



(12)发明专利申请

(10)申请公布号 CN 110209602 A  
(43)申请公布日 2019.09.06

(21)申请号 201910412026.X

(22)申请日 2019.05.17

(71)申请人 北京航空航天大学  
地址 100191 北京市海淀区学院路37号

(72)发明人 肖利民 宋尧 秦广军 霍志胜  
韦冰 周汉杰

(74)专利代理机构 北京海虹嘉诚知识产权代理  
有限公司 11129  
代理人 吴小灿

(51) Int. Cl.  
G06F 12/10(2016.01)  
G06F 16/22(2019.01)  
G06F 12/0802(2016.01)

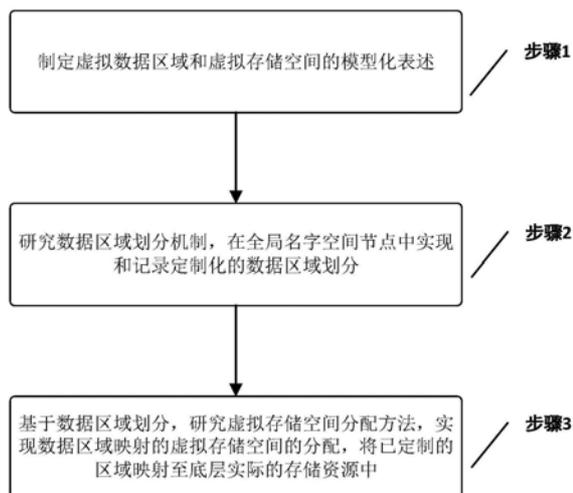
权利要求书2页 说明书6页 附图5页

(54)发明名称

跨域虚拟数据空间中区域划分及空间分配方法

(57)摘要

本发明提出一种跨域虚拟数据空间中区域划分及空间分配方法,基于由分散自治存储资源聚合成的全局名字空间,通过跨域的统一区域划分方法实现全局数据的统一管理和访问,并基于虚拟视图实现优化全局数据放置的空间分配。首先,制定虚拟数据区域和虚拟存储空间的模型化表述,为区域划分及空间分配提供理论基础;然后研究数据区域划分机制,在全局名字空间节点中实现和记录定制化的数据区域划分,以支持全局数据的统一管理和访问;最后,研究虚拟存储空间分配方法,实现数据区域映射的虚拟存储空间的分配,将已定制的区域映射至底层实际的存储资源中。从而建立逻辑上的数据区域与实际物理资源间的映射关系,以支撑大型计算应用的全局统一调度。



1. 跨域虚拟数据空间中区域划分及空间分配方法,其特征在于,基于由分散自治存储资源聚合成的全局名字空间,通过跨域的统一区域划分方法实现全局数据的统一管理和访问,并基于虚拟视图实现优化全局数据放置的空间分配;首先,制定虚拟数据区域和虚拟存储空间的模型化表述,为区域划分及空间分配提供理论基础;然后研究数据区域划分机制,在全局名字空间节点中实现和记录定制化的数据区域划分,以支持全局数据的统一管理和访问;最后,研究虚拟存储空间分配方法,实现数据区域映射的虚拟存储空间的分配,将已定制的区域映射至底层实际的存储资源中。从而建立逻辑上的数据区域与实际物理资源间的映射关系,以支撑大型计算应用的全局统一调度。

2. 根据权利要求1所述的方法,其特征在于,包括以下步骤:

1) 制定虚拟数据区域和虚拟存储空间的模型化表述;

2) 研究数据区域划分机制,在全局名字空间节点中实现和记录定制化的数据区域划分;

3) 基于数据区域划分,研究虚拟存储空间分配方法,实现数据区域映射的虚拟存储空间的分配,将已定制的区域映射至底层实际的存储资源中。

3. 根据权利要求2所述的方法,其特征在于,步骤1) 包括以下步骤:

步骤(1.1) 提出一种可定制化的数据区域模型,以五元组形式记录区域信息,为数据区域划分的可定制化提供理论基础。用于刻画数据区域信息的五元组可设计为AREA (zoneID, onename, owner, {member}, {spaceID});

步骤(1.2) 提出一种空间的模型化表述,以四元组形式记录空间信息,为虚拟存储空间分配提供理论基础。用于记录空间信息的四元组可设计为SPACE (spaceID, name, size, spacemetadata {hostcenter, storagesrcname, spacepath})。

