(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2016 年 7 月 7 日 (07.07.2016)

WIPO | PCT

(10) 国际公布号
WO 2016/107610 A1

(51) 国际专利分类号:
G06F 17/30 (2006.01)

(21) 国际申请号: PCT/CN2016/070052

(22) 国际申请日: 2016 年 1 月 4 日 (04.01.2016)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201510003943.4 2015 年 1 月 4 日 (04.01.2015) CN

(71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人: 肖宇雷 (XIAO, Yulei); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 叶涛 (YE, Tao); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(81) 指定国 (除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明，要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

根据细则 4.17 的声明:
— 关于申请人有权申请并被授予专利(细则 4.17(ii))

本国际公布:
— 包括国际检索报告(条约第 21 条(3))。

(54) Title: RESOURCE COORDINATION METHOD, DEVICE, AND SYSTEM FOR DATABASE CLUSTER

(54) 发明名称 : 一种数据库集群的资源协调方法、装置及系统

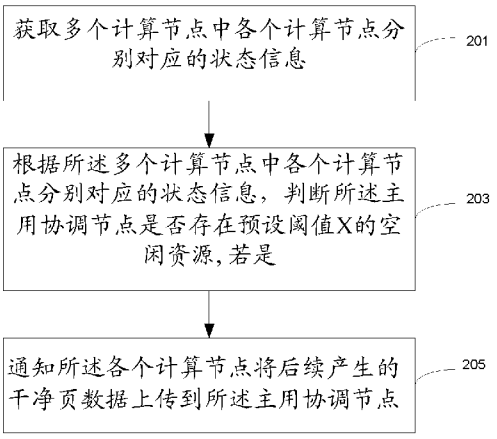


图 2 / Fig. 2

201 Acquire state information corresponding to each computing node in multiple computing nodes
203 Determine, according to the state information corresponding to each computing node in the multiple computing nodes, whether an active coordination node has idle resources with a preset threshold X
205 If yes, instruct each computing node to upload clean page data subsequently generated to the active coordination node

(57) Abstract: A resource coordination method, device, and system for a database cluster. The method comprises: an active coordination node acquires state information corresponding to each computing node in multiple computing nodes (201), the state information being used to indicate the running load state of the computing node; determine, according to the state information corresponding to each computing node in the multiple computing nodes, whether the active coordination node has idle resources with a preset threshold X (203); and if yes, instruct each computing node to upload clean page data subsequently generated to the active coordination node (205). By means of the method, more data generated by a computing node can be stored in a coordination node, and an idle resource on the coordination node and a feature of high speed communication between communication interfaces of the coordination node and the computing node are fully used, so that the computing node is started with a higher speed and higher efficiency.

(57) 摘要: 一种数据库集群的资源协调方法、设备及系统，方法包括：主用协调节点获取多个计算节点中各个计算节点分别对应的状态信息（201），状态信息用于表示计算节点的运行负载状态；并根据多个计算节点中各个计算节点分别对应的状态信息，判断主用协调节点是否存在预设阈值 X 的空闲资源（203）；若是则通知各个计算节点将后续产生的干净页数据上传到主用协调节点（205）。通过这样的方法可以使计算节点产生的数据有更多能够存储在协调节点上，充分利用协调节点上的空闲资源和协调节点与计算节点通讯接口高速通讯的特性，使得计算节点的启动更加快速高效。

WO 2016/107610 A1

一种数据库集群的资源协调方法、装置及系统

技术领域

本发明涉及计算机技术领域，具体涉及一种数据库集群的协调方法、装置及系统。

背景技术

在集群Share-Disk（共享磁盘）架构中，所有集群中的服务器共享一个存储。集群中的服务器按照功能，一般可以分为计算节点（PN）和协调节点（CN）。计算节点一般有多个，用来承担执行具体的功能，执行具体的SQL（结构化查询语言）和事务。当数据库性能需要提升时，可以通过增加计算节点来横向扩展，达到提升集群整体计算能力的目的。同时各个计算节点相互冗余，单个计算节点故障不影响其它计算节点。协调节点一般只有一个，是用来当多个计算节点并发执行时，对竞争的公共资源进行协调。

现有集群 Share-Disk 架构的典型代表是 Oracle RAC（Real Application Clusters）方案。该方案中，各个计算节点服务器通过高速私有网络直接互联，每个计算节点上的资源可以相互共享访问，通过 DML 机制控制并发，该 DML 是指数据锁，用于保护数据的完整性。现有技术方案(如 Oralce RAC)中的协调节点主要实现了不同节点间竞争的公共资源的协调，但是没有考虑到协调节点上存储空间的共享。因此对于只读或者读多写少的场景，协调节点缓存的数据就会非常少，因此无法充分利用协调节点上的存储空间和 Infiniband（协调节点与计算节点通讯接口）高速通讯的特性，存在浪费；从而对于计算节点故障重启，或者新添加计算节点等场景，无法从高速的协调节点获取资源，只能重新从磁阵读取，计算节点无法快速预热，效率低下。

发明内容

为了解决现有技术中计算节点故障重启或者新添加节点等场景下，无法从高速的协调节点快速获取资源，从而使得计算节点无法快速预热导致效率地下的问题，本发明实施例提供的一种数据库集群的资源协调方法，充分利用协调

节点上的空闲资源，从而使得计算节点产生的数据有更多能够存储在协调节点上，充分利用协调节点上的空闲资源和Infiniband（协调节点与计算节点通讯接口）高速通讯的特性，使得计算节点的启动更加快速高效。本发明实施例还提供了相应的装置及系统。

第一方面，本发明实施例提供一种数据库集群的资源协调方法，所述数据库集群包括多个计算节点和一个主用协调节点，所述方法包括：

获取所述多个计算节点中各个计算节点分别对应的状态信息，所述状态信息用于表示计算节点的运行负载状态；

根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点是否存在预设阈值X的空闲资源；若是

则通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点；或者，

根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据；若是

则通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点。

结合第一方面，在第一方面的第一种可能的实施方式中，在所述通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点之后，所述方法还包括：

实时监控所述主用协调节点的空闲资源，当所述主用协调节点的空闲资源小于预设阈值Z，则通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点，其中Z小于X。

结合第一方面的第一种可能的实施方式中，在第一方面的第二种可能的实施方式中，在所述通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点之后，所述方法还包括：

若所述主用协调节点收到新产生的脏页数据；

则优先淘汰所述主用协调节点上缓存的干净页数据。

结合第一方面的第一种可能的实施方式中，在第一方面的第三种可能的实施方式中，在所述通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点之后，所述方法还包括：

若所述主用协调节点收到新产生的脏页数据，且所述新产生的脏页数据的数据量大于所述主用协调节点上缓存的干净页数据的数据量；

则优先淘汰所述主用协调节点上缓存的干净页数据，并将所述新产生的脏页数据存储到所述数据库集群的存储设备上。

结合第一方面，在第一方面的第四种可能的实施方式中，在所述根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据；若是之后，所述方法还包括：

则根据所述多个计算节点中各个计算节点分别对应的状态信息，生成主用协调节点的缓存策略；所述主用协调节点的缓存策略包括：所述多个计算节点中各个计算节点向所述主用协调节点上传干净页数据的策略；

对应地，所述通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点包括：

通知所述各个计算节点将后续产生的干净页数据根据所述主用协调节点的缓存策略上传到所述主用协调节点。

结合第一方面的四种可能的实施方式中，在第一方面的第五种可能的实施方式中，所述根据所述多个计算节点中各个计算节点分别对应的状态信息，生成主用协调节点的缓存策略包括：

根据所述多个计算节点中各个计算节点分别对应的状态信息，计算所述多个计算节点中每一个计算节点当前单位时间脏页数据的产生比例；

根据所述每一个计算节点当前单位时间脏页数据的产生比例，生成主用协调节点的缓存策略。

结合第一方面的第五种可能的实施方式中，在第一方面的第六种可能的实施方式中，所述状态信息包括：所述状态信息对应的计算节点的单位时间事务数量TS、读写比例R/M、CPU利用率CP、IO（输入/输出）利用率IP、以及单

