点，包括：

[0017] 如果所述待调度任务的任务类型为Map任务，根据网络可用带宽信息中每个发送任务请求的工作节点的本地上行可用带宽和所述待调度任务的输入数据的存储位置选择执行节点；

[0018] 如果所述待调度任务的任务类型为Reduce任务，根据网络可用带宽信息中每两个工作节点之间的路径可用带宽选择执行节点；

[0019] 其中，所述每两个工作节点之间的路径可用带宽为每个发送任务请求的工作节点与上传所述待调度任务的输入数据的工作节点之间的路径可用带宽。

[0020] 具体地，所述根据网络可用带宽信息中每个发送任务请求的工作节点的本地上行可用带宽和所述待调度任务的输入数据的存储位置选择执行节点，包括：

[0021] 确定所述待调度任务的输入数据是否存储在每个工作节点的本地；

[0022] 如果所述待调度任务的输入数据存储在每个发送任务请求的工作节点的本地，则根据所述网络可用带宽信息选择将本地上行可用带宽达到第一预设阈值的工作节点作为执行节点；或者，

[0023] 如果所述待调度任务的输入数据未存储在每个发送任务请求的工作节点的本地，则确定与存储所述待调度任务的输入数据的存储节点之间的路径可用带宽达到第二预设阈值的工作节点，并根据所述网络可用带宽信息从确定的工作节点中选择本地上行可用带宽达到所述第一预设阈值的工作节点作为执行节点。

[0024] 具体地，所述根据网络可用带宽信息中每两个工作节点之间的路径可用带宽选择执行节点，包括：

[0025] 根据网络可用带宽信息选择与上传所述待调度任务的输入数据的工作节点之间的路径可用带宽达到第三预设阈值的发送任务请求的工作节点作为执行节点。

[0026] 另一方面，提供了一种任务调度的装置，所述装置包括：

[0027] 第一获取模块，用于获取待处理作业被分解后的多个任务，每个任务的任务类型为映射Map任务或者化简Reduce任务；

[0028] 第一选择模块，用于从所述第一获取模块获取到的多个任务中选择待调度任务；

[0029] 第二选择模块，用于在接收到工作节点发送的任务请求后，根据所述第一选择模

块选择的待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点；

[0030] 发送模块，用于将所述第一选择模块选择的待调度任务发送给所述第二选择模块执行节点，由所述执行节点执行所述待调度任务。

[0031] 进一步地，所述装置，还包括：

[0032] 第二获取模块，用于预先获取网络拓扑信息，所述网络拓扑信息至少包括交换机和工作节点之间的连接关系，每个交换机的端口速率和每个工作节点的端口速率；

[0033] 第三获取模块，用于按照预设周期获取每个工作节点的本地可用带宽和每个交换机的可用端口速率，所述每个工作节点的本地可用带宽包括每个工作节点的本地上行可用带宽和本地下行可用带宽；

[0034] 第一确定模块，用于根据所述第二获取模块获取到的网络拓扑信息和所述第三获取模块获取到的每个工作节点的本地可用带宽和每个交换机的可用端口速率确定每两个工作节点之间的路径可用带宽；

[0035] 第二确定模块，用于将所述第二获取模块获取到的每个工作节点的本地可用带宽和所述第一确定模块确定的每两个工作节点之间的路径可用带宽确定为网络可用带宽信息。

[0036] 具体地，所述第二选择模块，包括：

[0037] 第一选择单元，用于在所述待调度任务的任务类型为Map任务时，根据网络可用带宽信息中每个发送任务请求的工作节点的本地上行可用带宽和所述待调度任务的输入数据的存储位置选择工作节点；

[0038] 第二选择单元，用于在所述待调度任务的任务类型为Reduce任务时，根据网络可用带宽信息中每两个工作节点之间的路径可用带宽选择执行节点；

[0039] 其中，所述每两个工作节点之间的路径可用带宽为每个发送任务请求的工作节点与上传所述待调度任务的输入数据的工作节点之间的路径可用带宽。

[0040] 具体地，所述第一选择单元，包括：

[0041] 第一确定子单元，用于确定所述待调度任务的输入数据是否存储在每个工作节点的本地；

[0042] 第一选择子单元，用于在所述第一确定子单元确定所述待调度任务的输入数据存储在每个发送任务请求的工作节点的本地时，根据所述网络可用带宽信息选择将本地上行可用带宽达到第一预设阈值的工作节点作为执行节点；

[0043] 第二确定子单元，用于在所述第一确定子单元确定所述待调度任务的输入数据未存储在每个发送任务请求的工作节点的本地时，确定与存储所述待调度任务的输入数据的存储节点之间的路径可用带宽达到第二预设阈值的工作节点；

[0044] 第二选择子单元，用于根据所述网络可用带宽信息从所述第二确定子单元确定的工作节点中选择本地上行可用带宽达到所述第一预设阈值的工作节点作为执行节点。

[0045] 具体地，所述第二选择单元，用于根据网络可用带宽信息选择与上传所述待调度任务的输入数据的工作节点之间的路径可用带宽达到第三预设阈值的发送任务请求的工作节点作为执行节点。

[0046] 本发明实施例提供的技术方案带来的有益效果是：

[0047] 通过根据获取到多个任务中待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点，将待调度任务发送给执行节点执行，由于在执行任务时需要在网络中传输大量的数据，根据网络可用带宽信息而选择执行节点能够提高网络中数据传输的速率，从而减少任务执行所需的时间，提高了系统的整体性能。

附图说明

[0048] 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单地介绍，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，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[0049] 图1是本发明实施例一提供的一种任务调度的方法流程图；

[0050] 图2是本发明实施例二提供的一种任务调度的方法流程图；

[0051] 图3是本发明实施例二提供的一种MapReduce系统的结构示意图；

[0052] 图4是本发明实施例三提供的一种任务调度的装置的结构示意图；

[0053] 图5是本发明实施例三提供的另一种任务调度的装置的结构示意图；

[0054] 图6是本发明实施例三提供的一种第二选择模块的结构示意图；

[0055] 图7是本发明实施例三提供的一种第一选择单元的结构示意图；

[0056] 图8是本发明实施例四提供的一种任务调度的系统结构示意图。

具体实施方式

[0057] 为使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明实施方式作进一步地详细描述。

[0058] 实施例一

[0059] 本发明实施例提供了一种任务调度的方法，参见图1，方法流程包括：

[0060] 101：获取待处理作业被分解后的多个任务，每个任务的任务类型为Map任务或者Reduce任务。

[0061] 102：从多个任务中选择待调度任务，并在接收到工作节点发送的任务请求后，根据待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点。