步骤(1.3) 以此为基础,设计区域信息表、空间信息表等辅助记录区域和空间信息。在广域存储资源聚合完成的基础上,在区域空间服务器上构建数据库用于存放上述表中数据。

4. 根据权利要求2所述的方法,其特征在于,步骤2) 包括以下步骤:

步骤(2.1) 虚拟数据空间为虚拟数据空间用户提供一种可定制的数据区域划分机制,使得用户能按需自定义数据区域;

步骤(2.2) 通过区域划分机制,将聚合的分散自治存储资源划分为跨域的、用户可统一管理和访问的区域;

步骤(2.3) 以区域和空间模型为基础,将一个逻辑上的数据区域映射到一个或多个虚拟存储空间,实现逻辑区域到物理存储资源的映射,并结合安全管控机制,实现数据区域间的隔离与安全。

5. 根据权利要求2所述的方法,其特征在于,步骤3) 包括以下步骤:

步骤(3.1) 实现一种空间位置选择策略,基于用户定制信息,综合考虑虚拟存储空间性能距离、超算中心实际存储空间余量、已分配容量信息等因素,结合数据和计算任务的全局协同调度需求,根据空间位置选择算法选择分配空间的位置;

步骤(3.2) 基于定制的数据区域划分,提出一种有效的空间分配方法,在物理存储资源上构建虚拟存储空间,并建立与数据区域间的映射关系;

步骤(3.3) 将虚拟存储空间与实际的存储资源联结起来,支撑用户的存储、共享、读写

文件等操作,实现分散自治物理存储资源的统一透明分配,同时优化数据的全局放置策略,以支持大型计算应用的全局统一调度。

## 跨域虚拟数据空间中区域划分及空间分配方法

### 技术领域：

[0001] 本发明公开了一种跨域虚拟数据空间中区域划分及空间分配方法，涉及广域高性能计算面临的挑战，属于计算机技术领域。

### 背景技术：

[0002] 当前，国家高性能计算环境中存储资源广域分散且隔离自治，大型计算应用迫切需要可支持跨域统一访问的全局数据空间。在此基础上，实现跨域的统一区域划分方法及基于虚拟视图的空间分配方法，按需分配和布局存储资源并实现虚拟区域与物理资源的映射转换，以屏蔽底层资源的异构性和跨域分散问题，从而有效支撑存储资源从物理上分散自治到逻辑上统一聚合的全局数据管理和访问。当前在面向高性能计算领域的区域划分和空间分配方法的研究领域中，许多学者进行了大量的研究，总结有如下两类：面向网格计算的区域划分和空间分配方法、面向云计算的区域划分和空间分配方法。

[0003] 云计算是指通过互联网为用户提供动态虚拟化的计算和存储等资源，由于其便捷性和可用性，目前在互联网领域被广泛应用和发展如Windows Azure Storage、SPANStore、OceanStore等。在云计算系统的基础上，云存储技术也有了长足的发展，云存储系统通过广域网将多种异构的存储资源聚合起来，共同为用户提供一种便捷可用的数据存储和访问功能。

[0004] Windows Azure Storage (WAS) 提供了一种全局可扩展的、低成本的云存储系统。WAS提供了一种全局可扩展的名字空间，将许多不同用户的工作负载与不同的资源需求组合在一起，这样在任何时间点提供的存储需求都比在它们自己专用的硬件上运行这些服务要少得多。在区域划分和空间分配方法上，WAS在广域范围内提供多个跨域的StorageStamp，各StorageStamp之间提供备份机制，并在StorageStamp上WAS为每个用户提供一个或多个存储名称空间，并将存储名称空间分为三部分：帐户名称、分区名称和对象名称。其中，帐户名称用于实现不同账户之间存储名称空间的隔离，分区名称用于提供跨存储节点的数据访问，对象名称提供实际资源的物理地址同时，WAS将DNS融入了存储名称空间，因此，所有数据都可以通过存储在WAS中存储名称空间表单的URI访问。WAS系统通过上述方法实现了对数据的跨域存储与访问，以及基于用户的存储空间隔离与安全。

[0005] SPANStore系统是一种低延迟、低成本的云存储系统，SPANStore联合了多个云存储服务提供商，提供了一种跨多个云的数据存储服务，包括Amazon S3、Windows Azure、GCS等。每个应用程序都使用一个单独的SPANStore部署。在部署应用程序的每个数据中心，应用程序都向应用程序链接的SPANStore库发出对对象的PUT和GET请求。在SPANStore中，每个数据中心都是一个访问集，各个数据中心分别存储元数据。SPANStore通过上述方法实现了云存储系统的低成本和低延迟。