位时间产生脏页数据量DS。

结合第一方面的第六种可能的实施方式，在第一方面的第七种可能的实施方式中，所述根据所述多个计算节点中各个计算节点分别对应的状态信息，计算所述多个计算节点中每一个计算节点当前单位时间脏页数据的产生比例，包括:

$$T=\frac{DS}{\frac{TWS}{JP}*QW}=\frac{DS*(R+W)*(Q1*CP+Q2*IP)}{TS*W*QW};$$

其中，$TWS=\frac{TS*W}{R+W}$；$JP=Q1*CP+Q2*IP$，其中QW为写事务产生的脏页数据权值；其中Q1为CPU利用率的权值，Q2为IO利用率的权值，其中Q1+Q2≤1。

第二方面，本发明实施例提供一种应用于数据库集群的主用协调节点，所述主用协调节点包括:

获取模块，用于获取所述数据库集群中多个计算节点中各个计算节点分别对应的状态信息，所述状态信息用于表示计算节点的运行负载状态;

第一判断模块，用于根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点是否存在预设阈值X的空闲资源;

执行模块，用于当所述第一判断模块判断结果为是时，通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点。

结合第二方面，在第二方面的第一种可能的实施方式中，所述主用协调节点还包括:

第二判断模块，用于当所述第一判断模块判断结果为是时，根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据;

对应地，所述执行模块用于当所述第二判断模块判断结果为是时，通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点。

结合第二方面或第二方面的第一种可能的实施方式，在第二方面的第二种可能的实施方式中，所述主用协调节点还包括监控模块，用于实时监控所述主用协调节点的空闲资源，当所述主用协调节点的空闲资源小于预设阈值Z，则通过所述执行模块通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点，其中Z小于X。

结合第二方面的第二种可能的实施方式中，在第二方面的第三种可能的实施方式中，当所述执行模块通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点之后，若所述主用协调节点收到新产生的脏页数据，则

所述执行模块还用于：优先淘汰所述主用协调节点上缓存的干净页数据。

结合第二方面的第二种可能的实施方式中，在第二方面的第四种可能的实施方式中，当所述执行模块通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点之后，若所述主用协调节点收到新产生的脏页数据，且所述新产生的脏页数据的数据量大于所述主用协调节点上缓存的干净页数据的数据量；则

所述执行模块还用于：则优先淘汰所述主用协调节点上缓存的干净页数据，并将所述新产生的脏页数据存储到所述数据库集群的存储设备上。

结合第二方面的第一种可能的实施方式中，在第二方面的第五种可能的实施方式中，当所述第二判断模块判断结果为是时，所述装置还包括：

缓存策略生成模块：用于根据所述多个计算节点中各个计算节点分别对应的状态信息，生成主用协调节点的缓存策略；所述主用协调节点的缓存策略包括：所述多个计算节点中各个计算节点向所述主用协调节点上传干净页数据的策略；

对应地，所述执行模块用于：通知所述各个计算节点将后续产生的干净页数据根据所述主用协调节点的缓存策略上传到所述主用协调节点。

结合第二方面的第五种可能的实施方式中，在第二方面的第六种可能的实施方式中，所述缓存策略生成模块具体用于：

根据所述多个计算节点中各个计算节点分别对应的状态信息，计算所述多个计算节点中每一个计算节点当前单位时间脏页数据的产生比例；

根据所述每一个计算节点当前单位时间脏页数据的产生比例，生成主用协调节点的缓存策略。

第三方面，本发明实施例提供一种数据库集群，包括多个计算节点和一个主用协调节点，所述主用协调节点获取所述多个计算节点中各个计算节点分别对应的状态信息，并根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点是否存在预设阈值X的空闲资源；若是，则通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点；其中，所述状态信息用于表示计算节点的运行负载状态；

所述各个计算节点接收所述主用协调节点发送的通知，当所述各个计算节点中任意一个计算节点产生干净页数据，则将所述产生的干净页数据上传到所述主用协调节点。

结合第三方面，在第三方面的第一种可能的实施方式中，所述主用协调节点获取所述多个计算节点中各个计算节点分别对应的状态信息，并根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点是否存在预设阈值X的空闲资源；若是之后，再根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据；若是则通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点；

所述各个计算节点接收所述主用协调节点发送的通知，当所述各个计算节点中任意一个计算节点产生干净页数据，则上传到所述主用协调节点。

结合第三方面的第一种可能的实施方式，在第三方面的第二种可能的实施方式中，所述主用协调节点判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据；若是则根据所述多个计算节点中各个计算节点分别对应的状态信息，生成主用协调节点的缓存策略；并通知所述各个计算节点将后续产生的干净页数据根据所述主用协调节点的缓存策略上传到所述主用协调节点；

对应地，所述各个计算节点接收所述主用协调节点发送的通知，当所述各个计算节点中任意一个计算节点产生干净页数据，则将所述产生的干净页数据根据所述主用协调节点的缓存策略上传到所述主用协调节点。

结合第三方面的第一种可能的实施方式或第三方面的第二种可能的实施方式，在第三方面的第三种可能的实施方式中，所述数据库集群还包括备用协调节点，用于为所述主用协调节点进行数据备份，以使得所述备用协调节点上的内容与所述主用协调节点上的内容相同。

本发明实施例根据多个计算节点中各个计算节点分别对应的状态信息，判断主用协调节点是否存在预设阈值 X 的空闲资源，当判断结果为是则通知各个计算节点在接下来的工作中将新产生的干净页数据上传到主用协调节点上来。从而将主用协调节点上的空闲资源充分利用，当计算节点重启或添加新的计算节点时，可以从高速的协调节点上迅速获取资源，充分利用协调节点上的存储空间和 Infiniband 高速通讯的特性，从而使计算节点能够快速预热提高工作效率。

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术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。

图1是本发明实施例提供的数据库集群的硬件结构图；

图2是本发明实施例一提供的一种数据库集群的资源协调方法流程图；

图3是本发明实施例提供的另一种数据库集群的资源协调方法流程图；

图4是本发明实施例提供的又一种数据库集群的资源协调方法流程图；

图5是本发明实施例提供的再一种数据库集群的资源协调方法流程图；

图6是本发明实施例提供的一种应用于数据库集群的主用协调节点的结构示意图；

图7是本发明实施例提供的另一种应用于数据库集群的主用协调节点的结构示意图；

图8是本发明实施例提供的又一种应用于数据库集群的主用协调节点的结构示意图；

图9是本发明实施例提供的再一种应用于数据库集群的主用协调节点的结构示意图；

图 10 是本发明实施例提供的应用于数据库集群中的主用协调节点的结构示意图。

具体实施方式

本发明实施例提供一种数据库集群的资源协调方法，根据多个计算节点中各个计算节点分别对应的状态信息，判断主用协调节点是否存在预设阈值X的空闲资源；若是则通知各个计算节点将后续产生的干净页数据上传到主用协调节点，充分利用协调节点上的存储空间和Infiniband高速通讯的特性，从而使计算节点能够快速预热提高工作效率。本发明实施例还提供了相应的装置及系统。以下分别进行详细说明。

为了使本技术领域的人员更好地理解本发明方案，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分的实施例，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

有其他实施例，都应当属于本发明保护的范围。

本发明实施例中的数据库集群包括：计算节点和主用协调节点，通常这个数据库集群还有备用协调节点，用于备份主用协调节点的数据并当主用协调节点故障的时候来承担主用协调节点的工作。计算节点一般有多个，主要用于承担执行具体的功能，例如执行具体的SQL（结构化查询语言）及具体事务；主用协调节点一般只有一个，当然也不排除多个主用协调节点一起工作的，主用协调节点是用来当多个计算节点并发执行时，对竞争的公共资源进行协调。如图1所示的数据库集群100可以为本发明实施例提供的数据库集群的资源协调方法提供硬件支持，包括主用协调节点102，备用协调节点104，高速网络通道（如：Infiniband，IB）112、计算节点112、114、116、118，光纤通道132，及共享存储磁盘142。在一个实施例中，共享存储磁盘142存储数据库集群100中的数据，计算节点112、114、116、118执行具体的事务，对共享存储磁盘142上存储的数据进行计算或读写操作。主用协调节点102管理计算节点112、114、116、118间的通讯。备用协调节点104在主用协调节点102故障时替代主用协调节点102，完成主用协调节点102在正常状态下的功能。计算节点112、114、116及118之间，主用协调节点102与备用协调节点104之间，以及与计算节点112、114、116、118与主用协调节点102和备用协调节点104之间，均通过高速网络通道112进行通讯，该高速网络通道可以是IB通道。计算节点112、114、116、118与共享存储磁盘142之间通过光纤通道132进行通讯。上述的主用协调节点102、备用协调节点104、计算节点112、114、116、118具体均可以为计算机，因此其内部的硬件结构可以和通用的计算机的硬件结构相同或相似，例如它们都有处理器，都有内存，还可以有硬盘（可以叫做外存）。本发明以下实施例提供的数据库集群的资源协调方法、装置及系统就是在这样的硬件基础上进行的。