[0062] 进一步地，根据待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点之前，还包括：

[0063] 预先获取网络拓扑信息，网络拓扑信息至少包括交换机和工作节点之间的连接关系，每个交换机的端口速率和每个工作节点的端口速率；

[0064] 按照预设周期获取每个工作节点的本地可用带宽和每个交换机的可用端口速率，每个工作节点的本地可用带宽包括每个工作节点的本地上行可用带宽和本地下行可用带宽；

[0065] 根据网络拓扑信息、每个工作节点的本地可用带宽和每个交换机的可用端口速率确定每两个工作节点之间的路径可用带宽；

[0066] 将每个工作节点的本地可用带宽和每两个工作节点之间的路径可用带宽确定为

网络可用带宽信息。

[0067] 具体地，根据待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点，包括：

[0068] 如果待调度任务的任务类型为Map任务，根据网络可用带宽信息中每个发送任务请求的工作节点的本地上行可用带宽和待调度任务的输入数据的存储位置选择执行节点；

[0069] 如果待调度任务的任务类型为Reduce任务，根据网络可用带宽信息中每两个工作节点之间的路径可用带宽选择执行节点；

[0070] 其中，每两个工作节点之间的路径可用带宽为每个发送任务请求的工作节点与上传待调度任务的输入数据的工作节点之间的路径可用带宽。

[0071] 具体地，根据网络可用带宽信息中每个发送任务请求的工作节点的本地上行可用带宽和待调度任务的输入数据的存储位置选择执行节点，包括：

[0072] 确定待调度任务的输入数据是否存储在每个发送任务请求的工作节点的本地；

[0073] 如果待调度任务的输入数据存储在每个发送任务请求的工作节点的本地，则根据网络可用带宽信息选择将本地上行可用带宽达到第一预设阈值的工作节点作为执行节点；或者，

[0074] 如果待调度任务的输入数据未存储在每个发送任务请求的工作节点的本地，则确定与存储待调度任务的输入数据的存储节点之间的路径可用带宽达到第二预设阈值的工作节点，并根据网络可用带宽信息从确定的工作节点中选择本地上行可用带宽达到第一预设阈值的工作节点作为执行节点。

[0075] 具体地，根据网络可用带宽信息中每两个工作节点之间的路径可用带宽选择执行节点，包括：

[0076] 根据网络可用带宽信息选择与上传待调度任务的输入数据的工作节点之间的路径可用带宽达到第三预设阈值的发送任务请求的工作节点作为执行节点。

[0077] 103：将待调度任务发送给执行节点，由执行节点执行待调度任务。

[0078] 综上所述，本发明实施例提供的方法，通过根据获取到多个任务中待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点，将待调度任务发送给执行节点执行，由于在执行任务时需要在网络中传输大量的数据，根据网络可用带宽信息而选择执行节点能够提高网络中数据传输的速率，从而减少任务执行所需的时间，提高了系统的整体性能。

[0079] 实施例二

[0080] 为了减少MapReduce系统中执行任务所需的时间，本发明实施例提供了一种任务调度的方法，该方法应用于MapReduce系统，MapReduce系统中包括终端、调度节点和多个工作节点。其中，调度节点可以为服务器，用于管理整个MapReduce系统的任务调度；工作节点为服务器或其他设备，用于执行调度节点发送的任务。结合上述实施例一的内容，参见图2，方法流程包括：

[0081] 201：调度节点获取待处理作业被分解后的多个任务，每个任务的任务类型为Map任务或者Reduce任务。

[0082] 其中，待处理作业可以为并行处理的作业。调度节点可以直接获取到待处理作业被分解后的多个任务；也可以先获取终端发送的待处理作业，并将待处理作业分解为多个

任务。在待处理作业被分解后的多个任务中，每个任务的任务类型为Map任务或者Reduce任务。待处理作业可以为一个或多个，对于每个待处理作业，都可以分解得到多个任务。将待处理作业分解为多个任务的方法与现有技术相同，在此不再赘述。多个任务中可以有多个Map任务和多个Reduce任务，多个Map任务具有相同的处理功能，但每个Map任务处理的数据不同，即每个Map任务的输入数据不同，且每个Map任务的输入数据为待处理作业需要处理的数据的一部分；多个Reduce任务也具有相同的处理功能，但每个Reduce任务处理的数据也不同，即每个Reduce任务的输入数据不同，且每个Reduce任务的输入数据为至少一个Map任务的输出数据。

[0083] 举例来说，在如图3所示的MapReduce系统中，包括终端、调度节点和工作节点A-E。终端将待处理作业发送给调度节点，调度节点在获取到待处理作业后，将待处理作业分解为5个任务，分别为Map任务1、Map任务2、Map任务3，Reduce任务1和Reduce任务2。

[0084] 需要说明的是，调度节点在将待处理作业分解为多个任务后，需要进行任务调度，将多个任务分配给多个工作节点执行，具体的任务调度方法参见以下步骤202-204。

[0085] 202：从多个任务中选择待调度任务。

[0086] 针对该步骤，调度节点在执行任务调度时，首先需要从多个任务中选择待调度任务，调度节点从多个任务中选择待调度任务的方法可以根据实际情况配置的，具体选择的方法包括但不限于：根据队列配置选择待调度任务。其中，队列配置是指调度节点中预先配置好的任务排序信息，按照队列排序选择任务。如果调度节点接收到的待处理作业为多个，还可以根据待处理作业优先级选择待调度任务，先选择出优先级较高的待处理作业，然后从待处理作业的多个任务中选择待调度任务。此外，调度节点从多个任务中选择待调度任务的方法还可以包括：对执行节点执行失败的任务重新调度；对瓶颈任务（执行难度较大的任务）冗余调度，即将瓶颈任务分配到多个执行节点上执行。

[0087] 举例来说，仍以上述步骤201中将待处理作业分解为5个任务为例，依据队列配置，从5个任务中选择Map任务1作为待调度任务。

[0088] 调度节点在从多个任务中选择待调度任务时，可以采用上述方法中的一种或综合采用上述多种方法从多个任务中选择待调度任务。除了上述方法之外，还可以采用其他方法从多个任务中选择待调度任务，例如，从多个任务中随机选择待调度任务。对于从多个任务中选择待调度任务的方法，本发明实施例在此不进行具体限定。