[0006] OceanStore系统旨在提供一种有更好的可用性、性能和可靠性的云存储系统，其核心是一个个跨域分布的Pool，各个Pool之间数据可相互流通，一个用户与一个或多个Pool相连接。OceanStore系统采用一种广域分布式数据位置的全局算法进行数据放置优

化,OceanStore中的每个可寻址实体(例如浮动副本、存档片段或客户机)都由一个或多个GUID标识。功能相同的实体,如相同对象的不同副本,由相同的GUID标识,客户端使用一系列协议消息与这些实体进行交互。OceanStore以跨域分布且相互连通的Pool为核心,以Pool连接下层存储资源,以GUID标识数据资源,形成了一种高可用的云存储系统。

[0007] MetaStorage系统是一种高可扩展、高可用、基于分布式哈希表的云存储系统。MetaStorage系统以多个跨域分布的Distributor连接到实际存储资源,一个Distributor可用连接到许多的本地存储节点、Amazon S3存储节点等。在存储文件数据时,MetaStorage系统基于hash通过Distributor实现对文件数据的位置分配,同时MetaStorage系统以一个基于仲裁的哈希键值首选项列表代替并消除了分区算法,作负载均衡策略。为保证多个跨域分布的Distributor间首选项列表的一致性,MetaStorage系统在Distributor的基础上构建了Coordinator以管理底层Distributor的配置及成员的最新状态。

[0008] 网格计算相较于云计算,其目的在于将广域分散的网格资源聚合起来,以支撑大型的计算应用。网格计算通过中间件屏蔽底层资源的异构性,同时通过网格将异构的资源整合利用,支持虚拟组织,提供高层次的服务,并提供巨大的处理能力。面向网格计算的存储系统如XSEDE系统的全局联合文件系统(GFFS)、EGI系统的Onedata等都需要提供一种屏蔽底层存储资源异构性的存储方法,同时,为满足高性能计算的需求,还需提供全局的高效、安全访问机制。

[0009] 美国XSEDE项目旨在将分布式资源和工具在研究和教育领域整合利用,为研究工作提供一个极端的科学和工程发现环境,以向更全面、更有凝聚力的数字服务体系迈进一步。XSEDE项目开发的存储系统GFFS基于RNS实现了全局名字空间,连通了资源供应商的存储资源、大学集群的存储资源、国家超算中心的存储资源,并支持导入个人目录作为存储资源,基于Genesis 2的container实现了存储资源的管理,以支持存储资源的加入、注册和退出。同时,GFFS支持个人计算机、大学校园存储节点、超算中心节点等多个场景的访问模式,支持在各种场景下,基于全局名字空间,对多种存储资源的统一访问和管理。通过统一名字空间,GFFS实现了对跨域异构存储资源的统一访问、管理和共享。

[0010] 欧洲EGI项目通过Onedata提供数据和元数据管理服务,Onedata专注于对分布式数据集的即时、透明访问,无需不必要的暂存和迁移,允许直接从本地计算机或工作节点访问数据。在Onedata中,所有数据都将被组织到相应的Space中,Space由存储资源提供商Provider提供存储能力。Space可以被视为虚拟文件夹或卷,它隶属于某个用户或某个组,可以在用户间共享,可以包含任意目录和文件层次结构,也可以同时分布在多个Provider中。Provider可以被视为超算中心的存储集群,他们直接为Space的所有者提供数据存储服务。如果多个Provider支持同一个Space,那么Space的元数据及数据将在这些Provider之间同步。

[0011] 日本筑波大学研发的Gfarm网格文件系统也是一种面向网络的存储系统,Gfarm可以联合多个PC集群的本地文件系统,整合现存本地文件系统的目录,这种聚合方式没有对底层存储资源额外的硬件需求,但数据访问性能可能劣于对象存储或块存储。各PC集群的文件、目录由各PC集群分别管理,存储数据时,Gfarm在顶层基于各PC集群CPU负载、可用容量、数据访问往返时延(RTT)选择文件存储位置。

[0012] 传统的区域划分和空间分配方法存在的问题主要为:

