(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 8 月 4 日 (04.08.2016)

WIPO | PCT

(10) 国际公布号
WO 2016/119618 A1

(51) 国际专利分类号:
G06F 12/02 (2006.01)

(21) 国际申请号: PCT/CN2016/071483

(22) 国际申请日: 2016 年 1 月 20 日 (20.01.2016)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201510041110.7 2015 年 1 月 27 日 (27.01.2015) 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: 李辉 (LI, Hui); 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: REMOTE MEMORY ALLOCATION METHOD, DEVICE AND SYSTEM

(54) 发明名称 : 一种远端内存分配方法、装置和系统

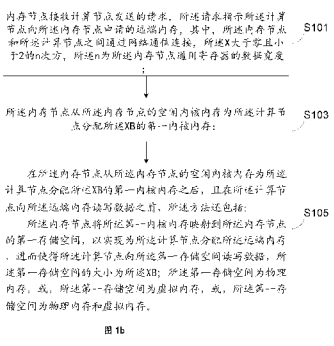

图 1b

S101 A memory node receives a request transmitted by a computing node, the request indicating the remote memory requested by the computing node from the memory node, wherein the memory node is in a communication connection to the computing node via a network. X is greater than zero and less than an *nth* power of 2, where *n* is a data width of a general register of the memory node

S103 The memory node allocates a first kernel memory of XB to the computing node from an idle kernel memory of the memory node

S104 After the memory node allocates the first kernel memory of XB to the computing node from the idle kernel memory of the memory node, and before the computing node reads and writes data to the remote memory, the method comprising: the memory node maps the first kernel memory to a first storage space of the memory node, so as to allocate the remote memory to the computing node, and thereby enables the computing node to read and write data to the first storage space, the capacity of the first storage space being the XB; the first storage space is a physical memory, a virtual memory, or a physical memory and a virtual memory

(57) Abstract: A remote memory allocation method, and a corresponding device and system, the method comprising: receiving, by a memory node (305), a request transmitted by a computing node (301); allocating, by the memory node (305), a first kernel memory of XB to the computing node (301) from an idle kernel memory of the memory node (305); after the memory node (305) allocates the first kernel memory of XB to the computing node (301) from the idle kernel memory of the memory node (305), and before the computing node (301) reads and writes data from the remote memory, the method further comprising: mapping, by the memory node (305), the first kernel memory to a first storage space of the memory node (305), so as to allocate the remote memory to the comput - ing node (301). The technical solution is free from a physical memory upper limit of the memory node (305) when allocating remote memory to the computing node (301), and allocates a larger remote memory to the computing node (301).

(57) 摘要:

[见续页]

WO 2016/119618 A1

一种远端内存分配方法，以及相应的装置和系统，该方法包括：内存节点（305）接收计算节点（301）发送的请求；所述内存节点（305）从所述内存节点（305）的空闲内核内存为所述计算节点（301）分配所述 XB 的第一内核内存；在所述内存节点（305）从所述内存节点（305）的空闲内核内存为所述计算节点（301）分配所述 XB 的第一内核内存之后，且在所述计算节点（301）向所述远端内存读写数据之前，所述方法还包括：所述内存节点（305）将所述第一内核内存映射到所述内存节点（305）的第一存储空间，以实现为所述计算节点（301）分配所述远端内存。上述技术方案能够实现对计算节点（301）分配远端内存时不受内存节点（305）中物理内存上限的限制，为计算节点（301）分配更大的远端内存。

一种远端内存分配方法、装置和系统

本申请要求于 2015 年 01 月 27 日提交中国专利局、申请号为 201510041110.7 、发明名称为“一种远端内存分配方法、装置和系统”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本发明涉及计算机技术领域，尤其涉及一种远端内存分配方法、装置和系统。

背景技术

物理内存是计算机中重要的部件之一，它是与 CPU 进行沟通的桥梁。计算机中所有程序的运行都是在内存中进行的，因此物理内存的性能对计算机的影响非常大。

当前，在第一终端设备的物理内存太小，不能满足需求的情况下，该第一终端设备通常会向其他终端设备申请远端内存，以扩展该第一终端设备的物理内存。现有技术中，其他终端设备（为便于陈述，下面以“第二终端设备”取代“其他终端设备”）会根据第一终端设备的请求，将第二终端设备的物理内存分出一部分给第一终端设备，以作为该第一终端设备的远端内存。这种方案的缺陷在于：由于第二终端设备的物理内存大小是固定的，所以分配给计算节点的远端内存不能超出该第二终端设备中空闲物理内存的上限。

发明内容

本发明提供一种远端内存分配方法，用于实现对计算节点分配远端内存时不受内存节点中物理内存上限的限制。

第一方面，本发明实施例提供一种远端内存分配方法，该方法包括：

内存节点接收计算节点发送的请求，所述请求指示所述计算节点向所述内存节点申请 XB 的远端内存，其中，所述内存节点和所述计算节点之间通过网络通信连接，所述 X 大于零且小于 2^n，所述 n 为所述内存节点 CPU 通用寄存器的数据宽度；

所述内存节点从所述内存节点的空闲内核内存为所述计算节点分配所述

XB 的第一内核内存；

在所述内存节点从所述内存节点的空闲内核内存为所述计算节点分配所述 *XB* 的第一内核内存之后，且在所述计算节点向所述远端内存读写数据之前，所述方法还包括：

所述内存节点将所述第一内核内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，所述第一存储空间的大小为所述 *XB*；所述第一存储空间为物理内存，或，所述第一存储空间为虚拟内存，或，所述第一存储空间为物理内存和虚拟内存。

结合第一方面，在第一方面的第一种实施方式下，

所述请求还指示所述远端内存为远端物理内存；或者，

所述请求还指示所述远端内存为远端虚拟内存；或者，

所述请求还指示所述远端内存为远端物理内存和远端虚拟内存。

结合第一方面的第一种实施方式，在第一方面的第二种实施方式下，

在所述请求还指示所述远端内存为远端物理内存的情况下，

所述内存节点将所述第一内核内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，所述第一存储空间的大小为所述 *XB*，具体包括：

所述内存节点将所述第一内核内存映射到所述内存节点的第一物理内存，以实现为所述计算节点分配所述远端物理内存，进而使得所述计算节点向所述第一物理内存读写数据，所述第一物理内存的大小为所述 *XB*。

结合第一方面、第一方面的第一种实施方式或第一方面的第二种实施方式，在第一方面的第三种实施方式下，

所述请求还指示第一起始地址，所述第一起始地址是指所述远端内存起始位置面向所述计算节点的地址；

所述方法还包括：

所述内存节点建立所述第一起始地址和第二起始地址之间的映射关系，所述第二起始地址是指所述空闲内核内存中所述第一内核内存起始位置的地址。

结合第一方面、第一方面的第一种实施方式或第一方面的第三种实施方式，在第一方面的第四种实施方式下，

所述内存节点从所述内存节点的空闲内核内存中为所述计算节点分配*XB*的第一内核内存，具体包括：

在所述内存节点的空闲内核内存大于或者等于所述*XB*的情况下，所述内存节点从所述空闲内核内存中为所述计算节点分配所述第一内核内存。

第二方面，本发明实施例提供一种远端内存分配方法，该方法包括：

内存节点接收计算节点发送的请求，所述请求指示所述计算节点向所述内存节点申请*XB*的远端内存，其中，所述X大于零且小于2^n，所述*n*为所述内存节点*CPU*通用寄存器的数据宽度，所述内存节点和所述计算节点之间通过网络通信连接；

所述内存节点根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述*XB*的第一逻辑内存，所述第一目标进程是指运行在所述内存节点中、具有与所述计算节点交互的第一接口且能通过所述第一接口为所述计算节点分配逻辑内存的进程，所述第一目标进程的空闲逻辑内存大于或者等于所述*XB*；

在所述内存节点从所述第一目标进程的空闲逻辑内存为所述计算节点分配所述*XB*的第一逻辑内存之后，且在所述计算节点向所述远端内存读写数据之前，所述方法还包括：

所述内存节点将所述第一逻辑内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据；其中，所述第一存储空间的大小为所述*XB*，所述第一存储空间为物理内存和虚拟内存，或，所述第一存储空间为物理内存，或，所述第一存储空间为虚拟内存。

结合第二方面，在第二方面的第一种实施方式下，

所述内存节点接收计算节点发送的请求之后，以及所述内存节点根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述*XB*的第一逻辑内存之前，所述方法还包括：

所述内存节点从运行在所述内存节点的所有进程中选择出所述第一目标进程。

结合第二方面，在第二方面的第二种实施方式下，

所述内存节点接收计算节点发送的请求之后，以及所述内存节点根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述 *XB* 的第一逻辑内存之前，所述方法还包括：