[0089] 203：在接收到工作节点发送的任务请求后，根据待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点。

[0090] 针对该步骤，调度节点在选择待调度任务之后，还需要从系统的多个工作节点中选择执行待调度任务的执行节点。在MapReduce系统中，每个工作节点中预先部署有配置文件，配置文件中记录该工作节点的所能执行的最大任务数。每个工作节点会实时判断当前执行的任务数是否达到配置文件中记录的最大任务数，如果没有达到配置文件中记录的最大任务数，则向调度节点发送任务请求，以请求执行新的任务。调度节点在接收到工作节点的任务请求后，根据预先选择的待调度任务的类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点。当然，调度节点也可以记录每个工作节点所能执行的最大任务数，并记录工作节点当前执行的任务数，从而无需接收工作节点发送的任务请求，而是直接从当前执行的任务数未达到最大任务数的工作节点中选择执行节点。

[0091] 在MapReduce系统中，终端、调度节点和多个工作节点通过交换机组成一个网络，具有一定的网络拓扑结构。在执行待调度任务时，执行待调度任务时所需的数据在网络中传输，且数据传输的速度与网络可用带宽信息有很大关系。为了提高执行节点任务执行的速度，进而提高MapReduce系统的性能，在从发送任务请求的工作节点中选择执行节点时，可以将网络可用带宽信息作为选择的依据。

[0092] 在根据待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点之前，还需要预先获取网络可用带宽信息，具体包括：预先获取网络拓扑信息，网络拓扑信息至少包括交换机和工作节点之间的连接关系，每个交换机的端口速率和每个工作节点的端口速率；按照预设周期获取每个工作节点的本地可用带宽和每个交换机的可用端口速率，每个工作节点的本地可用带宽包括每个工作节点的本地上行可用带宽和本地下行可用带宽；根据网络拓扑信息、每个工作节点的本地可用带宽和每个交换机的可用端口速率确定每两个工作节点之间的路径可用带宽；将每个工作节点的本地可用带宽和每两个工作节点之间的路径可用带宽确定为网络可用带宽信息。

[0093] 其中，网络拓扑信息中交换机和工作节点之间的连接关系能够表示网络的拓扑情况，工作节点和工作节点之间由交换机连接，一个工作节点的端口连接到一个交换机的端口。交换机的端口速率表示该交换机的端口所能传输数据的最大速率，工作节点的端口速率为该工作节点的端口所能传输数据的最大速率。交换机的端口速率和工作节点的端口速率将会影响到数据从一个工作节点传输到另一个工作节点的速率。

[0094] 此外，工作节点的本地可用带宽也影响到工作节点传输数据的速率，因此，还需要获取每个工作节点的本地可用带宽。工作节点的本地可用带宽为该工作节点当前的能够用于传输数据的带宽。例如，工作节点A的本地带宽为100Mbps，当前已使用40Mbps，则该工作节点A的本地可用带宽为60Mbps。每个工作节点的本地可用带宽包括每个工作节点的本地上行可用带宽和本地下行可用带宽，本地上行可用带宽是指该工作节点的本地网卡的上行可用速率，本地下行可用带宽是指该工作节点的本地网卡的下行可用速率。同样地，由于两个工作节点之间在传输数据时，需要由交换机进行中转，交换机的可用端口速率也影响到工作节点传输数据的速率，因此还需要获取每个交换机的可用端口速率。每个交换机的可用端口速率为该交换机的端口当前能够传输数据的速率。由于交换机可以有一个或多个端口，当交换机有多个端口时，需要获取交换机的每个可用端口速率。在按照预设周期获取每个工作节点的本地可用带宽和每个交换机的可用端口速率时，可以通过SNMP(Simple Network Management Protocol，简单网络管理协议)的方式来收集。预设周期可以根据实际情况设定，例如，可以设定为2s、3s或者4s等，本发明实施例在此不对预设周期进行具体限定。

[0095] 在获取到网络拓扑信息、每个工作节点的本地可用带宽和每个交换机的可用端口速率后，可以根据网络拓扑信息、每个工作节点的本地可用带宽和每个交换机的可用端口速率确定每两个工作节点之间的路径可用带宽。在网络拓扑结构中，从一个工作节点到另一个工作节点的路径中需要经过一个或多个交换机。因此，在确定每两个工作节点之间的路径可用带宽时，需要综合考虑两个工作节点的本地可用带宽、两个工作节点之间经过的交换机的数量、交换机的可用端口速率。

[0096] 当然，如果网络中工作节点较多，确定每两个工作节点之间的路径可用带宽的工

作量较大，则可以预先网络中每个工作节点与相邻的交换机的端口之间的链路可用带宽以及相邻两个交换机的端口之间的链路可用带宽，之后当需要确定其中任意两个工作节点之间的路径可用带宽时，再根据该任意两个工作节点之间的路径中，每个工作节点与相邻的交换机的端口之间的链路可用带宽以及相邻两个交换机的端口之间的链路可用带宽来确定两个工作节点之间的链路可用带宽。每个工作节点与相邻的交换机的端口之间的链路可用带宽以及相邻两个交换机的端口之间的链路可用带宽的确定方法较为简单，只要根据每个工作节点的本地可用带宽和交换机的可用端口速率确定即可。例如，以工作节点S1到S2的路径为S1→R1→R2→R3→S2为例，其中，S1至S2为工作节点，R1至R3为交换机。可以预先确定S1→R1，R1→R2，R2→R3，R3→S2的链路可用带宽。当需要确定工作节点S1→S2的路径可用带宽时，可以根据工作节点S1→S2的路径经过的交换机中，S1→R1，R1→R2，R2→R3，R3→S2的链路可用带宽确定工作节点S1→S2的路径可用带宽。

[0097] 在将每个工作节点的本地可用带宽和每两个工作节点之间的路径可用带宽确定为网络可用带宽信息之后，基于确定的网络可用带宽信息，可以进行根据待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点的操作。

[0098] 由于待调度任务的任务类型可以为Map任务或者Reduce任务，对于不同的任务类型，选择执行节点的方法也不同，具体分为以下方法一和方法二：

[0099] 方法一：如果待调度任务的任务类型为Map任务，根据网络可用带宽信息中每个发送任务请求的工作节点的本地上行可用带宽和待调度任务的输入数据的存储位置选择执行节点。