实施例一，如图2所示为本发明实施例提供的一种数据库集群的资源协调方法流程图，可以理解的是，该方法流程图可以运行在如图1提供的数据库集群100的硬件基础上。

在步骤201中，主用协调节点获取多个计算节点中各个计算节点分别对应

的状态信息，此处的状态信息用于表示所述状态信息对应的计算节点的运行负载状态。

需要说明的是，当该数据库集群中有备用协调节点时，备用协调节点为了保持和主用协调节点的数据一致，可以和主用协调节点一样，直接从多个计算节点上获取与主用协调节点所获取信息相同的信息，也可以从已经获取这些信息的主用协调节点上去获取这些信息，后面相关处亦是如此，不再赘述。进一步需要说明的是，此处的状态信息指每一个计算节点对应地其本身在执行具体事务过程中的缓存负载等情况，包括但不限于：单位时间事务数量（TS）、读写比例（R/W）、CPU利用率（CP）、IO利用率（IP）和单位时间产生脏页数据量（DS）等数据。

在步骤203中，主用协调节点根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点是否存在预设阈值X的空闲资源。

此处的空闲资源可以指主用协调节点上缓存空间，该缓存空间可以是指主用协调节点的内存，也可以指主用协调节点的内存及外存（即硬盘之类的存储部件）一起构成的存储空间。此处的预设阈值X可以是基于经验值设定的阈值。

基于步骤203的判断，若判断结果为是，则执行步骤205:

在步骤205中，主用协调节点通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点。

即当判断发现该主用协调节点上存在预设阈值X的的空闲资源后便通知各个计算节点在接下来执行具体事务的过程中产生的干净页数据上传到主用协调节点上去。

本发明实施例提供的数据库集群的资源协调方法，根据统筹多个计算节点的运行状态，根据各个计算节点的状态信息判断主用协调节点是否存在空闲资源，如存在则通知计算节点将后续产生的干净页数据上传到主用协调节点上，这样使得主用协调节点在协调各计算节点的竞争资源之余，还能将其多余的空闲资源充分利用来帮助计算几点缓存非竞争资源的干净页数据，从而当计算节点重启或添加新的计算节点时，可以从高速的协调节点上迅速获取数据资源，充分利用协调节点上的存储空间和Infiniband高速通讯的特性，从而使计算节

点能够快速预热提高工作效率。

如图3所示，在图2的基础上，在步骤203判断为是之后又增加步骤204。

在步骤204中，主用协调节点根据多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据。

需要说明的是，此处的预设时间可以是根据经验值来设定，可以是以小时级也可以以天为单位，此处不做限制。

当步骤204判断结果为是，则进一步执行步骤205，即：主用协调节点通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点。

添加步骤204的判断动作后，在实施例一的基础上，经过步骤203和步骤204的两次判断，在确定主用协调节点存在预设阈值X的空闲资源且在预设时间内各个计算节点将上传的脏也数据小于该预设阈值X，则说明在该预设时间内，主用协调节点上仍将存在空闲资源可以用于容纳计算节点的干净页数据，因此进一步确保了主用协调节点上空闲资源的可用性，使本发明实施例提供的方案具备更精确的实用性，并进一步提高了计算节点的工作效率。

如图4所示，在图3及上述实施例的基础上，本发明的一个实施例还可以在步骤205之后添加步骤206。

在步骤206中，主用协调节点实时监控它自己的空闲资源，当它的空闲资源小于预设阈值Z时，则通知各个计算节点不再将产生的干净页数据上传到所述主用协调节点，其中Z小于X。值得说明的是此处的预设阈值Z可以根据经验值或者硬件设备的要求来取得，例如Z可以为主用协调节点上总存储空间的10%，这里的总存储空间可以为主用协调节点的内存空间总值，还可以为主用协调节点上内存空间与硬盘等外存空间的和值，此处仅为举例不做限定。

可选的，在步骤206之后，本发明的另一个实施例还添加步骤207a。

在步骤207a中，若所述主用协调节点收到新产生的脏页数据；则优先淘汰所述主用协调节点上缓存的干净页数据。

或者，本发明的另一个实施例还可以在步骤206之后添加步骤207b。

在步骤207b中，若所述主用协调节点收到新产生的脏页数据，且所述新产

生的脏页数据的数据量大于所述主用协调节点上缓存的干净页数据的数据量；则优先淘汰所述主用协调节点上缓存的干净页数据，并将所述新产生的脏页数据存储到所述数据库集群的存储设备上。

如图5所示，在图3的基础上，本发明的又一个实施例在步骤204判断为是的基础上，又添加步骤204'。

在步骤204'中，根据所述多个计算节点中各个计算节点分别对应的状态信息，生成主用协调节点的缓存策略；所述主用协调节点的缓存策略包括：所述多个计算节点中各个计算节点向所述主用协调节点上传干净页数据的策略；

其中所述的缓存策略可以是由各个计算节点分别对应的缓存策略组成的一个综合策略，用于指示各个计算节点将自身产生的干净页数据在主用协调节点上进行缓存，缓存时间（什么时间开始上传什么时间结束上传）和缓存的量的多少都可以在这个缓存策略中体现。

对应地，步骤205可以进一步为步骤205'：

在步骤205'中，主用协调节点通知所述各个计算节点将后续产生的干净页数据根据所述主用协调节点的缓存策略上传到所述主用协调节点。

可选的，关于上述的缓存策略计算方法可以有很多种，本发明实施例提供一种缓存策略的计算方法，具体可以为：根据所述多个计算节点中各个计算节点分别对应的状态信息，计算所述多个计算节点中每一个计算节点当前单位时间脏页数据的产生比例；

根据所述每一个计算节点当前单位时间脏页数据的产生比例，生成主用协调节点的缓存策略。

需要说明的是，此处的状态信息指每一个计算节点对应地其本身在执行具体事务过程中的缓存负载等情况，包括但不限于：单位时间事务数量（TS）、读写比例（R/W）、CPU利用率（CP）、IO利用率（IP）和单位时间产生脏页数据量（DS）等数据。

进一步可选的，所述计算节点当前单位时间脏页数据的产生比例为T，具体的计算方法可以根据下述方法获取，当然，下述方法只是本发明实施例中提供的一种可选方法，不对本方案构成限制。

$$T=\frac{DS}{\frac{TWS}{JP}*QW}=\frac{DS*(R+W)*(Q1*CP+Q2*IP)}{TS*W*QW};$$

其中，$TWS=\frac{TS*W}{R+W}$；$JP=Q1*CP+Q2*IP$，其中QW为写事务产生的脏页数据权值；其中Q1为CPU利用率的权值，Q2为IO利用率的权值，其中Q1+Q2≤1。

在本发明实施例中提供了一种数据库集群的资源协调方法，在本发明实施例中，根据多个计算节点中各个计算节点分别对应的状态信息，生成主用协调节点的缓存策略，从而使计算节点可以根据这个缓存策略向主用协调节点上进行干净页数据的缓存，使得各个计算节点可以有针对性的向主用协调节点进行干净页数据的缓存，从而使本发明实施例提供的方案具备更精确的实用性，并进一步提高了计算节点的工作效率。

实施例二，如图6所示，本发明实施例提供一种应用于数据库集群的主用协调节点600。该主用协调节点600可以运行在如图1提供的数据库集群100的硬件基础上，例如该主用协调节点600可以是数据库集群100中的102。