[0013] 面向云计算的区域划分和空间分配方法对存储资源管理松散,主要以数据中心的形式提供底层资源的使用,以相对集中的资源运行分散的应用,承认资源的异构性且不强调上层的虚拟组织,对于数据的全局访问和统一管理存在一定局限性,不适用于当前国家高性能计算环境。面向网格计算的区域划分和空间分配方法可以有效的使用聚合的全局资源,但是需要专用网络支持提升性能,并且对数据的管理依然松散,不利于大型应用的全局统一调度,数据放置策略大多以数据中心为单位进行优化,缺乏基于虚拟视图的全局统一放置策略。因此,研究适用于国家高性能计算环境的虚拟数据空间中区域划分和空间分配方法,将分散资源聚合而成的虚拟数据空间按需划分为用户定制化的数据区域,并实现区域隔离与安全,以支持存储资源从物理上分散自治到逻辑上统一聚合的全局数据管理和访问,同时实现分散自治物理存储资源的统一透明分配,优化数据的全局放置策略,以支持大型计算应用的全局统一调度。

#### 发明内容:

[0014] 本发明的主要目的是提供一种跨域虚拟数据空间中区域划分及空间分配方法,针对大型计算应用使用资源的特点以及对数据安全的需求,实现了虚拟数据空间的区域划分方法,将虚拟数据空间按需划分为用户定制化的数据区域,并基于区域间访问控制机制实现区域隔离与安全,以支持存储资源从物理上分散自治到逻辑上统一聚合的全局数据管理和访问;针对数据和计算任务的全局调度需求,实现了虚拟数据空间中存储空间分配方法,实现分散自治物理存储资源的统一透明分配,同时优化数据的全局放置策略,以支持大型计算应用的全局统一调度。

[0015] 本发明的技术方案是:

[0016] 一种跨域虚拟数据空间中区域划分及空间分配方法,其特征在于,基于由分散自治存储资源聚合成的全局名字空间,通过跨域的统一区域划分方法实现全局数据的统一管理和访问,并基于虚拟视图实现优化全局数据放置的空间分配。首先,制定虚拟数据区域和虚拟存储空间的模型化表述,为区域划分及空间分配提供理论基础;然后研究数据区域划分机制,在全局名字空间节点中实现和记录定制化的数据区域划分,以支持全局数据的统一管理和访问;最后,研究虚拟存储空间分配方法,实现数据区域映射的虚拟存储空间的分配,将已定制的区域映射至底层实际的存储资源中。从而建立逻辑上的数据区域与实际物理资源间的映射关系,以支撑大型计算应用的全局统一调度。

[0017] 包括3个步骤,具体如下:

[0018] 1) 制定虚拟数据区域和虚拟存储空间的模型化表述;

[0019] 2) 研究数据区域划分机制,在全局名字空间节点中实现和记录定制化的数据区域划分;

[0020] 3) 基于数据区域划分,研究虚拟存储空间分配方法,实现数据区域映射的虚拟存储空间的分配,将已定制的区域映射至底层实际的存储资源中。

[0021] 其中,步骤1) 包括以下步骤:

[0022] 步骤(1.1) 提出一种可定制化的数据区域模型,以五元组形式记录区域信息,为数据区域划分的可定制化提供理论基础。用于刻画数据区域信息的五元组可设计为AREA (zoneID, onename, owner, {member}, {spaceID});

[0023] 步骤(1.2)提出一种空间的模型化表述,以四元组形式记录空间信息,为虚拟存储空间分配提供理论基础。用于记录空间信息的四元组可设计为SPACE(spaceID,name,size,spacemetadata{hostcenter,storagesrcname,spacepath});

[0024] 步骤(1.3)以此为基础,设计区域信息表、空间信息表等辅助记录区域和空间信息。在广域存储资源聚合完成的基础上,在区域空间服务器上构建数据库用于存放上述表中数据。

[0025] 其中,步骤2)包括以下步骤:

[0026] 步骤(2.1)虚拟数据空间为虚拟数据空间用户提供一种可定制的数据区域划分机制,使得用户能按需自定义数据区域;例如:定义只能由用户个人访问的数据区域,作为用户的home分区,存放用户个人的私有数据;又如:定义化学分区,作为一个由多个用户组成的化学工作组的数据区域,工作组内的用户都可访问分区数据,共享实验数据等;此外,还可定义公共数据分区,用于存放一些可供全体用户共享访问的数据,如:某些通用软件及其使用手册等,供所有用户共享;

[0027] 步骤(2.2)通过区域划分机制,将聚合的分散自治存储资源划分为跨域的、用户可统一管理和访问的区域;