[0100] 具体地，根据网络可用带宽信息中每个发送任务请求的工作节点的本地上行可用带宽和待调度任务的输入数据的存储位置选择工作节点的具体方式包括：确定待调度任务的输入数据是否存储在每个发送任务请求的工作节点的本地；如果待调度任务的输入数据存储在每个发送任务请求的工作节点的本地，则根据网络可用带宽信息选择将本地上行可用带宽达到第一预设阈值的工作节点作为执行节点；或者，如果待调度任务的输入数据未存储在每个发送任务请求的工作节点的本地，则确定与存储待调度任务的输入数据的存储节点之间的路径可用带宽达到第二预设阈值的工作节点，并根据网络可用带宽信息从确定的工作节点中选择本地上行可用带宽达到第一预设阈值的工作节点作为执行节点。

[0101] 在MapReduce系统中，待处理作业需要处理的数据预先被划分为多个数据块，采用分布式存储的方式存储在网络中，并由调度节点记录每个数据块的存储位置。每个数据块可能存储在工作节点中，也可能存储在存储节点中。因此，如果待调度任务的任务类型为Map任务，由于其输入数据为多个数据块中的一个，为了避免在执行待调度任务时获取输入数据所需要的传输时间，优先考虑待调度任务的输入数据的本地性。即先确定待调度任务的输入数据是否存储在每个发送任务请求的工作节点的本地，优先选择待调度任务的输入数据存储在本地的发送任务请求的工作节点；如果待调度任务的输入数据都未存储在每个发送任务请求的工作节点的本地，例如，待调度任务的输入数据存储在未发送任务请求的工作节点或存储在其他存储节点，则应当选择与存储待调度任务的输入数据的存储节点的之间的路径可用带宽达到第二预设阈值的工作节点，保证选择的发送任务请求的工作节点和存储待调度任务的输入数据的路径可用带宽较大。确定存储待调度任务的输入数据的节

点和每个发送任务请求的工作节点之间的路径可用带宽的具体方法与确定每两个工作节点的路径可用带宽的方法相同，在此不再赘述。其中，第二预设阈值可以根据实际情况设定，对于第二预设阈值的具体大小，本发明实施例在此不进行具体限定。

[0102] 此外，由于Map任务的输出数据还需要上传到执行相应的Reduce任务的工作节点，作为该Reduce任务的输入数据，因此，在优先考虑了待调度任务的输入数据的本地性之后，还需要考虑工作节点的本地上行可用带宽。如果确定的存储待调度任务的输入数据的工作节点有多个，或者是与存储待调度任务的输入数据的节点的路径可用带宽达到第二预设阈值的工作节点有多个，则需要从中选择本地上行可用带宽满足第一预设阈值的工作节点作为执行节点。当然，如果满足第一预设阈值的工作节点有多个，则可以选择其中任意一个工作节点作为执行节点或者选择其中本地上行可用带宽最大的工作节点作为执行节点。

[0103] 举例来说，仍以如图3所示的MapReduce系统，步骤202中选择的待调度任务为Map任务1，且工作节点A至C都发送任务请求给调度节点为例，由于待调度任务的任务类型为Map任务，则根据网络可用带宽信息中每个发送任务请求的工作节点的本地上行可用带宽和待调度任务的输入数据的存储位置选择工作节点，又由于Map任务1的输入数据存储在工作节点A的本地，且工作节点A的本地上行可用带宽满足第一预设阈值，因此选择工作节点A为执行Map任务1的执行节点。

[0104] 对于待调度任务的输入数据的本地性和工作节点的本地上行可用带宽两个因素中，除了上述可以先考虑待调度任务的输入数据的本地性，再考虑工作节点的本地上行可用带宽的方式之外，还也可以先考虑工作节点的本地上行可用带宽，再考虑待调度任务的输入数据的本地性；或者，综合考虑待调度任务的输入数据的本地性和工作节点的本地上行可用带宽。对于具体地如何根据待调度任务的输入数据的本地性和工作节点的本地上行可用带宽这两个因素选择工作节点，本发明实施例在此不进行具体限定。

[0105] 方法二：如果待调度任务的任务类型为Reduce任务，根据网络可用带宽信息中每两个工作节点之间的路径可用带宽选择执行节点；其中，每两个工作节点之间的路径可用带宽为每个发送任务请求的工作节点与上传待调度任务的输入数据的工作节点之间的路径可用带宽。

[0106] 具体地，根据网络可用带宽信息选择与上传待调度任务的输入数据的工作节点之间的路径可用带宽达到第三预设阈值的发送任务请求的工作节点作为执行节点。

[0107] 如果待调度任务的任务类型为Reduce任务，其输入数据为Map任务的输出数据。执行Map任务的工作节点在执行完毕之后，还需要将执行Map任务的输出数据作为待调度任务的输入数据，上传给执行待调度任务的工作节点。因此，为了保证任务执行的速度，需要与上传待调度任务的输入数据的工作节点之间的路径可用带宽达到第三预设阈值的发送任务请求的工作节点作为执行节点，即执行节点与上传待调度任务的输入数据的工作节点之间的路径可用带宽较大，从而保证待调度任务的输入数据的传输速率较快，以提高任务执行的速度。如果待调度任务的输入数据来自多个工作节点，则可以计算每个发送任务请求的工作节点和上传待调度任务的输入数据的多个工作节点之间的平均路径可用带宽，并选择平均路径可用带宽达到第三预设阈值的发送任务请求的工作节点。其中，第三预设阈值可以根据实际情况设定，对于第三预设阈值的具体大小，本发明实施例在此不进行具体限定。当然，如果有多个发送任务请求的工作节点都达到第三预设阈值，则可以选择其中任意

一个发送任务请求的工作节点作为执行节点或者选择其中平均路径可用带宽最大的发送任务请求的工作节点作为执行节点。

[0108] 举例来说，仍以MapReduce系统为如图3所示的系统，步骤202中选择的待调度任务为Reduce任务1，且工作节点D和E发送任务请求为例，由于待调度任务的任务类型为Reduce任务，且待调度任务的输入数据来自工作节点A、B、C，则根据网络可用带宽信息确定工作节点D和工作节点A、B、C之间的平均路径可用带宽，工作节点E和工作节点A、B、C之间的平均路径可用带宽，由于工作节点D和工作节点A、B、C之间的平均路径可用带宽达到第三预设阈值，则选择工作节点D为执行Reduce任务1的执行节点。