所述主用协调节点600包括:

获取模块602，用于获取所述数据库集群中多个计算节点中各个计算节点分别对应的状态信息，所述状态信息用于表示计算节点的运行负载状态;

第一判断模块604，用于根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点是否存在预设阈值X的空闲资源;

执行模块606，用于当所述第一判断模块判断结果为是时，通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点。

本发明实施例提供的主用协调节点，根据统筹多个计算节点的运行状态，

获取模块603获取各个计算节点的状态信息用于给第一判断模块604进行判断：主用协调节点是否存在空闲资源，如存在则由所述执行模块606通知计算节点将后续产生的干净页数据上传到主用协调节点上，这样使得主用协调节点600在协调各计算节点的竞争资源之余，还能将其多余的空闲资源充分利用来帮助计算几点缓存非竞争资源的干净页数据，从而当计算节点重启或添加新的计算节点时，可以从高速的协调节点上迅速获取数据资源，充分利用主用协调节点上的存储空间和Infiniband高速通讯的特性，从而使计算节点能够快速预热提高工作效率。

如图7所示，在上述实施例二的基础上，可选的，所述主用协调节点600还可以包括：

第二判断模块605，用于当所述第一判断模块604判断结果为是时，根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据；

对应地，所述执行模块606可以用于当所述第二判断模块605判断结果为是时，通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点。

如图8所示，进一步可选的，所述主用协调节点600还可以包括：

监控模块607，用于实时监控所述主用协调节点的空闲资源，当所述主用协调节点的空闲资源小于预设阈值Z，则通过所述执行模块通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点，其中Z小于X。

当所述执行模块606通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点之后，若所述主用协调节点600收到新产生的脏页数据，则所述执行模块606还用于：优先淘汰所述主用协调节点上缓存的干净页数据。

或者，当所述执行模块606通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点之后，若所述主用协调节点600收到新产生的脏页数

据，且所述新产生的脏页数据的数据量大于所述主用协调节点600上缓存的干净页数据的数据量；则所述执行模块606还用于：优先淘汰所述主用协调节点600上缓存的干净页数据，并将所述新产生的脏页数据存储到所述数据库集群的存储设备上（例如数据库集群100中的共享存储磁盘142）。

如图9所示，在上述实施例的基础上，本发明提供另一个实施例，在这个实施例里，所述主用协调节点600还包括：缓存策略生成模块605'：用于根据所述多个计算节点中各个计算节点分别对应的状态信息，生成主用协调节点600的缓存策略；所述主用协调节点600的缓存策略包括：所述多个计算节点中各个计算节点向所述主用协调节点600上传干净页数据的策略；

对应地，所述执行模块606还用于：通知所述各个计算节点将后续产生的干净页数据根据所述主用协调节点的缓存策略上传到所述主用协调节点600。

进一步可选的，所述缓存策略生成模块605'具体可以用于：

根据所述多个计算节点中各个计算节点分别对应的状态信息，计算所述多个计算节点中每一个计算节点当前单位时间脏页数据的产生比例；

根据所述每一个计算节点当前单位时间脏页数据的产生比例，生成主用协调节点的缓存策略。

实施例三，如图1所示，在图1的数据库集群100的硬件设备基础上，

本发明实施例提供一种数据库集群，包括多个计算节点如122、124、126、128等和一个主用协调节点102：

所述主用协调节点102获取所述多个计算节点如122、124、126、128等中各个计算节点分别对应的状态信息，并根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点102是否存在预设阈值X的空闲资

源；若是，则通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点102；其中，所述状态信息用于表示计算节点的运行负载状态；

所述各个计算节点接收所述主用协调节点102发送的通知，当所述各个计算节点中任意一个计算节点产生干净页数据，则将所述产生的干净页数据上传到所述主用协调节点102。

可选的，本发明实施例还提供一种数据库集群，包括包括多个计算节点如122、124、126、128等和一个主用协调节点102:

所述主用协调节点102获取所述多个计算节点如122、124、126、128等中各个计算节点分别对应的状态信息，并根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点102是否存在预设阈值X的空闲资源；若是之后，再根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据；若是则通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点102;

所述各个计算节点接收所述主用协调节点102发送的通知，当所述各个计算节点中任意一个计算节点产生干净页数据，则上传到所述主用协调节点102。

在上述实施例的基础上，进一步可选的，本发明实施例还提供另一种数据库集群，包括包括多个计算节点如122、124、126、128等和一个主用协调节点102:

所述主用协调节点102判断所述多个计算节点如122、124、126、128中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据；若是则根据所述多个计算节点中各个计算节点分别对应的状态信息，生成主用协调节点102的缓存策略；并通知所述各个计算节点将后续产生的干净页数据根据所述主用

协调节点的缓存策略上传到所述主用协调节点102;

对应地，所述各个计算节点接收所述主用协调节点102发送的通知，当所述各个计算节点中任意一个计算节点产生干净页数据，则将所述产生的干净页数据根据所述主用协调节点的缓存策略上传到所述主用协调节点102。

值得说明的是，如图1所示，所述数据库集群还可以包括备用协调节点104，所述备用协调节点104用于为所述主用协调节点102进行数据备份，以使得所述备用协调节点104上的内容与所述主用协调节点102上的内容相同。

实施例四

图 10 是本发明第六实施例的应用于数据库系统的图 1 中的主用协调节点 102 或图 6 至图 9 中的主用协调节点 600 的结构示意图。值得说明的是，主用协调节点 102 和主用协调节点 600 是站在不同角度描述的同一个物体，其内部的结构是相同的。以下以主用协调节点 102 为例进行说明。主协调节点 102 包括收发器 720，处理器（中央处理单元 Central Processing Unit， CPU）730、存储器 740 以及总线 750。收发器 720 用于接收数据库集群系统内的各个计算节点发送的事务日志。事务日志为记录了对应事务的数据库日志。具体地，收发器 720 用于通过高速网络接收数据库系统内的计算节点发送的事务日志。高速网络指传输时延不大于纳秒（ns）级的网络。例如，高速网络为 IB。

收发器 720 用于获取所述多个计算节点中各个计算节点分别对应的状态信息，所述状态信息用于表示计算节点的运行负载状态;

处理器 730 用于根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点是否存在预设阈值 X 的空闲资源；若是则通过收发器 720 通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点。

在另一个可选的实施例中，所述处理器 730 在根据多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点是否存在预设阈值 X 的空闲资源；若是之后，还用于根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设时间内上传

小于预设阈值 X 的脏页数据;

若是，则通过收发器 720 通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点。

在又一个可选的实施例中，在通过收发器 720 通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点之后，所述处理器 730 还用于实时监控所述主用协调节点的空闲资源，当所述主用协调节点的空闲资源小于预设阈值 Z，则通过收发器 720 通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点，其中 Z 小于 X。

在所述通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点之后，若所述主用协调节点收到新产生的脏页数据，处理器730还用于：优先淘汰所述主用协调节点的存储器740上缓存的干净页数据;

或者若所述主用协调节点收到新产生的脏页数据，且所述新产生的脏页数据的数据量大于所述主用协调节点上缓存的干净页数据的数据量；处理器730还用于优先淘汰所述主用协调节点的存储器740上缓存的干净页数据，并将所述新产生的脏页数据存储到所述数据库集群的存储设备（共享存储磁盘）上。

在再一个可选的实施例中，在根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据；若是之后，所述处理器730还用于根据所述多个计算节点中各个计算节点分别对应的状态信息，生成主用协调节点的缓存策略；所述主用协调节点的缓存策略包括：所述多个计算节点中各个计算节点向所述主用协调节点上传干净页数据的策略；并通过收发器720通知所述各个计算节点将后续产生的干净页数据根据所述主用协调节点的缓存策略上传到所述主用协调节点102上。

在主用协调节点102中，所述总线750用于处理器730，收发器720以及存储器740之间的通讯。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的程序可存储于一计算机可读取存储介质中，该程序在执行时，可包括如上述各方法的实施例的流程。

其中，所述的存储介质可为磁碟、光盘、只读存储记忆体（Read-Only Memory，ROM）或随机存储记忆体（Random Access Memory，RAM）等。