[0028] 步骤(2.3)以区域和空间模型为基础,将一个逻辑上的数据区域映射到一个或多个虚拟存储空间,实现逻辑区域到物理存储资源的映射,并结合安全管控机制,实现数据区域间的隔离与安全。

[0029] 其中,步骤3)包括以下步骤:

[0030] 步骤(3.1)实现一种空间位置选择策略,基于用户定制信息,综合考虑虚拟存储空间性能距离、超算中心实际存储空间余量、已分配容量信息等因素,结合数据和计算任务的全局协同调度需求,根据空间位置选择算法选择分配空间的位置;

[0031] 步骤(3.2)基于定制的数据区域划分,提出一种有效的空间分配方法,在物理存储资源上构建虚拟存储空间,并建立与数据区域间的映射关系;

[0032] 步骤(3.3)将虚拟存储空间与实际的存储资源联结起来,支撑用户的存储、共享、读写文件等操作,实现分散自治物理存储资源的统一透明分配,同时优化数据的全局放置策略,以支持大型计算应用的全局统一调度。

[0033] 本发明的优点包括:

[0034] 本发明提供了一种跨域虚拟数据空间中区域划分及空间分配方法,与现有的数据管理技术相比,其主要优点是:

[0035] 通过虚拟数据区域与数据存储空间的映射关系,可以实现对跨域分散数据的统一调度和管理,可以更有效的满足高性能计算环境中大型计算应用的资源需求;基于虚拟视图的空间分配方法,可以平衡分散存储资源的负载,并实现优化的全局数据放置策略,以优化应用计算过程中的存储与计算资源协同调度效率。

#### 附图说明:

[0036] 图1为跨域虚拟数据空间中区域划分及空间分配方法实施流程图

[0037] 图2为虚拟数据空间中区域与空间架构图

[0038] 图3为数据区域划分及虚拟存储空间分配流程图

[0039] 图4为区域共享操作步骤

[0040] 图5为区域空间定位信息访问流程图

### 具体实施方式：

[0041] 以下结合附图对本发明作进一步详细的说明。

[0042] 如图1所示,是本发明的跨域虚拟数据空间中区域划分及空间分配方法实施流程图。跨域虚拟数据空间中区域划分及空间分配方法,包括以下步骤:

[0043] 1) 制定虚拟数据区域和虚拟存储空间模型化表述;

[0044] 2) 研究数据区域划分机制,在全局名字空间节点中实现和记录定制化的数据区域划分;

[0045] 3) 基于数据区域划分,研究虚拟存储空间分配方法,实现数据区域映射的虚拟存储空间的分配,将已定制的区域映射至底层实际的存储资源中。

[0046] 虚拟数据空间中区域与空间的架构如图2所示,首先将数据区域和虚拟存储空间进行模型化表述,构建数据库记录数据区域和虚拟存储空间的信息;其次通过区域划分机制将由跨域分散自治存储资源聚合而成的全局数据空间划分为用户定制的可统一管理和访问数据的区域,实现逻辑区域到物理存储资源的映射,并结合安全管控机制,实现数据区域间的隔离与安全;然后通过虚拟存储空间分配方法为数据区域在物理存储资源上构建虚拟存储空间,并建立二者之间的映射关系,实现分散自治物理存储资源的统一透明分配,同时优化了数据的全局放置策略,提升了大型计算应用以及数据和计算任务的全局协同调度策略的访问性能。

[0047] 数据区域划分及虚拟存储空间分配的流程如图3所示,包括以下步骤:

[0048] 1) 用户使用虚拟数据空间软件时,需要首先申请一个自定义的区域,由客户端发出申请区域请求并包含区域定制信息;

[0049] 2) 申请区域请求发送到管理节点的虚拟数据空间区域划分模块,并由区域划分模块接受用户的申请后,根据用户的定制信息,选择区域划分策略;

[0050] 3) 管理节点上区域划分模块调用虚拟数据空间的区域分配模块进行虚拟存储空间的分配,以将数据区域映射到虚拟存储空间;

[0051] 4) 由空间位置选择模块基于空间位置选择算法,选定区域对应空间要创建在哪个存储资源上;

[0052] 5) 通过空间分配模块,将创建目录的请求发送到选定超算中心的存储集群客户端中,将创建好的目录作为虚拟存储空间,并设置初始权限;