[0109] 需要说明的是，针对现有技术的任务调度方法中，由于调度节点在选择执行节点时仅考虑执行节点的计算资源，从而导致任务执行的速度较慢的问题。而本发明实施例提供的方法，在调度节点在选择执行节点时考虑了网络可用带宽信息，由于MapReduce系统在执行任务时需要传输大量的数据，根据网络可用带宽信息而执行节点能够缩短数据传输的时间，很大程度地降低了执行任务的时间，提高了MapReduce系统的性能。当然，本发明实施例在选择执行节点时，在考虑网络可用带宽信息的同时，还可以综合考虑执行节点的计算资源，使得执行节点不仅网络可用带宽较好，且计算能力较强，从而更大程度地提高MapReduce系统的性能。

[0110] 204：将待调度任务发送给执行节点，由执行节点执行待调度任务。

[0111] 在该步骤中，执行节点在执行待调度任务时，根据待调度任务的不同而采用不同的执行方法。例如，如果待调度任务的任务类型为Map任务，对于不同的Map任务，需要获取不同的输入数据；执行节点在执行完毕之后，还需要将Map任务的输出数据作为Reduce任务的输入数据，上传到执行Reduce任务的执行节点。如果待调度任务的任务类型为Reduce任务，需要接收执行Map任务的执行节点上传的数据作为输入数据，用以执行该待调度任务；执行节点在执行完毕之后，可以将输出数据上传到分布式存储系统中。

[0112] 执行节点在执行完毕待调度任务之后，还可以将执行状态汇报给调度节点，执行状态包括执行成功或者执行失败。调度节点根据执行节点汇报的执行状态确定待调度任务是否执行成功，如果待调度任务执行失败，则需要继续为该调度任务选择执行节点，用于继续执行该调度任务。当然，如果调度节点在预设时间内没有接收到执行节点汇报的执行状态，则可以直接确定该调度任务执行失败，并继续为该调度任务选择执行节点。

[0113] 需要说明的是，对于待处理作业分解得到的多个任务中的每一个任务，均可以采用上述步骤202-204的方法选择待调度任务、选择执行节点、由执行节点执行待调度任务。当待处理作业分解得到的多个任务被调度且执行完毕，则该待处理作业处理完毕。其中，在步骤202中选择待调度任务时，对于一个待处理作业的多个任务来说，由于多个任务包括Map任务和Reduce任务，且Reduce任务的输入数据为Map任务的输出数据，即需要在执行Map任务并得到Map任务的输出数据后，Reduce任务才能执行。因此，在选择待调度任务时，调度节点也可以先将多个任务中的Map任务依次选择为待调度任务，并通过后续步骤203-204执行完毕之后，再继续将多个任务中的Reduce任务依次选择为待调度任务，并通过步骤203-204执行。当然，调度节点也可以首次选择多个任务中的Map任务为待调度任务，后续随机选择多个任务中的Map任务或是Reduce任务为待调度任务，也能够正常执行完毕待调度作业的多个任务。

[0114] 举例来说，在如图3所示的MapReduce系统中，采用上述步骤202至步骤204的方法进行任务调度，并且采用先调度Map任务后调度Reduce任务的策略。首先将Map任务1调度给工作节点A执行，且执行Map任务1所需的输入数据为数据块1；将Map任务2调度给工作节点B执行，且执行Map任务2所需的输入数据为数据块2；将Map任务3调度给工作节点C执行，且执行Map任务3所需的输入数据为数据块3。

[0115] 在Map任务调度完成之后，开始调度Reduce任务，将Reduce任务1调度给工作节点D执行，且执行Reduce任务1所需的输入数据为Map任务1至Map任务3的输出数据1至输出数据3；将Reduce任务2调度给工作节点E执行，且执行Reduce任务2所需的输入数据为Map任务1至Map任务3的输出数据1至输出数据3。在执行完毕之后，将Reduce任务1和Reduce任务2的输出数据1和输出数据2存储到分布式存储系统中，MapReduce系统处理完毕该待处理作业。

[0116] 在具体实施时，调度节点可以包括任务调度的装置和存储装置，任务调度装置用于执行上述步骤201-204的任务调度的方法，存储装置用于存储执行上述步骤201-204的任务调度的方法的操作指令，还可以用于存储任务调度的过程中所需的数据。

[0117] 综上所述，本发明实施例提供的方法，通过根据获取到多个任务中待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点，将待调度任务发送给执行节点执行，由于在执行任务时需要在网络中传输大量的数据，根据网络可用带宽信息而选择执行节点能够提高网络中数据传输的速率，从而减少任务执行所需的时间，提高了系统的整体性能。

[0118] 实施例三

[0119] 本发明实施例提供了一种任务调度的装置，该装置用于执行上述实施例一或实施例二提供的任务调度的方法。参见图4，该装置包括：

[0120] 第一获取模块401，用于获取待处理作业被分解后的多个任务，将待处理作业分解为多个任务，每个任务的任务类型为Map任务或者Reduce任务；

[0121] 第一选择模块402，用于从第一获取模块401获取到的多个任务中选择待调度任务；

[0122] 第二选择模块403，用于在接收到工作节点发送的任务请求后，根据第一选择模块402选择的待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点；

[0123] 发送模块404，用于将第一选择模块402选择的待调度任务发送给第二选择模块403执行节点，由执行节点执行待调度任务。

[0124] 进一步地，参见图5，该任务调度的装置还包括：

[0125] 第二获取模块405，用于预先获取网络拓扑信息，所述网络拓扑信息至少包括交换机和工作节点之间的连接关系，每个交换机的端口速率和每个工作节点的端口速率；

[0126] 第三获取模块406，用于按照预设周期获取每个工作节点的本地可用带宽和每个交换机的可用端口速率，所述每个工作节点的本地可用带宽包括每个工作节点的本地上行可用带宽和本地下行可用带宽；

[0127] 第一确定模块407，用于根据第二获取模块405获取到的网络拓扑信息和第三获取模块406获取到的每个工作节点的本地可用带宽和每个交换机的可用端口速率确定每两个工作节点之间的路径可用带宽；