以上所揭露的仅为本发明较佳实施例而已，当然不能以此来限定本发明之权利范围，因此依本发明权利要求所作的等同变化，仍属本发明所涵盖的范围。

5

权 利 要 求

1、一种数据库集群的资源协调方法，所述数据库集群包括多个计算节点和一个主用协调节点，其特征在于:

获取所述多个计算节点中各个计算节点分别对应的状态信息，所述状态信息用于表示计算节点的运行负载状态;

根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点是否存在预设阈值X的空闲资源；若是

则通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点；或者，

根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据；若是

则通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点。

2、根据权利要求1所述的方法，其特征在于，在所述通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点之后，所述方法还包括:

实时监控所述主用协调节点的空闲资源，当所述主用协调节点的空闲资源小于预设阈值Z，则通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点，其中Z小于X。

3、根据权利要求2所述的方法，其特征在于，在所述通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点之后，所述方法还包括:

若所述主用协调节点收到新产生的脏页数据;

则优先淘汰所述主用协调节点上缓存的干净页数据。

4、根据权利要求2所述的方法，其特征在于，在所述通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点之后，所述方法还包括:

若所述主用协调节点收到新产生的脏页数据，且所述新产生的脏页数据的数据量大于所述主用协调节点上缓存的干净页数据的数据量;

则优先淘汰所述主用协调节点上缓存的干净页数据，并将所述新产生的脏

页数据存储到所述数据库集群的存储设备上。

5、根据权利要求1所述的方法，其特征在于，在所述根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据；若是之后，所述方法还包括：

则根据所述多个计算节点中各个计算节点分别对应的状态信息，生成主用协调节点的缓存策略；所述主用协调节点的缓存策略包括：所述多个计算节点中各个计算节点向所述主用协调节点上传干净页数据的策略；

对应地，所述通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点包括：

通知所述各个计算节点将后续产生的干净页数据根据所述主用协调节点的缓存策略上传到所述主用协调节点。

6、根据权利要求5所述的方法，其特征在于，所述根据所述多个计算节点中各个计算节点分别对应的状态信息，生成主用协调节点的缓存策略包括：

根据所述多个计算节点中各个计算节点分别对应的状态信息，计算所述多个计算节点中每一个计算节点当前单位时间脏页数据的产生比例；

根据所述每一个计算节点当前单位时间脏页数据的产生比例，生成主用协调节点的缓存策略。

7、一种应用于数据库集群的主用协调节点，其特征在于，所述主用协调节点包括：

获取模块，用于获取所述数据库集群中多个计算节点中各个计算节点分别对应的状态信息，所述状态信息用于表示计算节点的运行负载状态；

第一判断模块，用于根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点是否存在预设阈值X的空闲资源；或者，当判断存在预设阈值X的空闲资源时，进一步根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设

时间内上传小于预设阈值X的脏页数据;

执行模块，用于当所述第一判断模块判断结果为是时，通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点。

8、根据权利要求7所述的主用协调节点，其特征在于，所述主用协调节点还包括监控模块，用于实时监控所述主用协调节点的空闲资源，当所述主用协调节点的空闲资源小于预设阈值Z，则通过所述执行模块通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点，其中Z小于X。

9、根据权利要求8所述的主用协调节点，其特征在于，

当所述执行模块通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点之后，若所述主用协调节点收到新产生的脏页数据，则

所述执行模块还用于：优先淘汰所述主用协调节点上缓存的干净页数据。

10、根据权利要求8所述的主用协调节点，其特征在于，

当所述执行模块通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点之后，若所述主用协调节点收到新产生的脏页数据，且所述新产生的脏页数据的数据量大于所述主用协调节点上缓存的干净页数据的数据量；则

所述执行模块还用于：则优先淘汰所述主用协调节点上缓存的干净页数据，并将所述新产生的脏页数据存储到所述数据库集群的存储设备上。

11、根据权利要求7所述的主用协调节点，其特征在于，当所述第二判断模块判断结果为是时，所述主用协调节点还包括：

缓存策略生成模块：用于根据所述多个计算节点中各个计算节点分别对应

的状态信息，生成主用协调节点的缓存策略；所述主用协调节点的缓存策略包括：所述多个计算节点中各个计算节点向所述主用协调节点上传干净页数据的策略；

对应地，所述执行模块用于：通知所述各个计算节点将后续产生的干净页数据根据所述主用协调节点的缓存策略上传到所述主用协调节点。

12、根据权利要求11所述的主用协调节点，其特征在于，所述缓存策略生成模块具体用于：

根据所述多个计算节点中各个计算节点分别对应的状态信息，计算所述多个计算节点中每一个计算节点当前单位时间脏页数据的产生比例；

根据所述每一个计算节点当前单位时间脏页数据的产生比例，生成主用协调节点的缓存策略。

13、一种数据库集群，包括多个计算节点和一个主用协调节点，其特征在于：

所述主用协调节点获取所述多个计算节点中各个计算节点分别对应的状态信息，并根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点是否存在预设阈值X的空闲资源；若是则通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点；或者，若是则进一步根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据；若是则通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点。其中，所述状态信息用于表示计算节点的运行负载状态；

所述各个计算节点接收所述主用协调节点发送的通知，当所述各个计算节点中任意一个计算节点产生干净页数据，则将所述产生的干净页数据上传到所述主用协调节点。

14、根据权利要求13所述的数据库集群，其特征在于，所述主用协调节点判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据；若是则根据所述多个计算节点中各个计算节点分别对应的状态信息，生成主用协调节点的缓存策略；并通知所述各个计算节点将后续产生的干净页数据根据所述主用协调节点的缓存策略上传到所述主用协调节点；

对应地，所述各个计算节点接收所述主用协调节点发送的通知，当所述各个计算节点中任意一个计算节点产生干净页数据，则将所述产生的干净页数据根据所述主用协调节点的缓存策略上传到所述主用协调节点。

15、根据权利要求 13 或 14 所述的数据库集群，其特征在于，所述数据库集群还包括备用协调节点，用于为所述主用协调节点进行数据备份，以使得所述备用协调节点上的内容与所述主用协调节点上的内容相同。

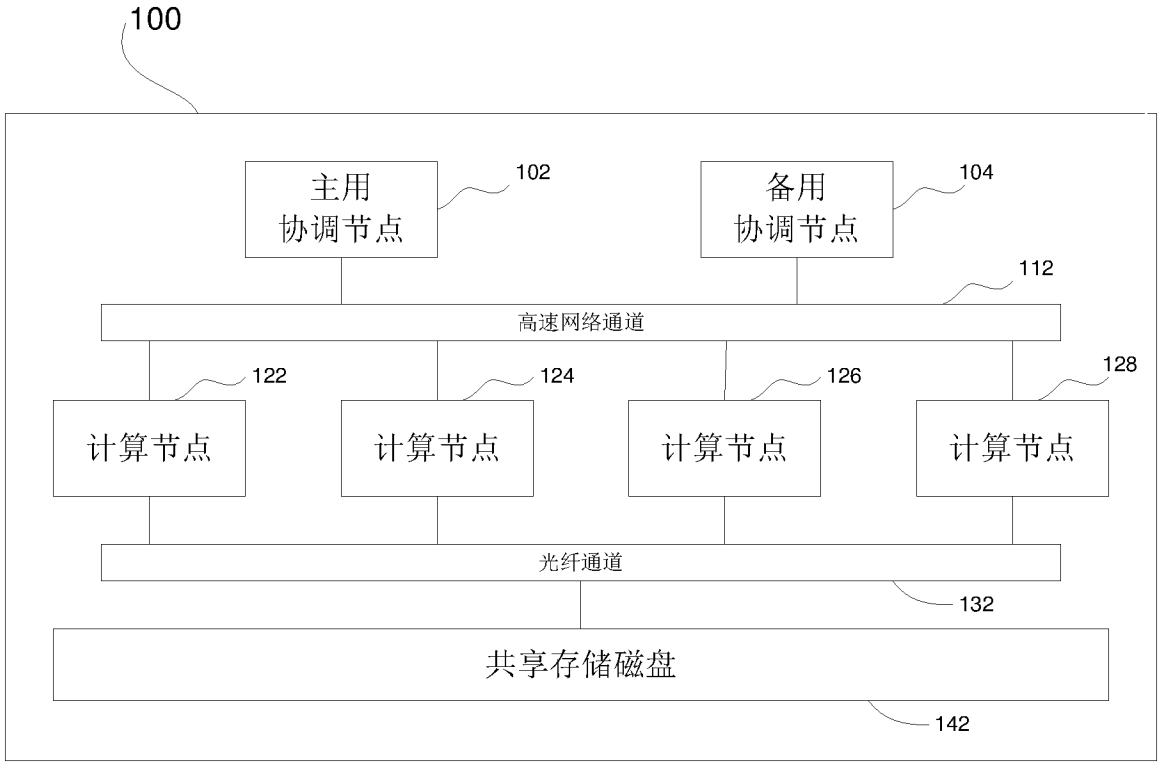

100
主用
协调节点
102
备用
协调节点
104
112
高速网络通道
122
124
126
128
计算节点
计算节点
计算节点
计算节点
光纤通道
132
共享存储磁盘
142