[0053] 6) 把创建的虚拟存储空间与数据区域进行映射,并将空间信息以四元组的形式记录在区域空间服务器上的空间信息表中,将区域的权限信息(如用户的读、写、执行权限等)记录在全局区域权限表中;

[0054] 7) 将用户数据区域信息以五元组的形式记录在区域空间服务器上的区域信息表中,通过数据库保障各超算节点间区域和空间信息的弱一致性,至此,区域划分和空间分配完成。

[0055] 在构建的数据区域以及虚拟存储空间基础上,区域共享操作步骤如图4所示,包括以下步骤:

- [0056] 1) 用户通过客户端进行区域共享申请,包括增加/删除组区域成员;
- [0057] 2) 客户端选择最优管理节点进行连接,并发送区域共享请求,管理节点上的区域共享模块接收到请求,并查询区域对应的全部空间以及空间定位信息;
- [0058] 3) 区域共享模块调用权限增加/权限删除模块,根据空间路径信息在区域映射的全部空间所对应的存储集群中更改用户权限;
- [0059] 4) 权限变更完成后,在数据库中更新相应的区域成员信息,并返还客户端区域共享操作结果。
- [0060] 基于数据区域划分及虚拟存储空间分配,区域空间定位信息访问流程如图5所示。包括以下步骤:
- [0061] 1) 客户端发送获取区域定位信息请求;
- [0062] 2) 管理节点中区域定位模块接受请求,并向区域空间服务器获取区域映射的空间信息;
- [0063] 3) 区域定位模块向空间定位模块发送获取空间定位信息的请求;
- [0064] 4) 空间定位模块向区域空间服务器获取区域映射的全部空间的定位信息;
- [0065] 5) 区域定位模块向客户端返回定位信息结果。
- [0066] 最后所应说明的是:本发明还可有其它多种应用场景,在不背离本发明精神及其实质的情况下,熟悉本领域的技术人员当可根据本发明做出各种相应的改变和变形,但这些相应的改变和变形都应属于本发明所附的权利要求的保护范围。