[0128] 第二确定模块408，用于将第二获取模块405获取到的每个工作节点的本地可用带宽和第一确定模块407确定的每两个工作节点之间的路径可用带宽确定为网络可用带宽信息。

[0129] 具体地，参见图6，第二选择模块403，包括：

[0130] 第一选择单元4031，用于在待调度任务的任务类型为Map任务时，根据网络可用带宽信息中每个发送任务请求的工作节点的本地上行可用带宽和待调度任务的输入数据的存储位置选择工作节点；

[0131] 第二选择单元4032，用于在待调度任务的任务类型为Reduce任务时，根据网络可用带宽信息中每两个工作节点之间的路径可用带宽选择执行节点；

[0132] 其中，每两个工作节点之间的路径可用带宽为每个发送任务请求的工作节点与上传待调度任务的输入数据的工作节点之间的路径可用带宽。

[0133] 具体地，参见图7，第一选择单元4031，包括：

[0134] 第一确定子单元4031a，用于确定待调度任务的输入数据是否存储在每个发送任务请求的工作节点的本地；

[0135] 第一选择子单元4031b，用于在第一确定子单元4031a确定待调度任务的输入数据存储在每个发送任务请求的工作节点的本地时，根据网络可用带宽信息选择将本地上行可用带宽达到第一预设阈值的工作节点作为执行节点；

[0136] 第二确定子单元4031c，用于在第一确定子单元4031a确定待调度任务的输入数据未存储在每个发送任务请求的工作节点的本地时，确定与存储待调度任务的输入数据的存储节点之间的路径可用带宽达到第二预设阈值的工作节点；

[0137] 第二选择子单元4031d，用于根据网络可用带宽信息从第二确定子单元4031c确定的工作节点中选择本地上行可用带宽达到第一预设阈值的工作节点作为执行节点。

[0138] 具体地，第二选择单元4032，用于根据网络可用带宽信息选择与上传待调度任务的输入数据的工作节点之间的路径可用带宽达到第三预设阈值的发送任务请求的工作节点作为执行节点。

[0139] 综上所述，本发明实施例提供的任务调度的装置，通过根据获取到多个任务中待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点，将待调度任务发送给执行节点执行，由于在执行任务时需要在网络中传输大量的数据，根据网络可用带宽信息而选择执行节点能够提高网络中数据传输的速率，从而减少任务执行所需的时间，提高了系统的整体性能。

[0140] 实施例四

[0141] 本发明实施例提供了一种任务调度的系统，参见图8，该系统包括：终端801、调度节点802、多个工作节点803和存储节点804；

[0142] 终端801，用于将待处理作业发送给调度节点802；

[0143] 调度节点802如上述实施例三所述的任务调度的装置；

[0144] 工作节点803，用于接收调度节点802发送的待调度任务，并执行待调度任务；

[0145] 存储节点804，用于存储任务调度过程中所需的数据。

[0146] 综上所述，本发明实施例提供的系统，通过根据获取到多个任务中待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点，

将待调度任务发送给执行节点执行，由于在执行任务时需要在网络中传输大量的数据，根据网络可用带宽信息而选择执行节点能够提高网络中数据传输的速率，从而减少任务执行所需的时间，提高了系统的整体性能。

[0147] 需要说明的是：上述实施例提供的任务调度的装置在任务调度时，仅以上述各功能模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能模块完成，即将任务调度的装置的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。另外，上述实施例提供的任务调度的装置与任务调度的方法实施例属于同一构思，其具体实现过程详见方法实施例，这里不再赘述。

[0148] 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。

[0149] 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，所述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。

[0150] 以上所述仅为本发明的较佳实施例，并不用以限制本发明，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。

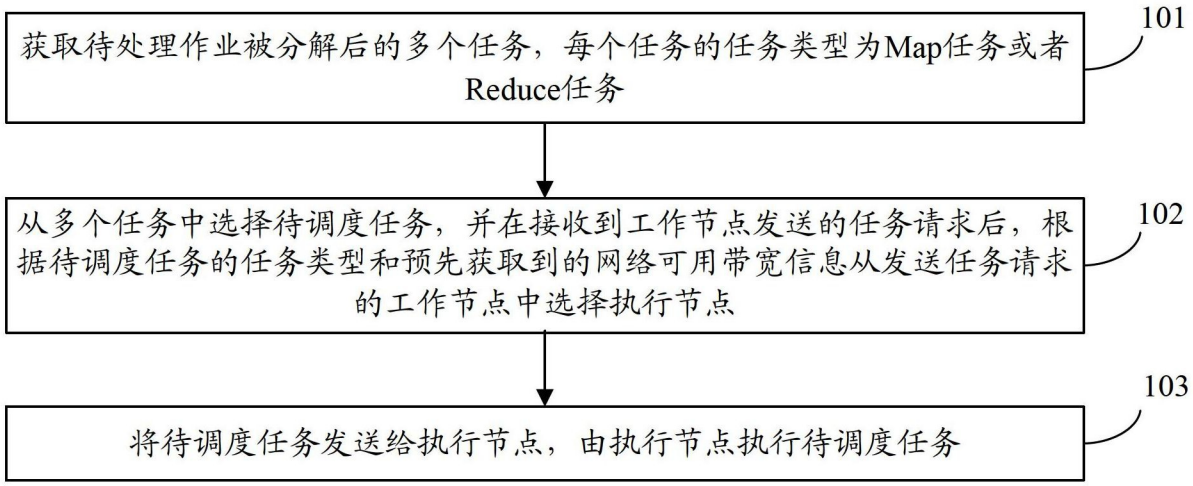

获取待处理作业被分解后的多个任务，每个任务的任务类型为Map任务或者Reduce任务
101
从多个任务中选择待调度任务，并在接收到工作节点发送的任务请求后，根据待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点
102
将待调度任务发送给执行节点，由执行节点执行待调度任务
103