图 1

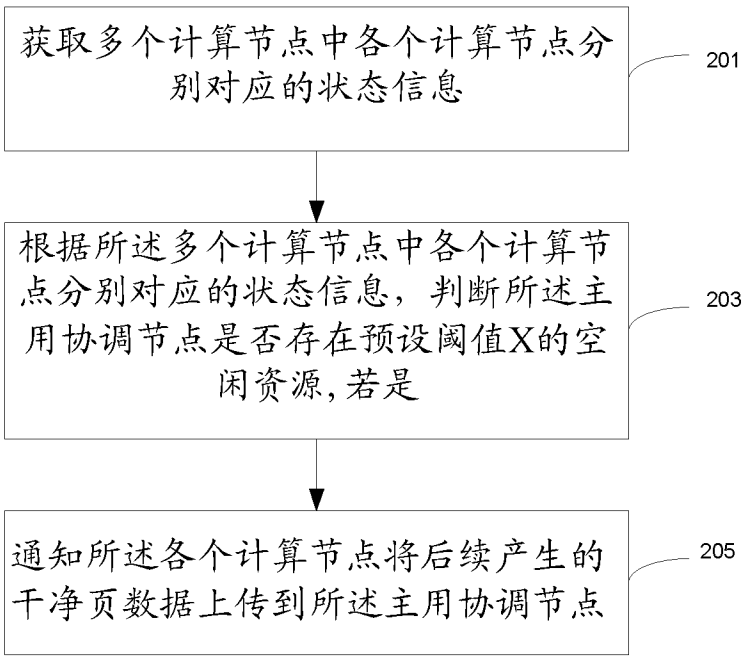

获取多个计算节点中各个计算节点分别对应的状态信息
201
根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点是否存在预设阈值X的空闲资源，若是
203
通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点
205

图 2

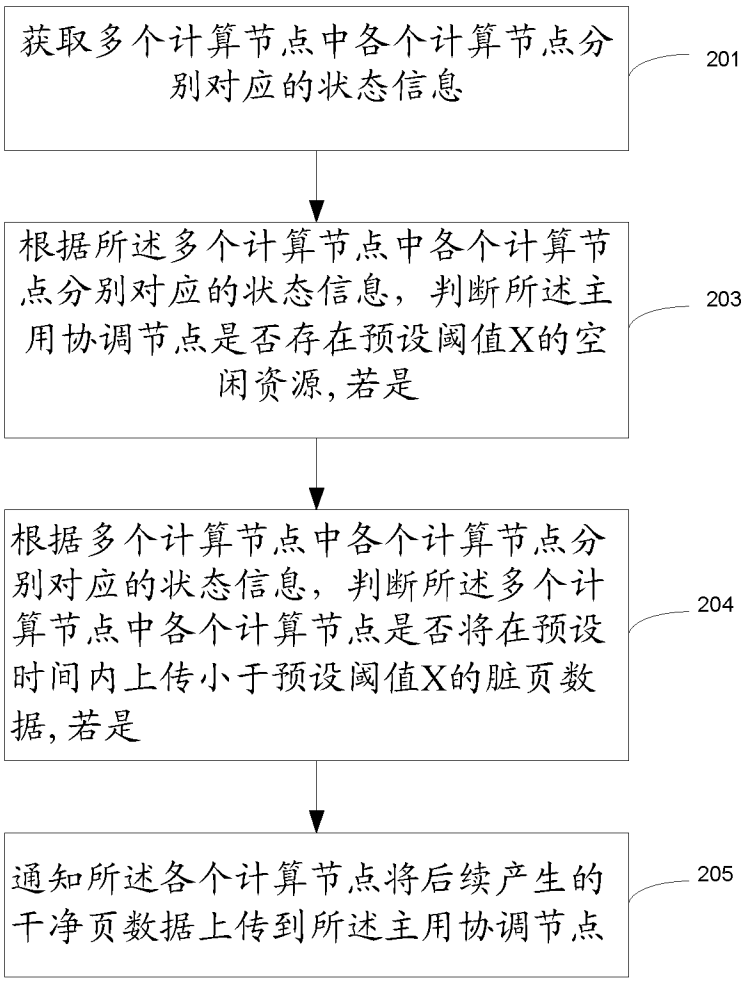

获取多个计算节点中各个计算节点分别对应的状态信息
201
根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点是否存在预设阈值X的空闲资源，若是
203
根据多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据，若是
204
通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点
205