所述内存节点创建所述第一目标进程，并为所述第一目标进程分配逻辑内存。

结合第二方面、第二方面的第一种实施方式或第二方面的第二种实施方式，在第二方面的第三种实施方式下，

所述请求还指示所述远端内存为远端物理内存；或者，

所述请求还指示所述远端内存为远端虚拟内存；或者，

所述请求还指示所述远端内存为远端物理内存和远端虚拟内存。

结合第二方面的第三种实施方式，在第二方面的第四种实施方式下， 在所述请求还指示所述远端内存为远端物理内存的情况下，

所述内存节点将所述第一逻辑内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，具体包括：

所述内存节点将所述第一逻辑内存映射到所述内存节点的第一物理内存，以实现为所述计算节点分配所述远端物理内存，进而使得所述计算节点向所述第一物理内存读写数据，所述第一物理内存的大小为所述 *XB* 。

结合第二方面或第二方面的第一种实施方式至第二方面的第四种实施方式中任一种实施方式，在第二方面的第五种实施方式下，

所述请求还指示第一起始地址，所述第一起始地址是指所述远端内存起始位置面向所述计算节点的地址；

所述方法还包括：

所述内存节点建立所述第一起始地址和第二起始地址之间的映射关系，所述第二起始地址是指在所述空闲逻辑内存中所述第一逻辑内存起始位置的地址。

结合第二方面的第二种实施方式至第二方面的第五种实施方式中任一种实施方式，在第二方面的第六种实施方式下，

所述内存节点的内核数量大于创建所述第一目标进程之前运行所述内存节点上的进程的数量的情况下，

创建的所述第一目标进程之后，运行在所述内存节点上的进程的数量不超过

所述内存节点的内核数量。

第三方面，本发明实施例提供一种远端内存分配装置，该装置包括：

接收单元，用于接收计算节点发送的请求，所述请求指示所述计算节点向内存节点申请*XB*的远端内存，所述X大于零且小于2^n，所述*n*为所述内存节点*CPU*通用寄存器的数据宽度；

分配单元，用于从所述内存节点的空闲内核内存为所述计算节点分配所述*XB*的第一内核内存；

在所述分配单元从所述内存节点的空闲内核内存为所述计算节点分配所述*XB*的第一内核内存之后，且在所述计算节点向所述远端内存读写数据之前，

映射单元，用于将所述第一内核内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，所述第一存储空间的大小为所述*XB*；所述第一存储空间为物理内存，或，所述第一存储空间为虚拟内存，或，所述第一存储空间为物理内存和虚拟内存。

结合第三方面，在第三方面的第一种实施方式下，

所述请求还指示所述远端内存为远端物理内存；或者，

所述请求还指示所述远端内存为远端虚拟内存；或者，

所述请求还指示所述远端内存为远端物理内存和远端虚拟内存。

结合第三方面的第一种实施方式，在第三方面的第二种实施方式下，

在所述请求还指示所述远端内存为远端物理内存的情况下，

所述映射单元具体用于将所述第一内核内存映射到所述内存节点的第一物理内存，以实现为所述计算节点分配所述远端物理内存，进而使得所述计算节点向所述第一物理内存读写数据，所述第一物理内存的大小为所述*XB*。

结合第三方面、第三方面的第一种实施方式或第三方面的第二种实施方式，在第三方面的第三种实施方式下，所述请求还指示第一起始地址，所述第一起始地址是指所述远端内存起始位置面向所述计算节点的地址；

所述装置还包括：

建立单元用于建立所述第一起始地址和第二起始地址之间的映射关系，所述第二起始地址是指所述空闲内核内存中所述第一内核内存起始位置的地址。

第四方面，本发明实施例提供一种远端内存分配装置，该装置包括：

接收单元，用于接收计算节点发送的请求，所述请求指示所述计算节点向内存节点申请*XB*的远端内存，其中，所述X大于零且小于2^n，所述*n*为所述内存节点*CPU*通用寄存器的数据宽度，所述内存节点和所述计算节点之间通过网络通信连接；

分配单元，用于根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述*XB*的第一逻辑内存，所述第一目标进程是指运行在所述内存节点中、具有与所述计算节点交互的第一接口且能通过所述第一接口为所述计算节点分配逻辑内存的进程，所述第一目标进程的空闲逻辑内存大于或者等于所述*XB*；

在所述分配单元从所述第一目标进程的空闲逻辑内存为所述计算节点分配所述*XB*的第一逻辑内存之后，且在所述计算节点向所述远端内存读写数据之前，

映射单元，用于将所述第一逻辑内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，所述第一存储空间的大小为所述*XB*；所述第一存储空间为物理内存和虚拟内存，或，所述第一存储空间为物理内存，或，所述第一存储空间为虚拟内存。

结合第四方面，在第四方面的第一种实施方式下，

在所述接收单元接收计算节点发送的请求之后，以及所述分配单元根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述*XB*的第一逻辑内存之前，所述装置还包括：

选择单元，用于从运行在所述内存节点的所有进程中选择出所述第一目标进程。

结合第四方面，在第四方面的第二种实施方式下，

在所述接收单元接收计算节点发送的请求之后，以及所述分配单元根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述*XB*的第一逻辑内存之前，所述装置还包括：

创建单元，用于创建所述第一目标进程，并为所述第一目标进程分配逻辑

内存。

结合第四方面、第四方面的第一种实施方式或第四方面的第二种实施方式，在第四方面的第三种实施方式下，

所述请求还指示所述远端内存为远端物理内存；或者，

所述请求还指示所述远端内存为远端虚拟内存；或者，

所述请求还指示所述远端内存为远端物理内存和远端虚拟内存。

结合第四方面的第三种实施方式，在第四方面的第四种实施方式下，在所述请求还指示所述远端内存为远端物理内存的情况下，

所述映射单元具体用于将所述第一逻辑内存映射到所述内存节点的第一物理内存，以实现为所述计算节点分配所述远端物理内存，进而使得所述计算节点向所述第一物理内存读写数据，所述第一物理内存的大小为所述 *XB* 。

结合第四方面或第四方面的第一种实施方式至第四方面的第四种实施方式中任一种实施方式，在第四方面的第五种实施方式下，

所述请求还指示第一起始地址，所述第一起始地址是指所述远端内存起始位置面向所述计算节点的地址；

所述装置还包括：

建立单元，用于建立所述第一起始地址和第二起始地址之间的映射关系，所述第二起始地址是指在所述空闲逻辑内存中所述第一逻辑内存起始位置的地址。

结合第四方面的第二种实施方式至第四方面的第五种实施方式中任一种实施方式，在第四方面的第六种实施方式下，

所述内存节点的内核数量大于创建所述第一目标进程之前运行所述内存节点上的进程的数量的情况下，

所述创建单元创建的所述第一目标进程之后，运行在所述内存节点上的进程的数量不超过所述内存节点的内核数量。

第五方面，本发明实施例提供一种远端内存分配系统，该系统包括计算节点、内存节点和如第三方面或第三方面的第一种实施方式至第三方面的第三种实施方式中任一种实施方式所述的远端内存分配装置：

所述计算节点用于发送所述远端内存分配装置接收的所述请求；

所述内存节点用于提供所述第一存储空间；

所述计算节点还用于获取所述第一存储空间为所述远端内存。

结合第五方面，在第五方面的第一种实施方式下，所述远端内存分配装置集成在所述内存节点。

第六方面，本发明实施例提供一种远端内存分配系统，该系统包括计算节点、内存节点和如第四方面或第四方面的第一种实施方式至第四方面的第六种实施方式中任一种实施方式所述的远端内存分配装置：

所述计算节点用于发送所述远端内存分配装置接收的所述请求；

所述内存节点用于提供所述第一存储空间；

所述计算节点还用于获取所述第一存储空间为所述远端内存。

结合第六方面，在第六方面的第一种实施方式下，

所述远端内存分配装置集成在所述内存节点。可知，在本发明实施例提供的远端内存分配方法中，内存节点根据计算节点的请求，从内存节点的空闲内核内存中为计算节点分配第一内核内存，再将该第一内核内存映射到第一存储空间，以实现为计算节点分配远端内存。作为本领域的公知常识，在内存节点中，内核内存是操作系统为内核对象分配的内存，是逻辑内存，内核内存的大小为 $2^n B$，n 为所述内存节点 CPU 通用寄存器的数据宽度，也即内核内存要远远大于内存节点的物理内存，基于虚拟内存技术，内核内存不仅可以映射到物理内存，也可以映射到虚拟内存（即从外存上扩展的虚拟内存）。所以，采用本发明实施例提供的技术方案，计算节点向内存节点申请远端内存时，不受该内存节点中空闲物理内存上限的限制，计算节点可以向内存节点申请更大的远端内存。

附图说明

为了更清楚地说明本发明实施例的技术方案，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 1a 为本发明实施例提供的远端内存分配方法的应用场景示意图；