图1

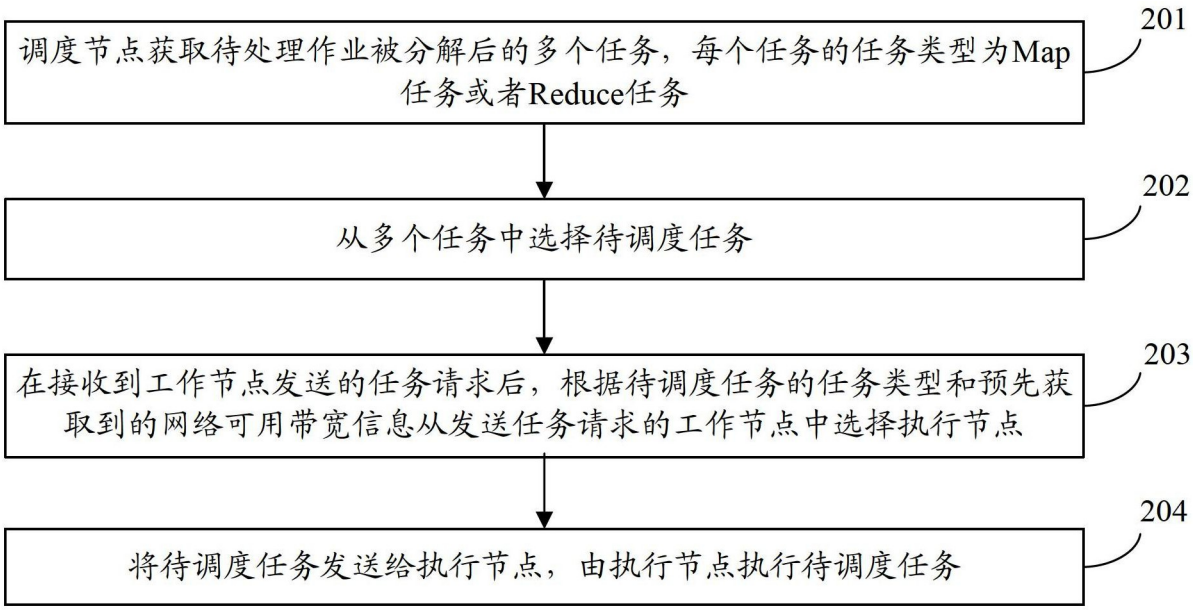

调度节点获取待处理作业被分解后的多个任务，每个任务的任务类型为Map任务或者Reduce任务
201
从多个任务中选择待调度任务
202
在接收到工作节点发送的任务请求后，根据待调度任务的任务类型和预先获取到的网络可用带宽信息从发送任务请求的工作节点中选择执行节点
203
将待调度任务发送给执行节点，由执行节点执行待调度任务
204

图2

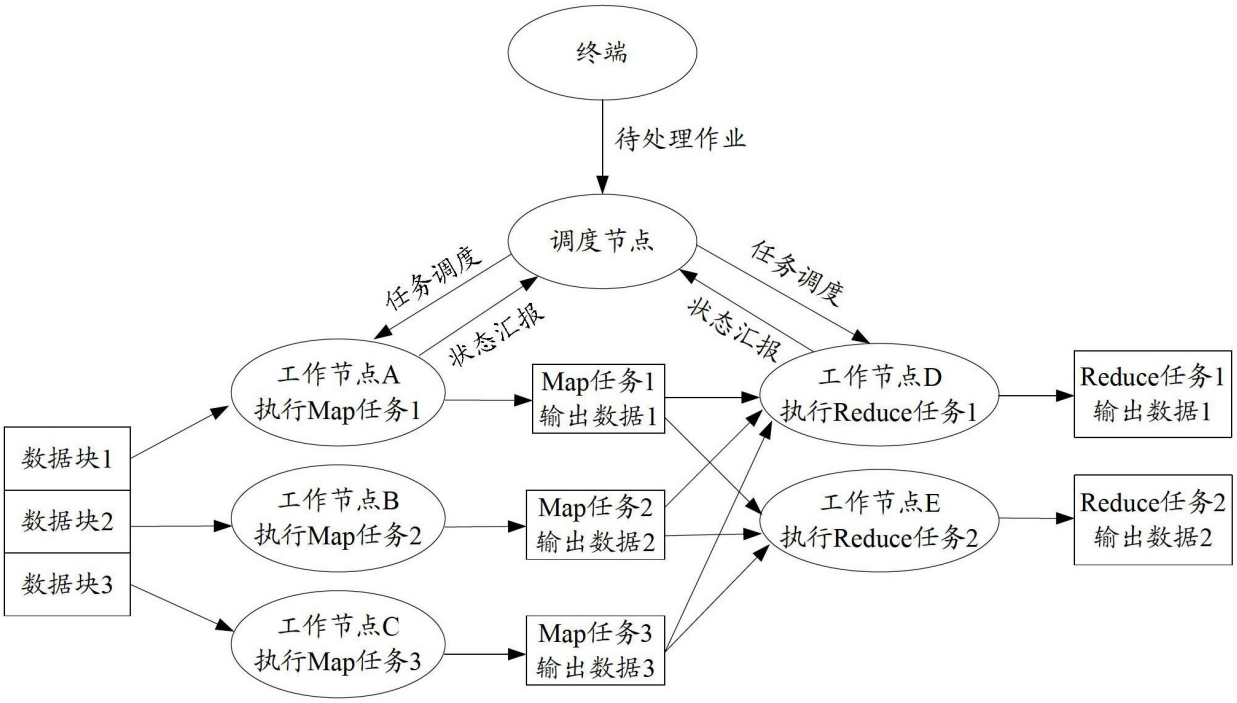

终端
待处理作业
调度节点
任务调度
状态汇报
任务调度
状态汇报
工作节点A
执行Map任务1
Map任务1
输出数据1
工作节点D
执行Reduce任务1
Reduce任务1
输出数据1
数据块1
数据块2
数据块3
工作节点B
执行Map任务2
Map任务2
输出数据2
工作节点E
执行Reduce任务2
Reduce任务2
输出数据2
工作节点C
执行Map任务3
Map任务3
输出数据3