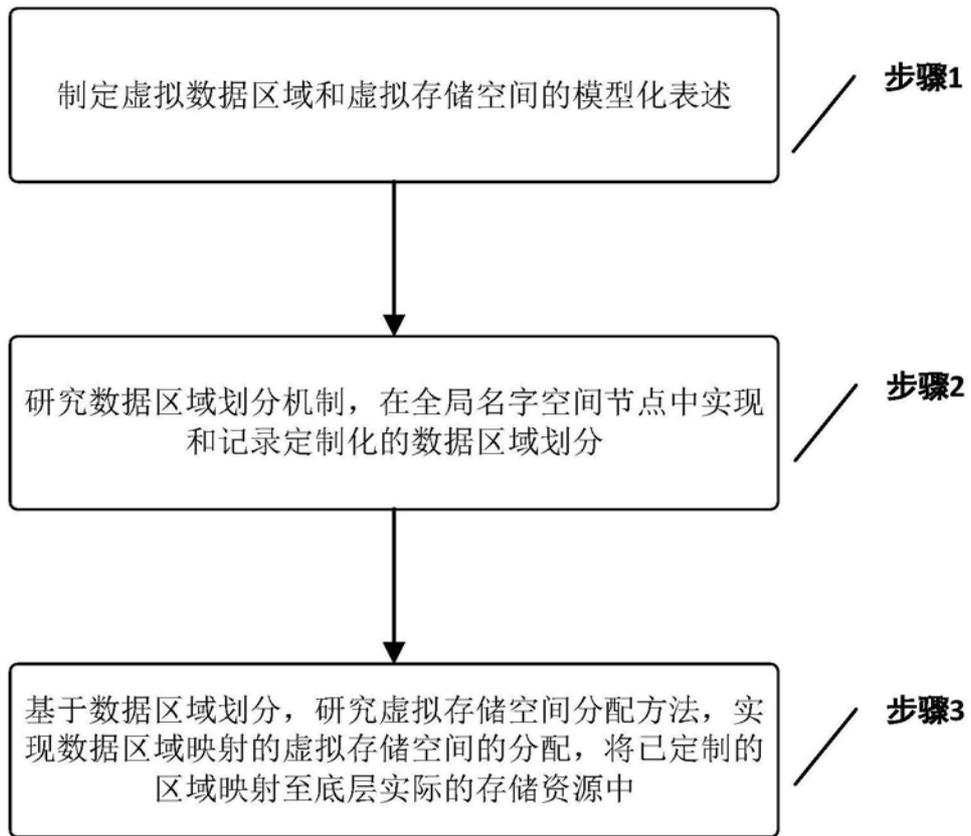


图1

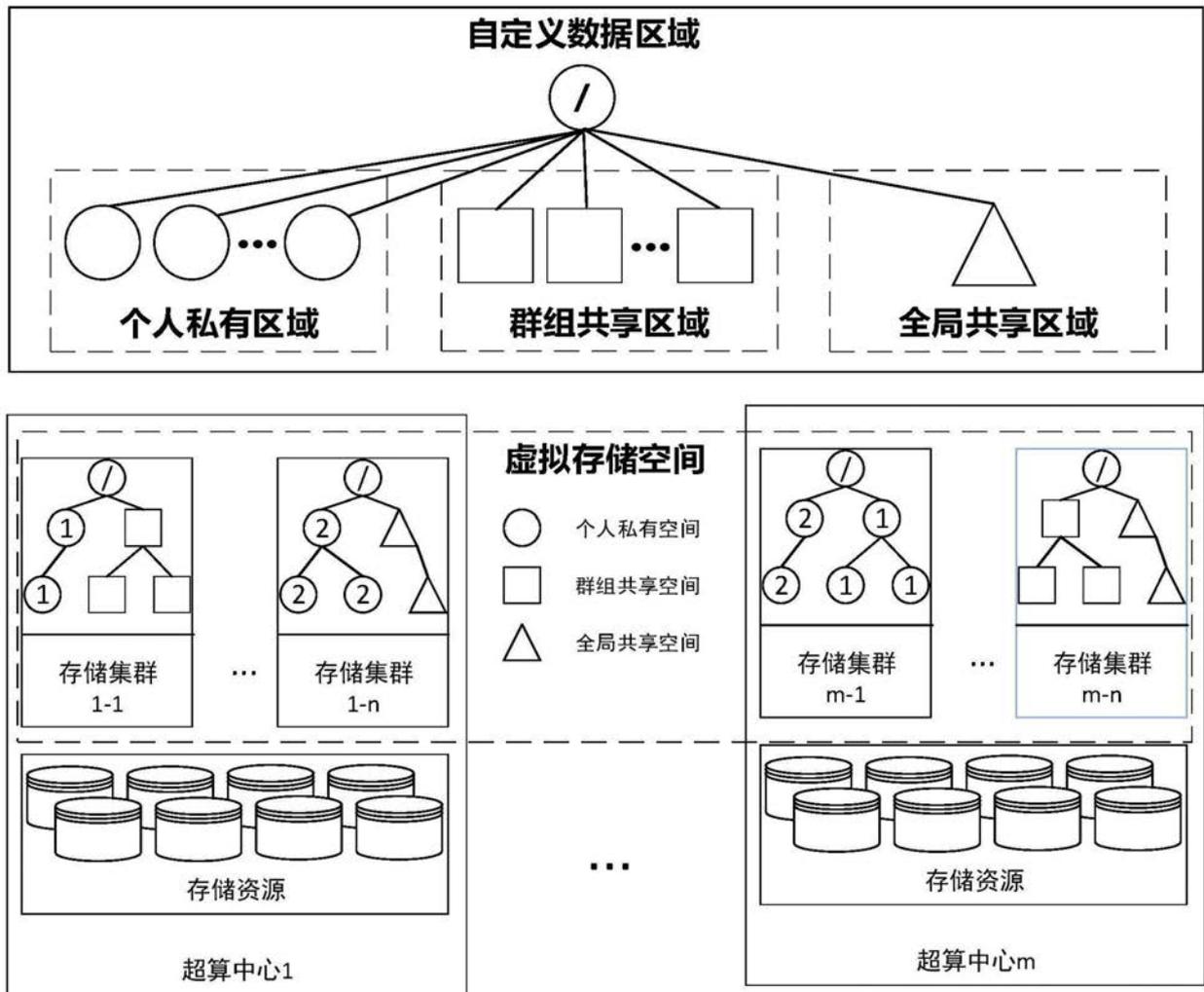


图2