图 1b 为本发明实施例提供的远端内存分配方法的流程示意图；

图 1c 为本发明实施例提供的远端内存分配方法的流程示意图；

图 2a 为本发明实施例提供的远端内存分配装置的结构示意图；

图 2b 为本发明实施例提供的另一种远端内存分配装置的结构示意图；

图 2c 为本发明实施例提供的再一种远端内存分配装置的结构示意图；

图 2d(1)为本发明实施例提供的再一种远端内存分配装置的结构示意图；

图 2d(2)为本发明实施例提供的再一种远端内存分配装置的结构示意图；

图 2e 为本发明实施例提供的再一种远端内存分配装置的结构示意图；

图 3a 为本发明实施例提供的一种远端内存分配系统的结构示意图；

图 3b 为本发明实施例提供的另一种远端内存分配系统的结构示意图。

具体实施方式

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，所描述的实施例是本发明一部分实施例，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作出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

在对本发明所述的技术方案进行解释之前，先明确两个成熟的现有技术：

1、基于内存分页机制实现逻辑地址到物理地址的映射。具体是指，基于分页机制，每个进程能够获得独立的逻辑地址空间，进而通过逻辑地址访问物理地址，并且能够提供逻辑地址空间的保护。为了提升地址转换性能，一般处理器中都包括内存管理单元（*MMU*，Memeory Management Unit）和传输后备缓存器（*TLB*，Translation Lookside Buffer）来优化内存访问性能。

2、虚拟内存。是指基于虚拟内存技术实现对物理内存的动态扩展，具体的，通过使用外部存储设备（HDD/SDD/NVM 等），实现对物理内存的扩展。

如图 1a 所示，为本发明实施例所述的远端内存分配方法的应用场景示意图，具体的，包括内存节点 11 和计算节点 13，内存节点 11 和计算节点 13 之间通过高速互连接口连接，以实现内存节点 11 和计算节点 13 之间的信息交互。

实施例一

参阅附图 1b，本发明实施例所述的远端内存分配方法的流程图，本发明实施例所述的方法应用于图 1a 所示的应用场景中。具体的，包括下述步骤：

S101、内存节点接收计算节点发送的请求，所述请求指示所述计算节点向所述内存节点申请 *XB* 的远端内存，其中，所述内存节点和所述计算节点之间通过

网络通信连接，所述 X 大于零且小于 2^n，所述 n 为所述内存节点 *CPU* 通用寄存器的数据宽度；

需要说明的是，本发明实施例中所述的内存节点是指具有物理存储空间和网络通信功能的节点，在本发明实施例中称之为内存节点是为了便于表述，不应理解为限制性规定。同样的，本发明实施例中所述的计算节点是指具有运算能力和网络通信功能的节点，在本发明实施例中称之为计算节点也是为了便于表述，不应理解为限制性规定。

作为本发明的另一个实施例，所述请求还指示所述远端内存为远端物理内存；或者，所述请求还指示所述远端内存为远端虚拟内存；或者，所述请求还指示所述远端内存为远端物理内存和远端虚拟内存。

作为本发明的再一个实施例，所述请求还指示还指示第一起始地址，所述第一起始地址是指所述远端内存起始位置面向所述计算节点的地址。

S103、所述内存节点从所述内存节点的空闲内核内存为所述计算节点分配所述 *XB* 的第一内核内存；

需要说明的是，本发明实施例所述的技术方案针对的是所述内存节点的 CPU 运行在内核态的情况下。应当理解的是，所述内存节点从所述内存节点的空闲内核内存为所述计算节点分配所述 *XB* 的第一内核内存，具体包括：所述内存节点判断所述内存节点的空闲内核内存是否大于或者等于所述远端内存，并在所述空闲内核内存大于或者等于所述远端内存的情况下，所述内存节点从所述空闲内核内存中为所述计算节点分配所述 *XB* 的第一内核内存。应当理解的是，在所述空闲内核内存小于所述远端内存的情况下，则不再继续进行下述步骤，也即所述内存节点不为所述计算节点分配所述远端内存。

需要说明的是，为了更好的实现本方案，所述内存节点的操作系统在为内核设置参数时，为该内存节点的内核配备足够的内核内存。例如，对于 64 位 CPU，可以将内核内存配置为 8EB，这样可以提供充足的内核内存以分配给计算节点。

S105、在所述内存节点从所述内存节点的空闲内核内存为所述计算节点分配所述 *XB* 的第一内核内存之后，且在所述计算节点向所述远端内存读写数据之前，所述方法还包括：

所述内存节点将所述第一内核内存映射到所述内存节点的第一存储空间，以

实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，所述第一存储空间的大小为所述 XB；所述第一存储空间为物理内存，或，所述第一存储空间为虚拟内存，或，所述第一存储空间为物理内存和虚拟内存。

值得注意的是，所述内存节点从所述内存节点的空闲内核内存为所述计算节点分配所述 XB 的第一内核内存之后，所述计算节点被分配的是大小为 XB 的逻辑内存，因为在第一内核内存映射的真正的存储空间之前，它只是一段逻辑内存，由于逻辑内存不能用来读取数据，所以所述计算节点不能将数据读取到所述第一内核内存。结合步骤 S105 可知，在所述计算节点被分配所述第一内核内存之后，且在所述第一内核内存映射到所述第一存储空间之前，所述计算节点没有向所述第一内核内存读写数据。所述计算节点是在所述第一内核内存映射到所述第一存储空间之后，向所述第一存储空间读写数据的。应当理解的是，步骤 S105 所说的“所述内存节点将所述第一内核内存映射到所述内存节点的第一存储空间”是基于内存分页机制实现逻辑地址到物理地址的映射。

需要说明的是，在所述请求还指示所述远端内存为远端物理内存的情况下，所述内存节点将所述第一内核内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，所述第一存储空间的大小为所述 XB，具体包括：所述内存节点将所述第一内核内存映射到所述内存节点的第一物理内存，以实现为所述计算节点分配所述远端物理内存，进而使得所述计算节点向所述第一物理内存读写数据，所述第一物理内存的大小为所述 XB。应当理解的是，在所述内存节点将所述第一内核内存映射到所述内存节点的第一物理内存之前，本发明实施例所述的方法还包括：所述内存节点判断所述内存节点的空闲物理内存是否大于或者等于所述远端内存，并在所述空闲物理内存大于或者等于所述远端内存的情况下，所述内存节点从所述空闲物理内存中为所述计算节点分配所述第一物理内存。应当理解的是，在所述空闲物理内存小于所述远端内存的情况下，则不再继续进行下述步骤，也即所述内存节点不为所述计算节点分配所述远端内存。

进一步需要说明的是，在所述请求还指示所述远端内存为远端物理内存和远端虚拟内存，且所述远端物理内存的大小为 KB，所述远端虚拟内存的大小为

$XB-KB$的情况下，所述内存节点将所述第一内核内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，所述第一存储空间的大小为所述XB，具体包括：所述内存节点将所述第一内核内存中大小为KB的部分映射到所述内存节点的物理内存，所述内存节点将所述第一内核内存中大小为$XB-KB$的部分映射到所述内存节点的虚拟内存，进而使得所述计算节点向所述KB的物理内存和所述$XB-KB$的虚拟内存中读写数据。类似的，在所述内存节点将所述第一内核内存中大小为KB的部分映射到所述内存节点的物理内存之前，所述内存节点还判断所述内存节点的空闲物理内存是否大于或者等于所述KB，并在所述空闲物理内存大于或者等于所述KB的情况下，所述内存节点将所述第一内核内存中大小为KB的部分映射到所述内存节点的物理内存。值得注意的是，在所述请求还指示第一起始地址的情况下，本发明实施例提供的技术方案还包括：所述内存节点建立所述第一起始地址和第二起始地址之间的映射关系，所述第二起始地址是指所述空闲内核内存中所述第一内核内存起始位置的地址。 需要说明的是，所述计算节点向所述远端内存读取数据A时，会将所述数据A和所述数据A的起始位置面向所述计算节点的地址A_1发送给所述内存节点，所述内存节点根据所述映射关系，判断出所述数据A的起始位置在所述空闲内核内存中的地址A_2，在将所述A_2映射到所述内存节点的存储空间A_3，以将所述数据A读写到所述A_3中。其中所述A_3可以是物理内存，也可以是虚拟内存，或者所述A_3的一部分为物理内存，另一部分为虚拟内存。

可知，本发明实施例提供的技术方案中，内存节点根据计算节点的请求，从内存节点的空闲内核内存中为计算节点分配第一内核内存，再将该第一内核内存映射到第一存储空间，以实现为计算节点分配远端内存。作为本领域的公知常识，在内存节点中，内核内存是操作系统为内核对象分配的内存，是逻辑内存，内核内存要远远大于内存节点的物理内存，基于虚拟内存技术，内核内存不仅可以映射到物理内存，也可以映射到虚拟内存（即从外存上扩展的虚拟内存）。所以，采用本发明实施例提供的技术方案，计算节点向内存节点申请远端内存时，不受该内存节点中空闲物理内存上限的限制，计算节点可以向内存节点申请更大

的远端内存。

实施例二

参阅附图 1c，本发明实施例所述的远端内存分配方法的流程图，本发明实施例所述的方法应用于图 1a 所示的应用场景中。具体的，包括下述步骤：

S111、内存节点接收计算节点发送的请求，所述请求指示所述计算节点向所述内存节点申请 XB 的远端内存，其中，所述 X 大于零且小于 2^n，所述 n 为所述内存节点 CPU 通用寄存器的数据宽度，所述内存节点和所述计算节点之间通过网络通信连接；