图 3

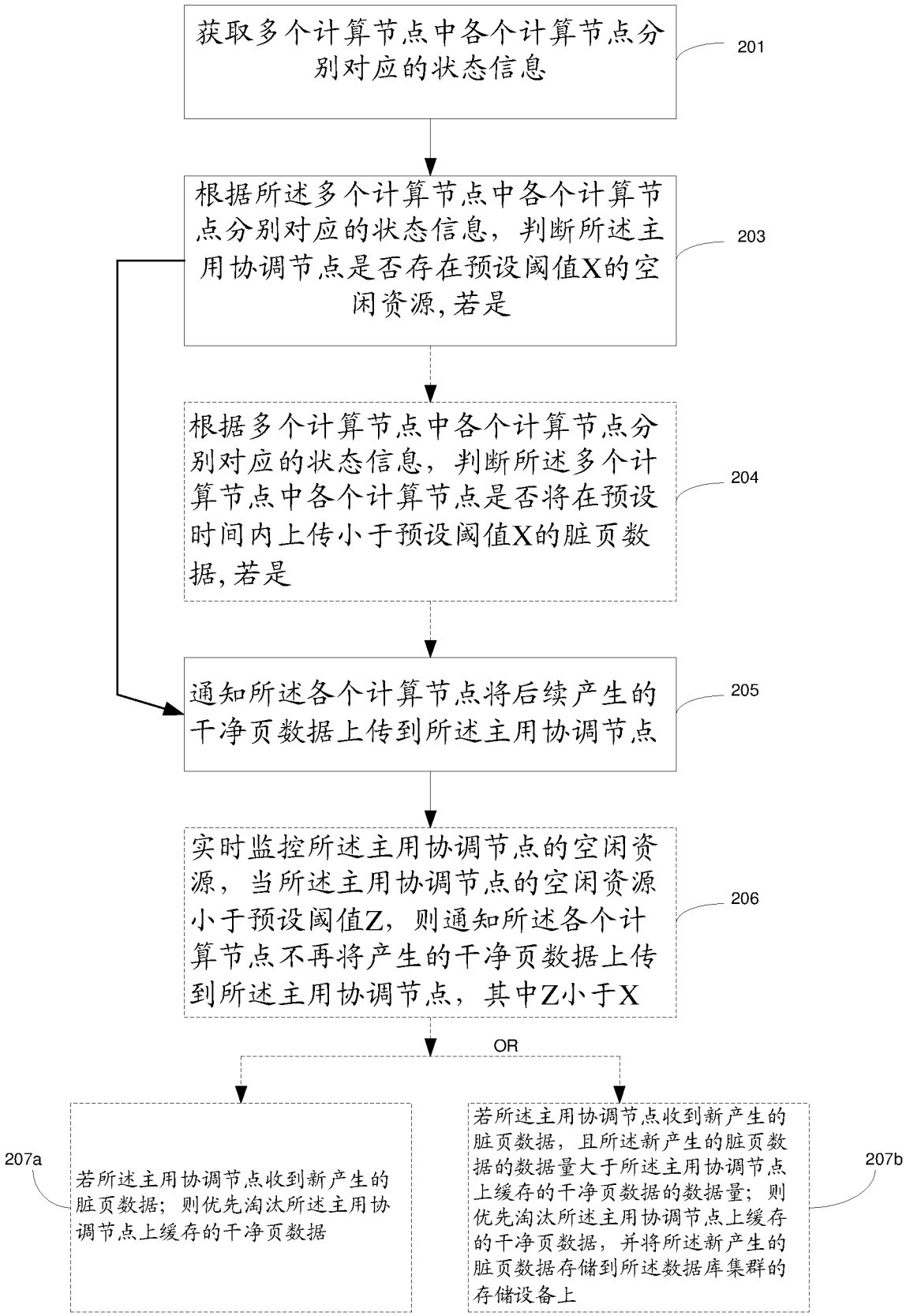

获取多个计算节点中各个计算节点分别对应的状态信息
201
根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点是否存在预设阈值X的空闲资源，若是
203
根据多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据，若是
204
通知所述各个计算节点将后续产生的干净页数据上传到所述主用协调节点
205
实时监控所述主用协调节点的空闲资源，当所述主用协调节点的空闲资源小于预设阈值Z，则通知所述各个计算节点不再将产生的干净页数据上传到所述主用协调节点，其中Z小于X
206
OR
207a
若所述主用协调节点收到新产生的脏页数据；则优先淘汰所述主用协调节点上缓存的干净页数据
若所述主用协调节点收到新产生的脏页数据，且所述新产生的脏页数据的数据量大于所述主用协调节点上缓存的干净页数据的数据量；则优先淘汰所述主用协调节点上缓存的干净页数据，并将所述新产生的脏页数据存储到所述数据库集群的存储设备上
207b

图 4

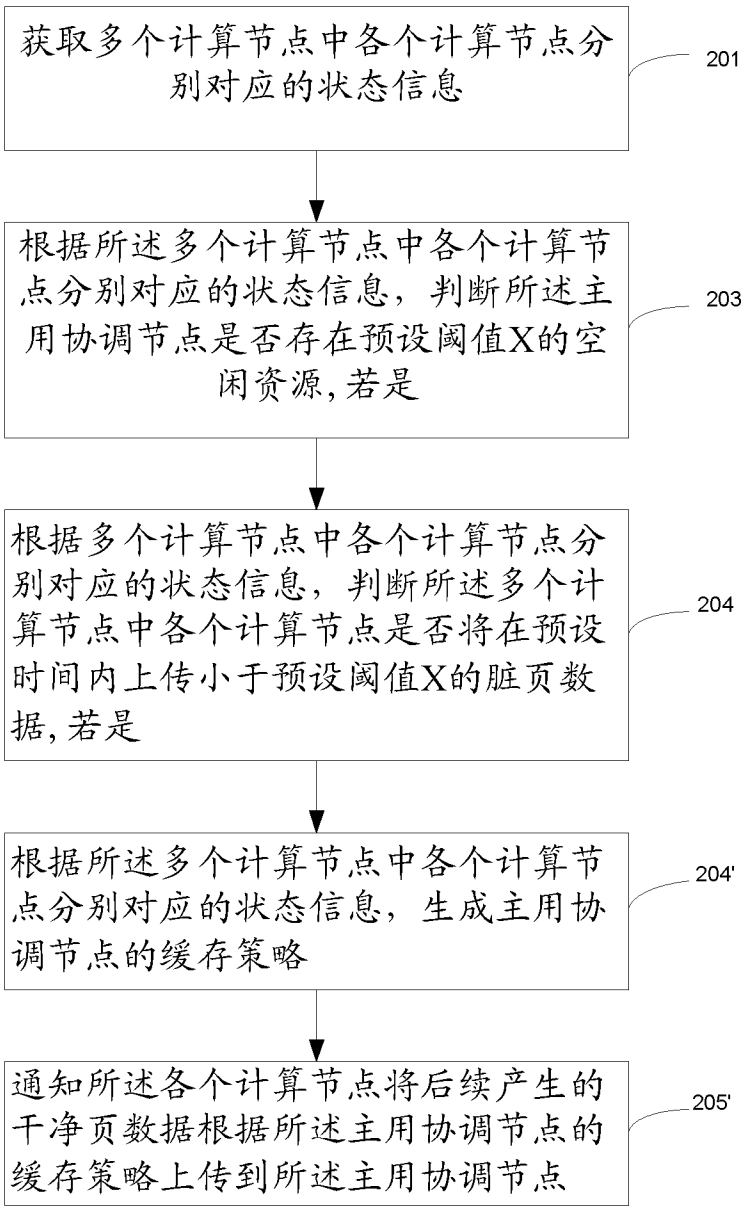

获取多个计算节点中各个计算节点分别对应的状态信息
201
根据所述多个计算节点中各个计算节点分别对应的状态信息，判断所述主用协调节点是否存在预设阈值X的空闲资源，若是
203
根据多个计算节点中各个计算节点分别对应的状态信息，判断所述多个计算节点中各个计算节点是否将在预设时间内上传小于预设阈值X的脏页数据，若是
204
根据所述多个计算节点中各个计算节点分别对应的状态信息，生成主用协调节点的缓存策略
204'
通知所述各个计算节点将后续产生的干净页数据根据所述主用协调节点的缓存策略上传到所述主用协调节点
205'