图3

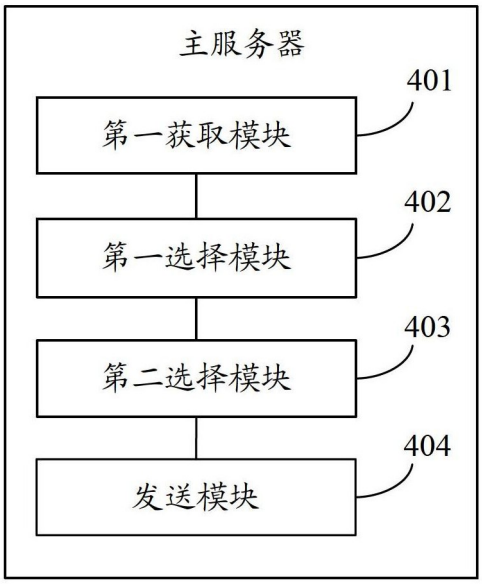

主服务器
401
第一获取模块
402
第一选择模块
403
第二选择模块
404
发送模块

图4

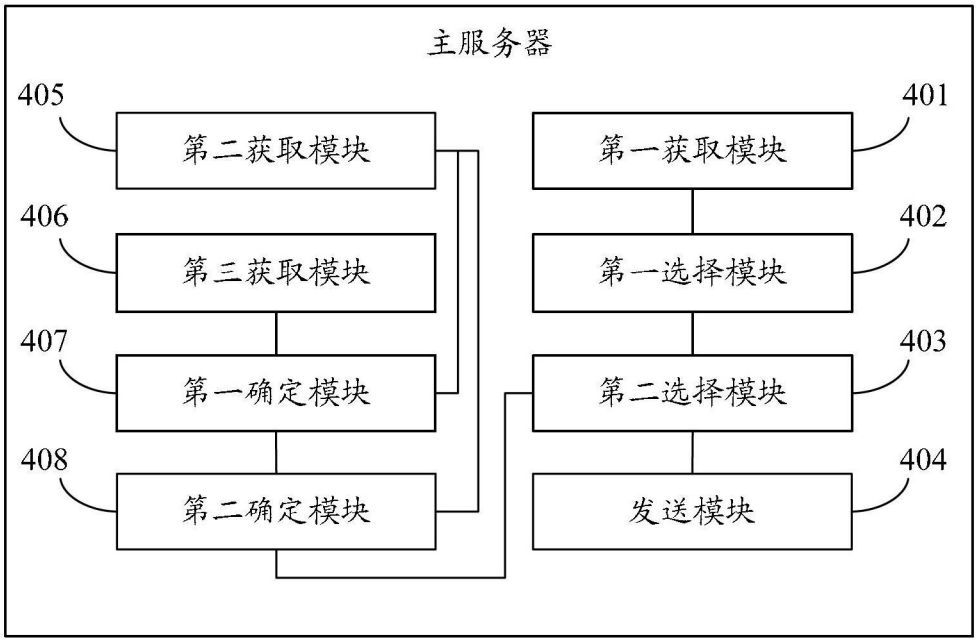

主服务器
405
第二获取模块
401
第一获取模块
406
第三获取模块
402
第一选择模块
407
第一确定模块
403
第二选择模块
408
第二确定模块
404
发送模块

图5

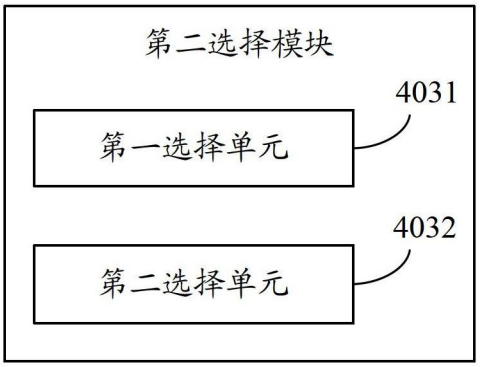

第二选择模块
4031
第一选择单元
4032
第二选择单元

图6

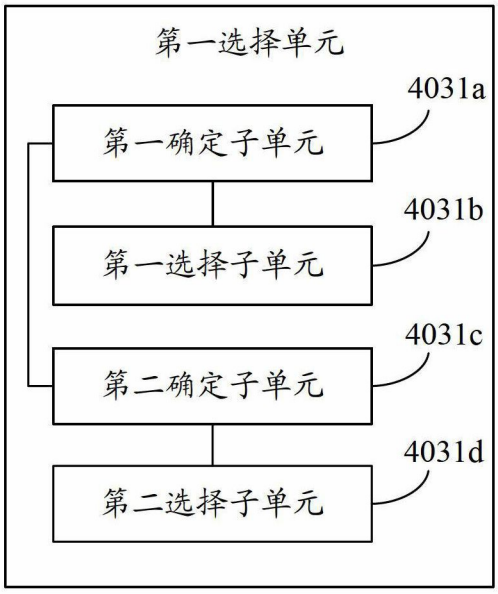

第一选择单元
4031a
第一确定子单元
4031b
第一选择子单元
4031c
第二确定子单元
4031d
第二选择子单元

图7

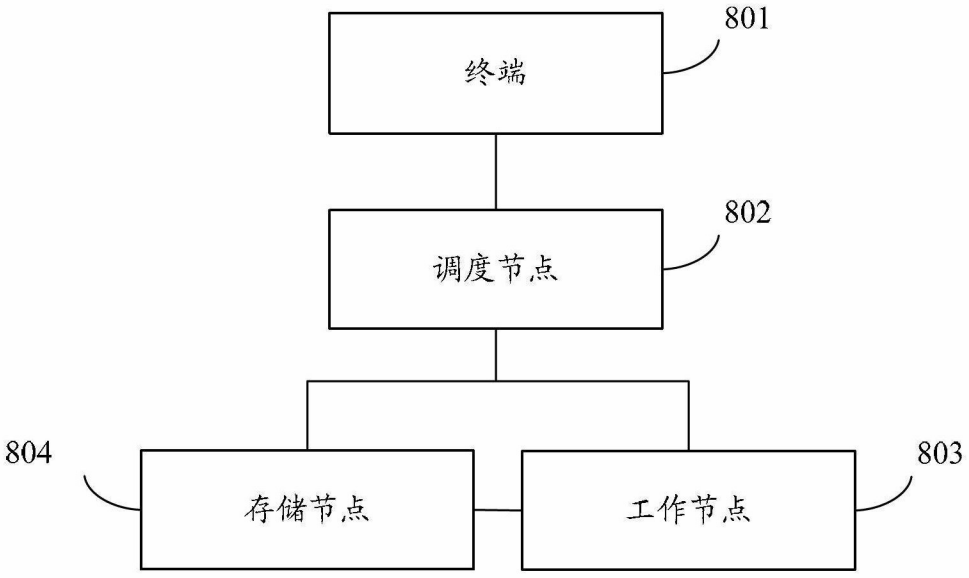


图8