需要说明的是，本发明实施例中所述的内存节点是指具有物理存储空间和网络通信功能的节点，在本发明实施例中称之为内存节点是为了便于表述，不应理解为限制性规定。同样的，本发明实施例中所述的计算节点是指具有运算能力和网络通信功能的节点，在本发明实施例中称之为计算节点也是为了便于表述，不应理解为限制性规定。

作为本发明的另一个实施例，所述请求还指示所述远端内存为远端物理内存；或者，所述请求还指示所述远端内存为远端虚拟内存；或者，所述请求还指示所述远端内存为远端物理内存和远端虚拟内存。

作为本发明的再一个实施例，所述请求还指示第一起始地址，所述第一起始地址是指所述远端内存起始位置面向所述计算节点的地址。

S113、所述内存节点根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述 XB 的第一逻辑内存，所述第一目标进程是指运行在所述内存节点中、具有与所述计算节点交互的第一接口且能通过所述第一接口为所述计算节点分配逻辑内存的进程，所述第一目标进程的空闲逻辑内存大于或者等于所述 XB；

值得注意的是，当程序被操作系统调用到内存以后，操作系统会给程序分配一定的资源（例如逻辑内存等），然后进行一系列的复杂操作，使程序变成进程以供系统调用。对于一个进程来说，它的逻辑内存和 CPU 的位数之间的关系通常为：进程的逻辑内存为 $2^r B$，其中 r 为 CPU 的位数。通常的，进程的逻辑内存远大于物理内存，且进程的逻辑内存中往往有很大一部分是空闲的。

需要说明的是，本发明实施例所述的第一目标进程是从运行在所述内存节点的多个进程中选择出来的，或者，所述第一目标进程是有所述内存节点创建的。也即，所述内存节点接收计算节点发送的请求之后，以及所述内存节点根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述 *XB* 的第一逻辑内存之前，所述方法还包括：所述内存节点从运行在所述内存节点的所有进程中选择出所述第一目标进程；或者，所述内存节点创建所述第一目标进程，并为所述第一目标进程分配逻辑内存。优选的，所述内存节点创建所述第一目标进程，是在运行在所述内存节点的所有进程中不存在所述第一目标进程的情况下。

在有多个计算节点向所述内存节点请求远端内存，且所述内存节点需要创建两个以上所述第一目标进程的情况下，需要说明的是，所述内存节点的内核数量大于创建所述第一目标进程之前运行所述内存节点上的进程的数量的情况下，所述方法还包括：创建的所述第一目标进程之后，运行在所述内存节点上的进程的数量不超过所述内存节点的内核数量。这是为了保证进程的运行速度，应当知道的是，若运行在所述内存节点上的进程的数量超过所述内存节点的内核数量，则进程的运行速度要慢一些。

S115、在所述内存节点从所述第一目标进程的空闲逻辑内存为所述计算节点分配所述 *XB* 的第一逻辑内存之后，且在所述计算节点向所述远端内存读写数据之前，所述方法还包括：

所述内存节点将所述第一逻辑内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据；其中，所述第一存储空间的大小为所述 *XB*，所述第一存储空间为物理内存和虚拟内存，或，所述第一存储空间为物理内存，或，所述第一存储空间为虚拟内存。

值得注意的是，所述内存节点从所述第一目标进程的空闲逻辑内存为所述计算节点分配所述 *XB* 的第一逻辑内存之后，所述计算节点被分配的是大小为 *XB* 的逻辑内存，因为在所述第一逻辑内存映射的真正的存储空间之前，它只是一段逻辑内存，由于逻辑内存不能用来读取数据，所以所述计算节点不能将数据读取到所述第一逻辑内存。结合步骤 S115 可知，在所述计算节点被分配所述第

一逻辑内存之后，且在所述第一逻辑内存映射到所述第一存储空间之前，所述计算节点没有向所述第一逻辑内存读写数据。所述计算节点是在所述第一逻辑内存映射到所述第一存储空间之后，向所述第一存储空间读写数据的。应当理解的是，步骤 S115 所说的“所述内存节点将所述第一逻辑内存映射到所述内存节点的第一存储空间”是基于内存分页机制实现逻辑地址到物理地址的映射。

需要说明的是，在所述请求还指示所述远端内存为远端物理内存的情况下，所述内存节点将所述第一逻辑内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，具体包括：所述内存节点将所述第一逻辑内存映射到所述内存节点的第一物理内存，以实现为所述计算节点分配所述远端物理内存，进而使得所述计算节点向所述第一物理内存读写数据，所述第一物理内存的大小为所述 XB。应当理解的是，在所述内存节点将所述第一逻辑内存映射到所述内存节点的第一物理内存之前，本发明实施例所述的方法还包括：所述内存节点判断所述内存节点的空闲物理内存是否大于或者等于所述远端内存，并在所述空闲物理内存大于或者等于所述远端内存的情况下，所述内存节点从所述空闲物理内存中为所述计算节点分配所述第一物理内存。应当理解的是，在所述空闲物理内存小于所述远端内存的情况下，则不再继续进行下述步骤，也即所述内存节点不为所述计算节点分配所述远端内存。

值得注意的是，在所述请求还指示第一起始地址的情况下，本发明实施例提供的技术方案还包括：所述内存节点建立所述第一起始地址和第二起始地址之间的映射关系，所述第二起始地址是指在所述空闲逻辑内存中所述第一逻辑内存起始位置的地址。需要说明的是，所述计算节点向所述远端内存读取数据 A 时，会将所述数据 A 和所述数据 A 的起始位置面向所述计算节点的地址 A_1 发送给所述内存节点，所述内存节点根据所述映射关系，判断出所述数据 A 的起始位置在所述空闲内核内存中的地址 A_2，在将所述 A_2 映射到所述内存节点的存储空间 A_3，以将所述数据 A 读写到所述 A_3 中。其中所述 A_3 可以是物理内存，也可以是虚拟内存，或者所述 A_3 的一部分为物理内存，另一部分为虚拟内存。

可知，本发明实施例提供的技术方案中，内存节点根据计算节点的请求，

从第一目标进程的空闲逻辑内存为所述计算节点分配第一逻辑内存，所述第一目标进程运行在所述内存节点中；再将该第一逻辑内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配远端内存。作为本领域的公知常识，进程的逻辑内存要远远大于内存节点的物理内存，基于虚拟内存技术，进程的逻辑内存不仅可以映射到物理内存，也可以映射到虚拟内存（即从外存上扩展的虚拟内存）。所以，采用本发明实施例提供的技术方案，计算节点向内存节点申请远端内存时，不受该内存节点中空闲物理内存上限的限制，计算节点可以向内存节点申请更大的远端内存。

实施例三

参见附图 2a，为本发明实施例提供的一种远端内存分配装置 200 的结构示意图，优选的，本发明实施例所述的远端内存分配装置 200 集成在图 1a 所示的内存节点中。需要说明的是，本发明实施例所述的远端内存分配装置 200 是实施例一所述的方法的执行主体，可用于执行实施例一所述的方法。具体的，远端内存分配装置 200 包括：

接收单元 201，用于接收计算节点发送的请求，所述请求指示所述计算节点向内存节点申请 *XB* 的远端内存，所述 X 大于零且小于 2^n，所述 n 为所述内存节点 *CPU* 通用寄存器的数据宽度；

作为本发明的一个实施例，所述请求还指示所述远端内存为远端物理内存；或者，所述请求还指示所述远端内存为远端虚拟内存；或者，所述请求还指示所述远端内存为远端物理内存和远端虚拟内存。

作为本发明的另一个实施例，所述请求还指示第一起始地址，所述第一起始地址是指所述远端内存起始位置面向所述计算节点的地址。

分配单元 203，用于从所述内存节点的空闲内核内存为所述计算节点分配所述 *XB* 的第一内核内存；

需要说明的是，本发明实施例所述的技术方案针对的是所述内存节点的 CPU 运行在内核态的情况下。

在所述分配单元从所述内存节点的空闲内核内存为所述计算节点分配所述 *XB* 的第一内核内存之后，且在所述计算节点向所述远端内存读写数据之前，

映射单元 205，用于将所述第一内核内存映射到所述内存节点的第一存储空

间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据；其中，所述第一存储空间的大小为所述 XB，所述第一存储空间为物理内存，或，所述第一存储空间为虚拟内存，或，所述第一存储空间为物理内存和虚拟内存。

需要说明的是，在所述请求还指示所述远端内存为远端物理内存的情况下，映射单元 205 具体用于将所述第一内核内存映射到所述内存节点的第一物理内存，以实现为所述计算节点分配所述远端物理内存，进而使得所述计算节点向所述第一物理内存读写数据，所述第一物理内存的大小为所述 XB。