图 5

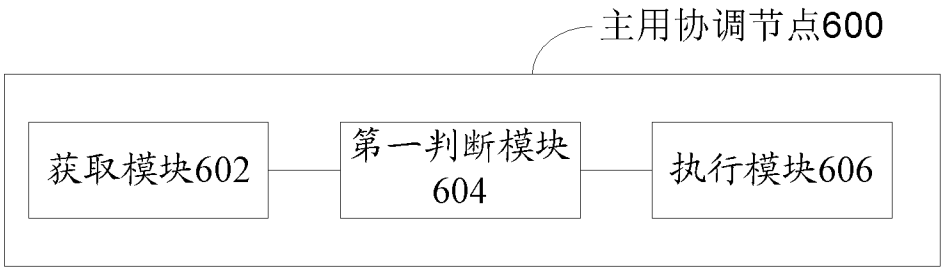

主用协调节点600
获取模块602
第一判断模块604
执行模块606

图 6

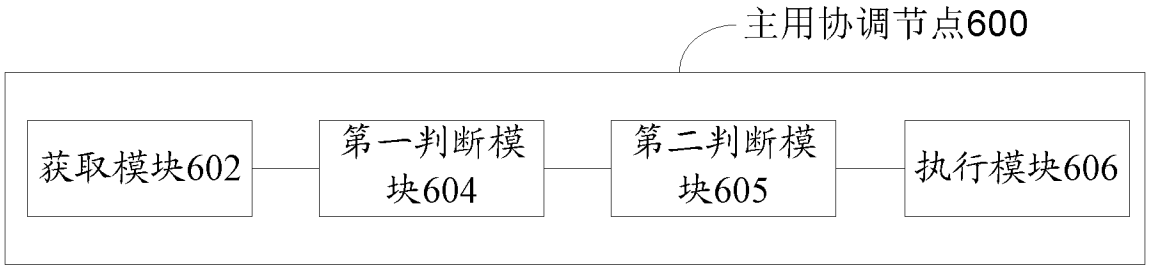

主用协调节点600
获取模块602
第一判断模块604
第二判断模块605
执行模块606

图 7

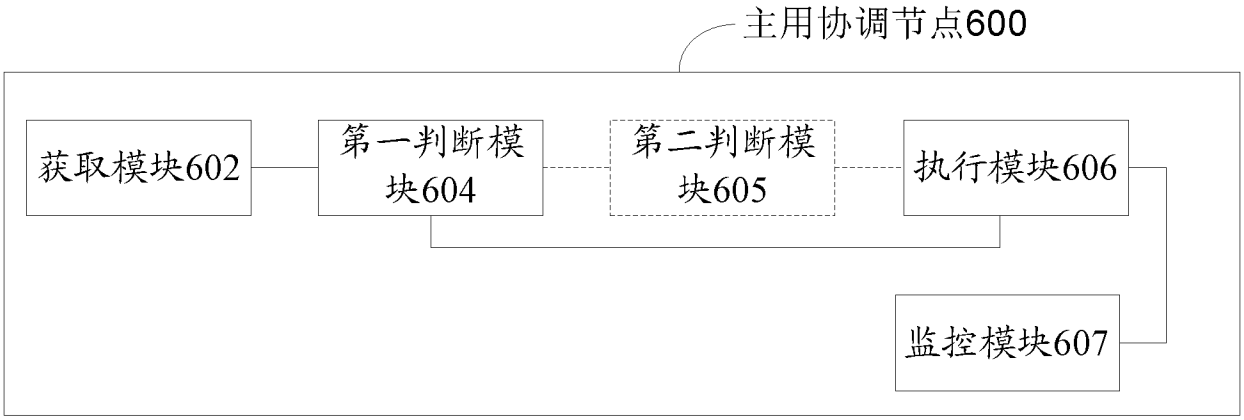

主用协调节点600
获取模块602
第一判断模块604
第二判断模块605
执行模块606
监控模块607

图 8

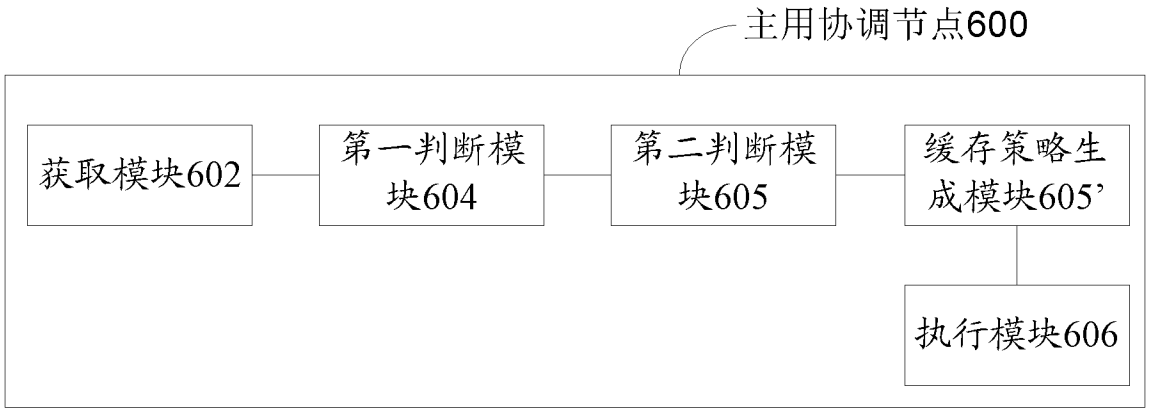

主用协调节点600
获取模块602
第一判断模块604
第二判断模块605
缓存策略生成模块605'
执行模块606

图 9

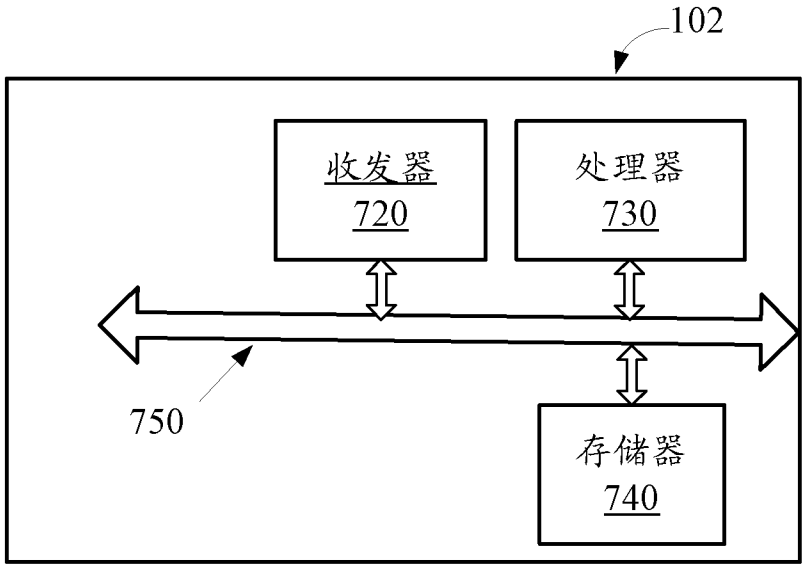


图 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/070052

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, CNPAT, WPI, EPODOC, GOOGLE: dispatching, centre, storage, free, surplus, load, cluster, node?, server?, coordinat+, schedul+, control+, manag+, shar+, resource

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 104504147 A (HUAWEI TECHNOLOGIES CO., LTD.), 08 April 2015 (08.04.2015), claims 1-15, and description, paragraphs [0070]-[0094] and [0102]-[0130]	1-15
PX	CN 104462225 A (HUAWEI TECHNOLOGIES CO., LTD.), 25 March 2015 (25.03.2015), description, paragraphs [0005]-[0013] and [0069]-[0072]	1-15
A	CN 101778002 A (INSPUR (BEIJING) ELECTRONIC INFORMATION INDUSTRY CO., LTD.), 14 July 2010 (14.07.2010), description, paragraphs [0033]-[0036] and [0039]	1-15
A	CN 104239227 A (H3C TECHNOLOGIES CO., LIMITED), 24 December 2014 (24.12.2014), the whole document	1-15
A	CN 102591679 A (COMPUTER NETWORK INFORMATION CENTER, CHINESE ACADEMY OF SCIENCES), 18 July 2012 (18.07.2012), the whole document	1-15
A	US 2013117526 A1 (SYBASE, INC.), 09 May 2013 (09.05.2013), the whole document	1-15

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 March 2016 (24.03.2016)	**01 April 2016 (01.04.2016)**
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer **WANG, Wei** Telephone No.: (86-10) **62413695**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/070052

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104504147 A	08 April 2015	None	
CN 104462225 A	25 March 2015	None	
CN 101778002 A	14 July 2010	None	
CN 104239227 A	24 December 2014	None	
CN 102591679 A	18 July 2012	None	
US 2013117526 A1	09 May 2013	None	

国际检索报告

国际申请号
PCT/CN2016/070052

A. 主题的分类

G06F 17/30(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

CNKI, CNPAT, WPI, EPODOC, GOOGLE: 集群，节点，结点，服务器，协调，调度，控制，管理，中心，共享，存储，资源，空闲，剩余，负载，负荷，cluster, node?, server?, coordinat+, schedul+, control+, manag+, shar+, resource

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
PX	CN 104504147 A (华为技术有限公司) 2015年 4月 8日 (2015-04-08) 权利要求1-15，说明书第[0070]-[0094]，[0102]-[0130]段	1-15
PX	CN 104462225 A (华为技术有限公司) 2015年 3月 25日 (2015-03-25) 说明书第[0005]-[0013]，[0069]-[0072]段	1-15
A	CN 101778002 A (浪潮北京电子信息产业有限公司) 2010年 7月 14日 (2010-07-14) 说明书第[0033]-[0036]，[0039]段	1-15
A	CN 104239227 A (杭州华三通信技术有限公司) 2014年 12月 24日 (2014-12-24) 全文	1-15
A	CN 102591679 A (中国科学院计算机网络信息中心) 2012年 7月 18日 (2012-07-18) 全文	1-15
A	US 2013117526 A1 (SYBASE, INC.) 2013年 5月 9日 (2013-05-09) 全文	1-15

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2016年 3月 24日	2016年 4月 1日
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 王伟 电话号码 (86-10)62413695

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2016/070052

检索报告引用的专利文件	公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN 104504147 A	2015年 4月 8日	无	
CN 104462225 A	2015年 3月 25日	无	
CN 101778002 A	2010年 7月 14日	无	
CN 104239227 A	2014年 12月 24日	无	
CN 102591679 A	2012年 7月 18日	无	
US 2013117526 A1	2013年 5月 9日	无	