在所述请求还指示第一起始地址的情况下，参见附图 2b 所述的远端内存分配装置 210，所述远端内存分配装置 210 还包括建立单元 212，其中，建立单元 212 用于建立所述第一起始地址和第二起始地址之间的映射关系，所述第二起始地址是指所述空闲内核内存中所述第一内核内存起始位置的地址。

需要说明的是，本实施例和实施例一所述的内容可互相参见，重复的内容不再赘述。

可知，在本发明实施例提供的远端内存分配装置中，内存节点根据计算节点的请求，从内存节点的空闲内核内存中为计算节点分配第一内核内存，再将该第一内核内存映射到第一存储空间，以实现为计算节点分配远端内存。作为本领域的公知常识，在内存节点中，内核内存是操作系统为内核对象分配的内存，是逻辑内存，内核内存要远远大于内存节点的物理内存，基于虚拟内存技术，内核内存不仅可以映射到物理内存，也可以映射到虚拟内存（即从外存上扩展的虚拟内存）。所以，采用本发明实施例提供的远端内存分配装置，计算节点向内存节点申请远端内存时，不受该内存节点中空闲物理内存上限的限制，计算节点可以向内存节点申请更大的远端内存。

实施例四

参见附图 2c，为本发明实施例提供的一种远端内存分配装置 220 的结构示意图，优选的，本发明实施例所述的远端内存分配装置 220 集成在图 1a 所示的内存节点中。需要说明的是，本发明实施例所述的远端内存分配装置 220 是实施例二所述的方法的执行主体，可用于执行实施例二所述的方法。具体的，远端内存分配装置 220 包括：

接收单元 221，用于接收计算节点发送的请求，所述请求指示所述计算节点

向内存节点申请*XB*的远端内存，其中，所述X大于零且小于2^n，所述*n*为所述内存节点*CPU*通用寄存器的数据宽度，所述内存节点和所述计算节点之间通过网络通信连接；

作为本发明的一个实施例，所述请求还指示所述远端内存为远端物理内存；或者，所述请求还指示所述远端内存为远端虚拟内存；或者，所述请求还指示所述远端内存为远端物理内存和远端虚拟内存。

作为本发明的另一个实施例，所述请求还指示第一起始地址，所述第一起始地址是指所述远端内存起始位置面向所述计算节点的地址。

分配单元223，用于根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述*XB*的第一逻辑内存，所述第一目标进程是指运行在所述内存节点中、具有与所述计算节点交互的第一接口且能通过所述第一接口为所述计算节点分配逻辑内存的进程，所述第一目标进程的空闲逻辑内存大于或者等于所述*XB*；

需要说明的是，本发明实施例所述的第一目标进程是从运行在所述内存节点的多个进程中选择出来的，或者，所述第一目标进程是有所述内存节点创建的。

具体的，参见图2d(1)所示的远端内存分配装置230，远端内存分配装置230还包括选择单元232，在接收单元231接收计算节点发送的请求之后，以及分配单元233根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述*XB*的第一逻辑内存之前，选择单元232用于从运行在所述内存节点的所有进程中选择出所述第一目标进程。

或者，参见图2d(2)所示的远端内存分配装置240，远端内存分配装置240还包括创建单元242，在接收单元241接收计算节点发送的请求之后，以及分配单元243根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述*XB*的第一逻辑内存之前，创建单元242用于创建所述第一目标进程，并为所述第一目标进程分配逻辑内存。

需要说明的是，在所述内存节点的内核数量大于创建所述第一目标进程之前运行所述内存节点上的进程的数量的情况下，创建单元242创建所述第一目标进程之后，运行在所述内存节点上的进程的数量不超过所述内存节点的内核数量。这是为了保证进程的运行速度，应当知道的是，若运行在所述内存节点上的进程

的数量超过所述内存节点的内核数量，则进程的运行速度要慢一些。

在所述分配单元从所述第一目标进程的空闲逻辑内存为所述计算节点分配所述 XB 的第一逻辑内存之后，且在所述计算节点向所述远端内存读写数据之前，

映射单元 225，用于将所述第一逻辑内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，所述第一存储空间的大小为所述 XB；所述第一存储空间为物理内存和虚拟内存，或，所述第一存储空间为物理内存，或，所述第一存储空间为虚拟内存。

需要说明的是，在所述请求还指示所述远端内存为远端物理内存的情况下，映射单元 225 具体用于将所述第一逻辑内存映射到所述内存节点的第一物理内存，以实现为所述计算节点分配所述远端物理内存，进而使得所述计算节点向所述第一物理内存读写数据，所述第一物理内存的大小为所述 XB。

值得注意的是，参见附图 2e 所示的远端内存分配装置 250，远端内存分配装置 250 还包括建立单元 252，在所述请求还指示第一起始地址的情况下，建立单元 252 用于建立所述第一起始地址和第二起始地址之间的映射关系，所述第二起始地址是指在所述空闲逻辑内存中所述第一逻辑内存起始位置的地址。

需要说明的是，本实施例和实施例二所述的内容可互相参见，重复的内容不再赘述。

可知，在本发明实施例提供的远端内存分配装置中，内存节点根据计算节点的请求，从第一目标进程的空闲逻辑内存为所述计算节点分配第一逻辑内存，所述第一目标进程运行在所述内存节点中；再将该第一逻辑内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配远端内存。作为本领域的公知常识，进程的逻辑内存要远远大于内存节点的物理内存，基于虚拟内存技术，进程的逻辑内存不仅可以映射到物理内存，也可以映射到虚拟内存（即从外存上扩展的虚拟内存）。所以，采用本发明实施例提供的远端内存分配装置，计算节点向内存节点申请远端内存时，不受该内存节点中空闲物理内存上限的限制，计算节点可以向内存节点申请更大的远端内存。

实施例五

参见附图 3a 所示的远端内存分配系统，包括计算节点 301、内存节点 305 和如实施例三所述的远端内存分配装置 303。

具体的，计算节点 301 用于发送远端内存分配装置 303 接收的所述请求；

内存节点 305 用于提供所述第一存储空间；

计算节点 301 还用于获取所述第一存储空间为所述远端内存。

优选的，远端内存分配装置 303 集成在内存节点 305 中。

需要说明的是，远端内存分配装置 303 的功能参见实施例三的解释说明，此处不再赘述。

可知，基于实施例三中所述的远端内存分配装置 303 的有益效果，采用本发明实施例所述的远端内存分配系统，计算节点 301 向内存节点 305 申请远端内存时，不受该内存节点 305 中空闲物理内存上限的限制，计算节点 301 可以向内存节点 305 申请更大的远端内存。

实施例六

参见附图 3b 所示的远端内存分配系统，包括计算节点 311、内存节点 315 和如实施例四所述的远端内存分配装置 313。

具体的，计算节点 311 用于发送远端内存分配装置 313 接收的所述请求；

内存节点 315 用于提供所述第一存储空间；

计算节点 311 还用于获取所述第一存储空间为所述远端内存。

优选的，远端内存分配装置 313 集成在内存节点 315 中。

需要说明的是，远端内存分配装置 313 的功能参见实施例四的解释说明，此处不再赘述。

可知，基于实施例四中所述的远端内存分配装置 313 的效果，采用本发明实施例所述的远端内存分配系统，计算节点 311 向内存节点 315 申请远端内存时，不受该内存节点 315 中空闲物理内存上限的限制，计算节点 311 可以向内存节点 315 申请更大的远端内存。

其中上述实施例之间可以相互参见。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

另外，在本发明各个实施例提供的网络设备中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或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中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、磁碟或者光盘等各种可以存储程序代码的介质。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本发明的范围。

本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于系统实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

以上所述的本发明实施方式，并不构成对本发明保护范围的限定。任何在本发明的精神和原则之内所作的修改、等同替换和改进等，均应包含在本发明的保护范围之内。

权利要求

1、一种远端内存分配方法，其特征在于，包括：

内存节点接收计算节点发送的请求，所述请求指示所述计算节点向所述内存节点申请 *XB* 的远端内存，其中，所述内存节点和所述计算节点之间通过网络通信连接，所述 X 大于零且小于 2^n，所述 n 为所述内存节点 *CPU* 通用寄存器的数据宽度；

所述内存节点从所述内存节点的空闲内核内存为所述计算节点分配所述 *XB* 的第一内核内存；

在所述内存节点从所述内存节点的空闲内核内存为所述计算节点分配所述 *XB* 的第一内核内存之后，且在所述计算节点向所述远端内存读写数据之前，所述方法还包括：

所述内存节点将所述第一内核内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，所述第一存储空间的大小为所述 *XB*；所述第一存储空间为物理内存，或，所述第一存储空间为虚拟内存，或，所述第一存储空间为物理内存和虚拟内存。

2、根据权利要求 1 所述的方法，其特征在于：

所述请求还指示所述远端内存为远端物理内存；或者，

所述请求还指示所述远端内存为远端虚拟内存；或者，

所述请求还指示所述远端内存为远端物理内存和远端虚拟内存。

3、根据权利要求 2 所述的方法，其特征在于：

在所述请求还指示所述远端内存为远端物理内存的情况下，

所述内存节点将所述第一内核内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，所述第一存储空间的大小为所述 *XB*，具体包括：

所述内存节点将所述第一内核内存映射到所述内存节点的第一物理内存，以实现为所述计算节点分配所述远端物理内存，进而使得所述计算节点向所述第一物理内存读写数据，所述第一物理内存的大小为所述 *XB*。

4、根据权利要求 1 至 3 任一项所述的方法，其特征在于，所述请求还指示第一

起始地址，所述第一起始地址是指所述远端内存起始位置面向所述计算节点的地址；

所述方法还包括：

所述内存节点建立所述第一起始地址和第二起始地址之间的映射关系，所述第二起始地址是指所述空闲内核内存中所述第一内核内存起始位置的地址。

5、一种远端内存分配方法，其特征在于，包括：

内存节点接收计算节点发送的请求，所述请求指示所述计算节点向所述内存节点申请 *XB* 的远端内存，其中，所述 X 大于零且小于 2^n，所述 n 为所述内存节点 *CPU* 通用寄存器的数据宽度，所述内存节点和所述计算节点之间通过网络通信连接；

所述内存节点根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述 *XB* 的第一逻辑内存，所述第一目标进程是指运行在所述内存节点中、具有与所述计算节点交互的第一接口且能通过所述第一接口为所述计算节点分配逻辑内存的进程，所述第一目标进程的空闲逻辑内存大于或者等于所述 *XB*；

在所述内存节点从所述第一目标进程的空闲逻辑内存为所述计算节点分配所述 *XB* 的第一逻辑内存之后，且在所述计算节点向所述远端内存读写数据之前，所述方法还包括：

所述内存节点将所述第一逻辑内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据；其中，所述第一存储空间的大小为所述 *XB*，所述第一存储空间为物理内存和虚拟内存，或，所述第一存储空间为物理内存，或，所述第一存储空间为虚拟内存。

6、根据权利要求 5 所述的方法，其特征在于，

所述内存节点接收计算节点发送的请求之后，以及所述内存节点根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述 *XB* 的第一逻辑内存之前，所述方法还包括：

所述内存节点从运行在所述内存节点的所有进程中选择出所述第一目标进程。

7、根据权利要求 5 所述的方法，其特征在于，

所述内存节点接收计算节点发送的请求之后，以及所述内存节点根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述 *XB* 的第一逻辑内存之前，所述方法还包括：

所述内存节点创建所述第一目标进程，并为所述第一目标进程分配逻辑内存。

8、根据权利要求 5 至 7 任一项所述的方法，其特征在于：

所述请求还指示所述远端内存为远端物理内存；或者，

所述请求还指示所述远端内存为远端虚拟内存；或者，

所述请求还指示所述远端内存为远端物理内存和远端虚拟内存。

9、根据权利要求 8 所述的方法，其特征在于：

在所述请求还指示所述远端内存为远端物理内存的情况下，

所述内存节点将所述第一逻辑内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，具体包括：

所述内存节点将所述第一逻辑内存映射到所述内存节点的第一物理内存，以实现为所述计算节点分配所述远端物理内存，进而使得所述计算节点向所述第一物理内存读写数据，所述第一物理内存的大小为所述 *XB* 。

10、根据权利要求 5 至 9 任一项所述的方法，其特征在于，

所述请求还指示第一起始地址，所述第一起始地址是指所述远端内存起始位置面向所述计算节点的地址；

所述方法还包括：

所述内存节点建立所述第一起始地址和第二起始地址之间的映射关系，所述第二起始地址是指在所述空闲逻辑内存中所述第一逻辑内存起始位置的地址。

11、根据权利要求 7 至 10 任一项所述的方法，其特征在于：

所述内存节点的内核数量大于创建所述第一目标进程之前运行所述内存节点上的进程的数量的情况下，

创建的所述第一目标进程之后，运行在所述内存节点上的进程的数量不超过所述内存节点的内核数量。

12、一种远端内存分配装置，其特征在于，包括：

接收单元，用于接收计算节点发送的请求，所述请求指示所述计算节点向内存节点申请*XB*的远端内存，所述X大于零且小于2^n，所述*n*为所述内存节点*CPU*通用寄存器的数据宽度；

分配单元，用于从所述内存节点的空闲内核内存为所述计算节点分配所述*XB*的第一内核内存；

在所述分配单元从所述内存节点的空闲内核内存为所述计算节点分配所述*XB*的第一内核内存之后，且在所述计算节点向所述远端内存读写数据之前，

映射单元，用于将所述第一内核内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，所述第一存储空间的大小为所述*XB*；所述第一存储空间为物理内存，或，所述第一存储空间为虚拟内存，或，所述第一存储空间为物理内存和虚拟内存。

13、根据权利要求12所述的装置，其特征在于：

所述请求还指示所述远端内存为远端物理内存；或者，

所述请求还指示所述远端内存为远端虚拟内存；或者，

所述请求还指示所述远端内存为远端物理内存和远端虚拟内存。

14、根据权利要求13所述的装置，其特征在于：

在所述请求还指示所述远端内存为远端物理内存的情况下，

所述映射单元具体用于将所述第一内核内存映射到所述内存节点的第一物理内存，以实现为所述计算节点分配所述远端物理内存，进而使得所述计算节点向所述第一物理内存读写数据，所述第一物理内存的大小为所述*XB*。

15、根据权利要求12至14任一项所述的装置，其特征在于，所述请求还指示第一起始地址，所述第一起始地址是指所述远端内存起始位置面向所述计算节点的地址；

所述装置还包括：

建立单元用于建立所述第一起始地址和第二起始地址之间的映射关系，所述第二起始地址是指所述空闲内核内存中所述第一内核内存起始位置的地址。

16、一种远端内存分配装置，其特征在于，包括：

接收单元，用于接收计算节点发送的请求，所述请求指示所述计算节点向

内存节点申请*XB*的远端内存，其中，所述X大于零且小于2^n，所述*n*为所述内存节点*CPU*通用寄存器的数据宽度，所述内存节点和所述计算节点之间通过网络通信连接；

分配单元，用于根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述*XB*的第一逻辑内存，所述第一目标进程是指运行在所述内存节点中、具有与所述计算节点交互的第一接口且能通过所述第一接口为所述计算节点分配逻辑内存的进程，所述第一目标进程的空闲逻辑内存大于或者等于所述*XB*；

在所述分配单元从所述第一目标进程的空闲逻辑内存为所述计算节点分配所述*XB*的第一逻辑内存之后，且在所述计算节点向所述远端内存读写数据之前，

映射单元，用于将所述第一逻辑内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，所述第一存储空间的大小为所述*XB*；所述第一存储空间为物理内存和虚拟内存，或，所述第一存储空间为物理内存，或，所述第一存储空间为虚拟内存。

17、根据权利要求16所述的装置，其特征在于，

在所述接收单元接收计算节点发送的请求之后，以及所述分配单元根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述*XB*的第一逻辑内存之前，所述装置还包括：

选择单元，用于从运行在所述内存节点的所有进程中选择出所述第一目标进程。

18、根据权利要求16所述的装置，其特征在于，

在所述接收单元接收计算节点发送的请求之后，以及所述分配单元根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述*XB*的第一逻辑内存之前，所述装置还包括：

创建单元，用于创建所述第一目标进程，并为所述第一目标进程分配逻辑内存。

19、根据权利要求16至18任一项所述的装置，其特征在于：

所述请求还指示所述远端内存为远端物理内存；或者，

所述请求还指示所述远端内存为远端虚拟内存；或者，

所述请求还指示所述远端内存为远端物理内存和远端虚拟内存。

20、根据权利要求 19 所述的装置，其特征在于：

在所述请求还指示所述远端内存为远端物理内存的情况下，

所述映射单元具体用于将所述第一逻辑内存映射到所述内存节点的第一物理内存，以实现为所述计算节点分配所述远端物理内存，进而使得所述计算节点向所述第一物理内存读写数据，所述第一物理内存的大小为所述 XB 。

21、根据权利要求 16 至 20 任一项所述的装置，其特征在于，

所述请求还指示第一起始地址，所述第一起始地址是指所述远端内存起始位置面向所述计算节点的地址；

所述装置还包括：

建立单元，用于建立所述第一起始地址和第二起始地址之间的映射关系，所述第二起始地址是指在所述空闲逻辑内存中所述第一逻辑内存起始位置的地址。

22、根据权利要求 18 至 21 任一项所述的装置，其特征在于：

所述内存节点的内核数量大于创建所述第一目标进程之前运行所述内存节点上的进程的数量的情况下，

所述创建单元创建的所述第一目标进程之后，运行在所述内存节点上的进程的数量不超过所述内存节点的内核数量。

23、一种远端内存分配系统，其特征在于，包括计算节点、内存节点和如权利要求 12 至 15 任一项所述的远端内存分配装置：

所述计算节点用于发送所述远端内存分配装置接收的所述请求；

所述内存节点用于提供所述第一存储空间；

所述计算节点还用于获取所述第一存储空间为所述远端内存。

24、根据权利要求 23 所述的系统，其特征在于：

所述远端内存分配装置集成在所述内存节点。

25、一种远端内存分配系统，其特征在于，包括计算节点、内存节点和如权利要求 16 至 22 任一项所述的远端内存分配装置：

所述计算节点用于发送所述远端内存分配装置接收的所述请求；

所述内存节点用于提供所述第一存储空间；

所述计算节点还用于获取所述第一存储空间为所述远端内存。

26、根据权利要求 25 所述的系统，其特征在于：

所述远端内存分配装置集成在所述内存节点。


计算节点
内存节点

图 1a

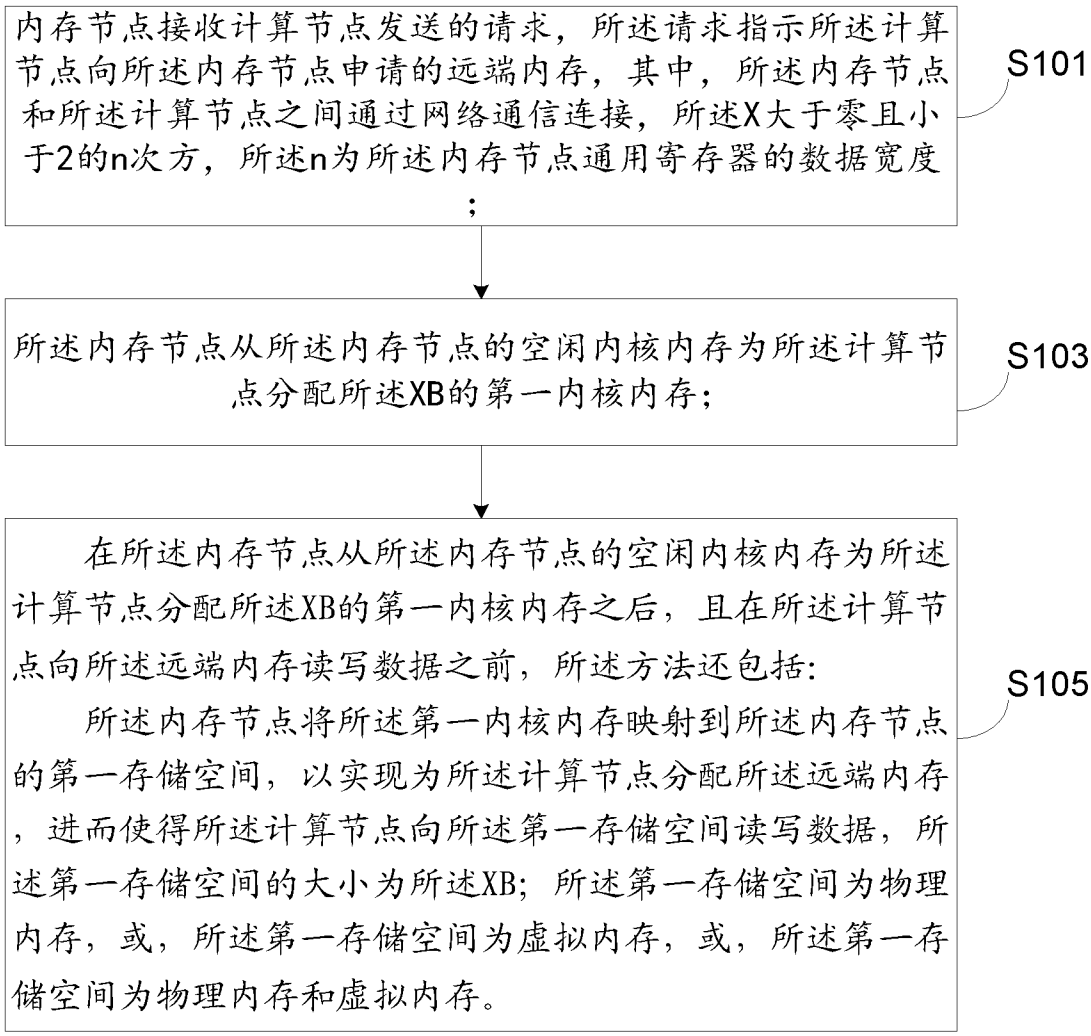
内存节点接收计算节点发送的请求，所述请求指示所述计算节点向所述内存节点申请的远端内存，其中，所述内存节点和所述计算节点之间通过网络通信连接，所述X大于零且小于2的n次方，所述n为所述内存节点通用寄存器的数据宽度；
S101
所述内存节点从所述内存节点的空闲内核内存为所述计算节点分配所述XB的第一内核内存；
S103
在所述内存节点从所述内存节点的空闲内核内存为所述计算节点分配所述XB的第一内核内存之后，且在所述计算节点向所述远端内存读写数据之前，所述方法还包括：
所述内存节点将所述第一内核内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据，所述第一存储空间的大小为所述XB；所述第一存储空间为物理内存，或，所述第一存储空间为虚拟内存，或，所述第一存储空间为物理内存和虚拟内存。
S105

图 1b

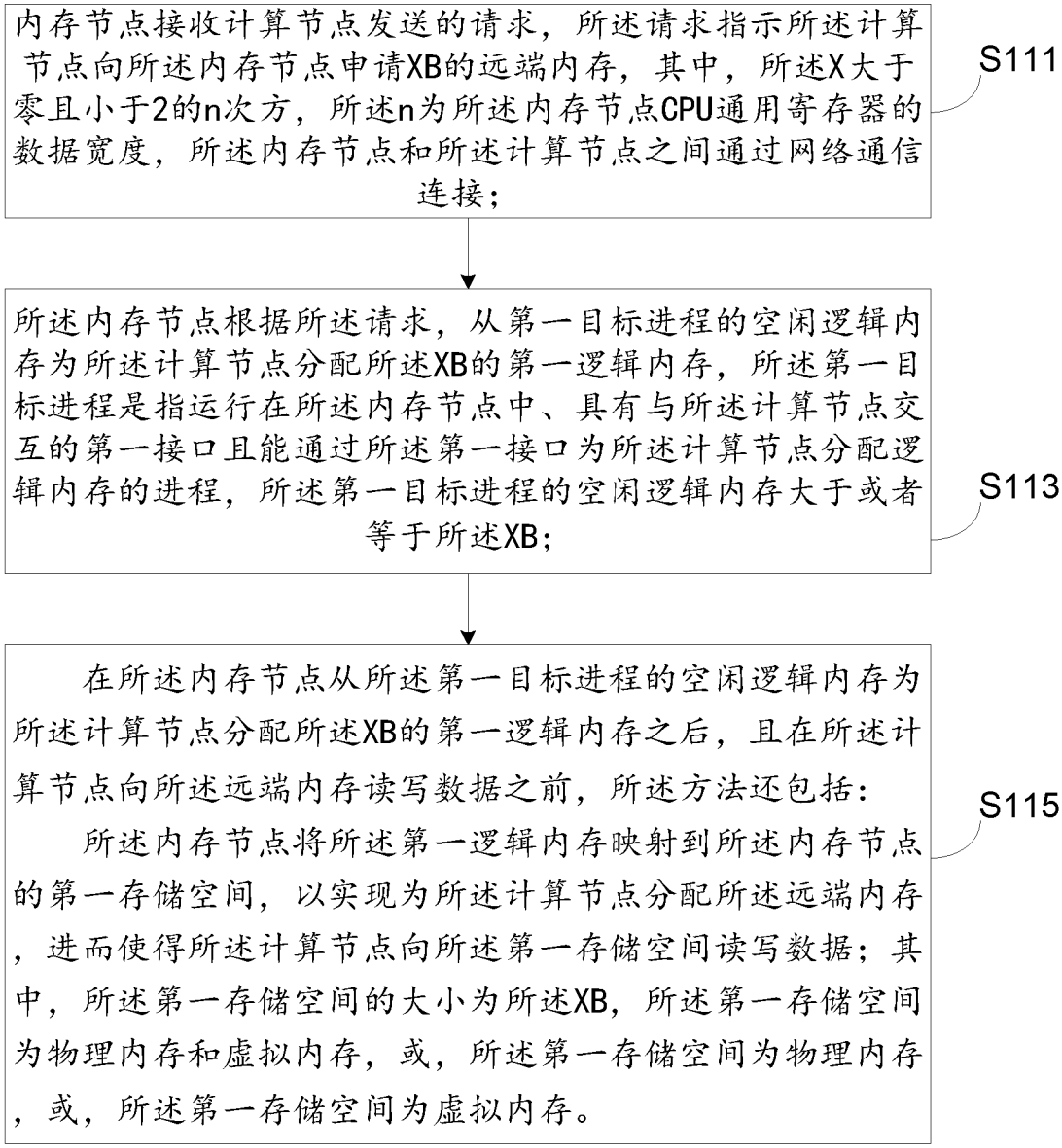
内存节点接收计算节点发送的请求，所述请求指示所述计算节点向所述内存节点申请XB的远端内存，其中，所述X大于零且小于2的n次方，所述n为所述内存节点CPU通用寄存器的数据宽度，所述内存节点和所述计算节点之间通过网络通信连接；
S111
所述内存节点根据所述请求，从第一目标进程的空闲逻辑内存为所述计算节点分配所述XB的第一逻辑内存，所述第一目标进程是指运行在所述内存节点中、具有与所述计算节点交互的第一接口且能通过所述第一接口为所述计算节点分配逻辑内存的进程，所述第一目标进程的空闲逻辑内存大于或者等于所述XB；
S113
在所述内存节点从所述第一目标进程的空闲逻辑内存为所述计算节点分配所述XB的第一逻辑内存之后，且在所述计算节点向所述远端内存读写数据之前，所述方法还包括：
所述内存节点将所述第一逻辑内存映射到所述内存节点的第一存储空间，以实现为所述计算节点分配所述远端内存，进而使得所述计算节点向所述第一存储空间读写数据；其中，所述第一存储空间的大小为所述XB，所述第一存储空间为物理内存和虚拟内存，或，所述第一存储空间为物理内存，或，所述第一存储空间为虚拟内存。
S115

图 1c

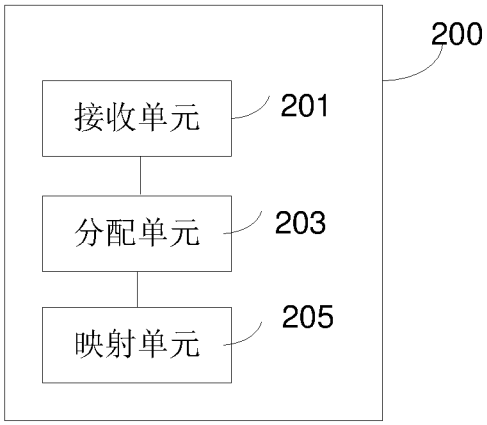
200
接收单元
201
分配单元
203
映射单元
205

图 2a

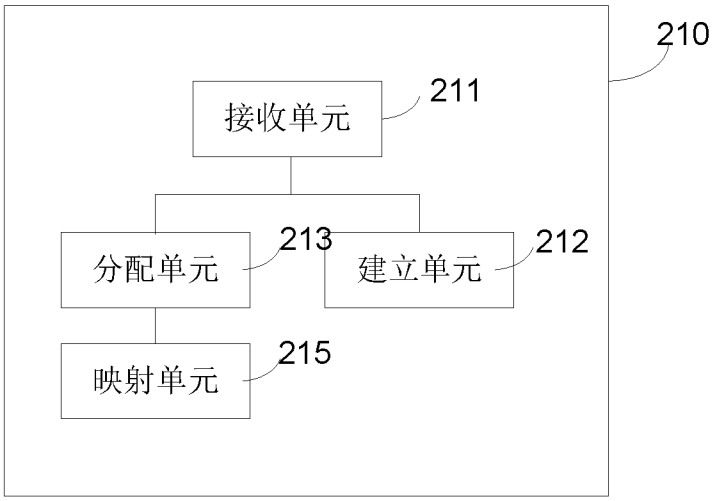
210
接收单元
211
分配单元
213
建立单元
212
映射单元
215

图 2b

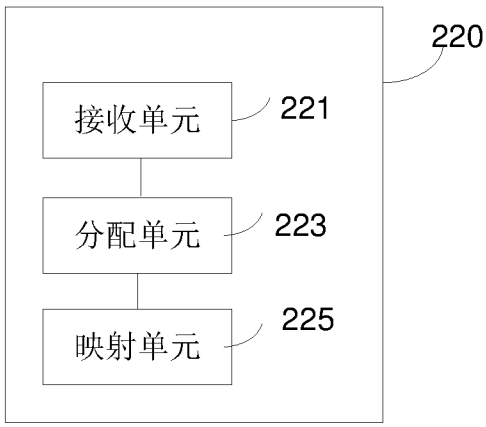
220
接收单元
221
分配单元
223
映射单元
225

图 2c


230
接收单元
231
选择单元
232
分配单元
233
映射单元
235

图 2d(1)

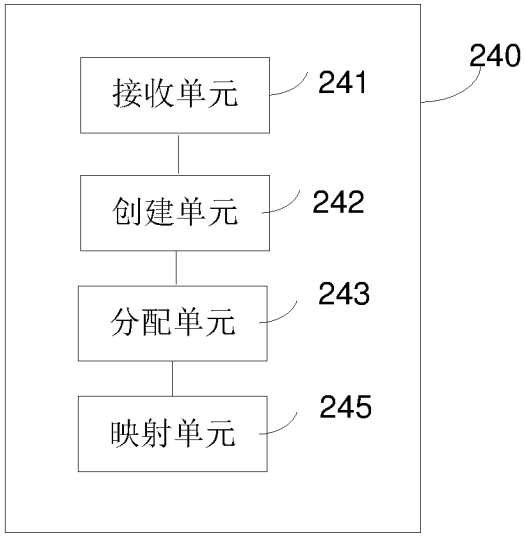

图 2d(2)

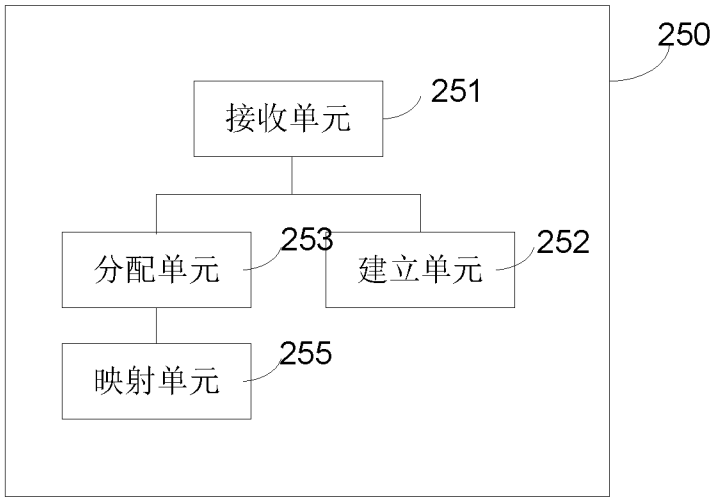

图 2e

图 3a

图 3b

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/071483

A. CLASSIFICATION OF SUBJECT MATTER

G06F 12/02 (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)

CNPAT; CNKI; WPI; EPODOC: linear 1w space, map+, applicat+, allocat+, logic 1w address+, logic 1w memory, logic 1w space, request+, remote, memory, storage, virtual 1w address, linear 1w address, virtual 1w memory, shar+, virtual 1w space

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 104216835 A (YANG, Liqun), 17 December 2014 (17.12.2014), description, paragraphs[0020]-[0026], and figures 1-3	1-26
A	CN 103942087 A (HUAWEI TECHNOLOGIES CO., LTD.), 23 July 2014 (23.07.2014), the whole document	1-26
A	CN 102110196 A (CHINA GREAT WALL COMPUTER SHENZHEN CO., LTD.), 29 June 2011 (29.06.2011), the whole document	1-26
A	CN 103077120 A (NEUSOFT CORPORATION), 01 May 2013 (01.05.2013), the whole document	1-26

☐ 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
28 March 2016 (28.03.2016)	**22 April 2016 (22.04.2016)**
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 **LIU, Guangde** Telephone No.: (86-10) **82245503**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/071483

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104216835 A	17 December 2014	None	
CN 103942087 A	23 July 2014	None	
CN 102110196 A	29 June 2011	None	
CN 103077120 A	01 May 2013	None	

国际检索报告

国际申请号
PCT/CN2016/071483

A. 主题的分类

G06F 12/02(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））

CNPAT; CNKI; WPI; EPODOC:线性空间，映射，申请，分配，逻辑地址，逻辑内存，逻辑空间，请求，远程，远端，内存，存储，虚拟地址，线性地址，虚拟内存，共享，虚拟空间, linear lw space, map+, applicat+, allocat+, logic lw address+, logic lw memory, logic lw space, request+, remote, memory, storage, virtual lw address, linear lw address, virtual lw memory, shar+, virtual lw space

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
X	CN 104216835 A (杨立群) 2014年 12月 17日 (2014-12-17) 说明书第[0020]-[0026]、附图1-3	1-26
A	CN 103942087 A (华为技术有限公司) 2014年 7月 23日 (2014-07-23) 全文	1-26
A	CN 102110196 A (中国长城计算机深圳股份有限公司) 2011年 6月 29日 (2011-06-29) 全文	1-26
A	CN 103077120 A (东软集团股份有限公司) 2013年 5月 1日 (2013-05-01) 全文	1-26

☐ 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 28日	2016年 4月 22日

ISA/CN的名称和邮寄地址	受权官员
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	刘光德
传真号 (86-10)62019451	电话号码 (86-10)82245503

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2016/071483

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 104216835 A	2014年 12月 17日	无	
CN 103942087 A	2014年 7月 23日	无	
CN 102110196 A	2011年 6月 29日	无	
CN 103077120 A	2013年 5月 1日	无	