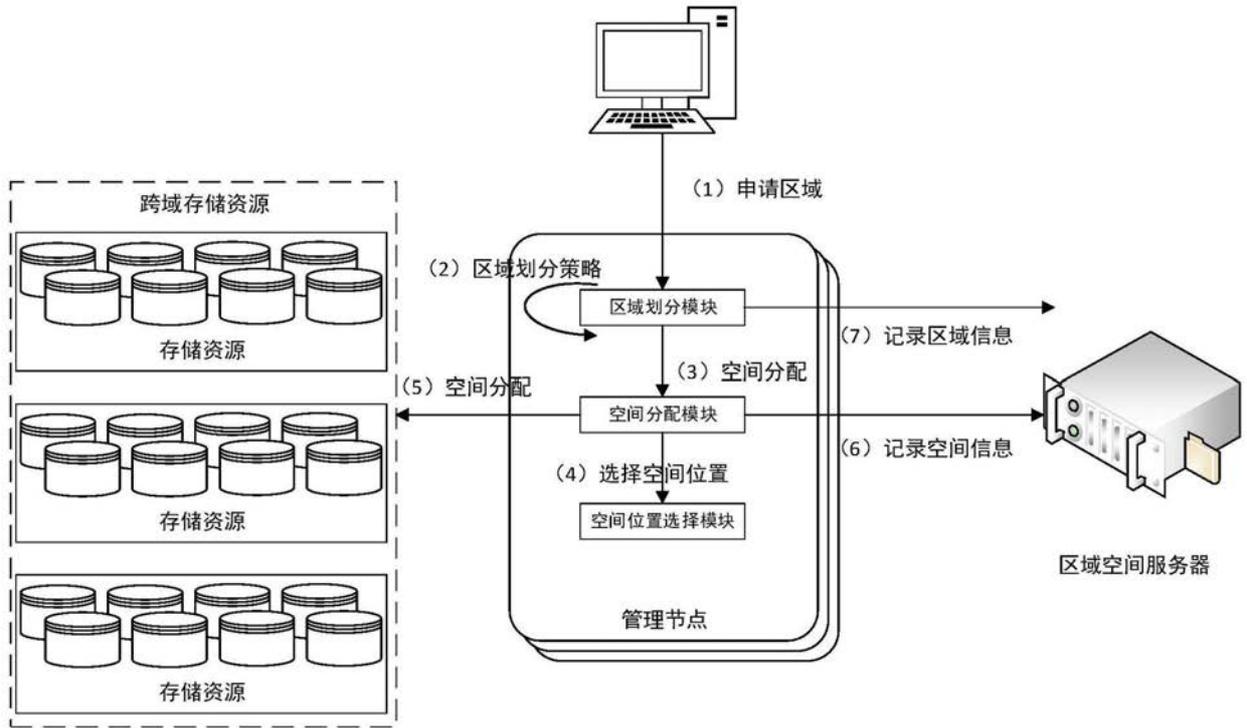


图3

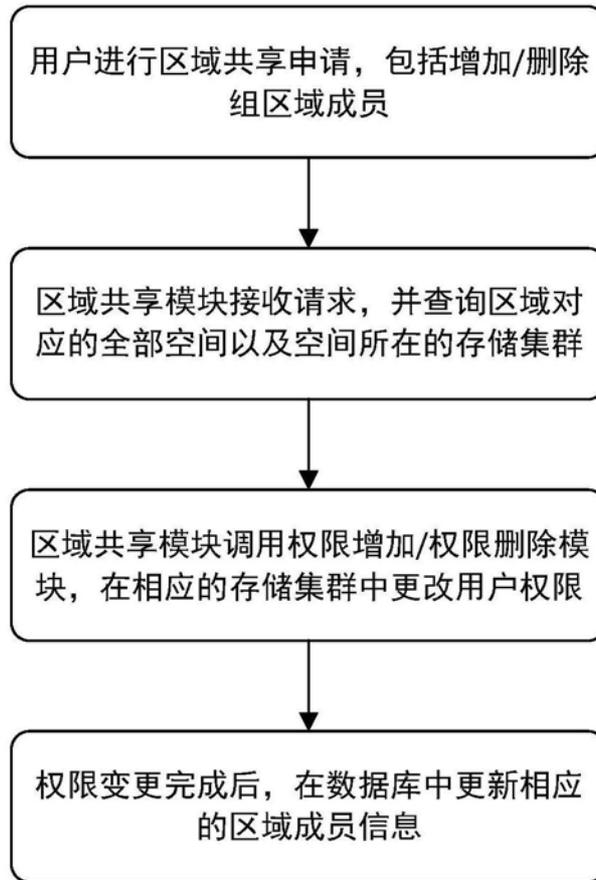


图4

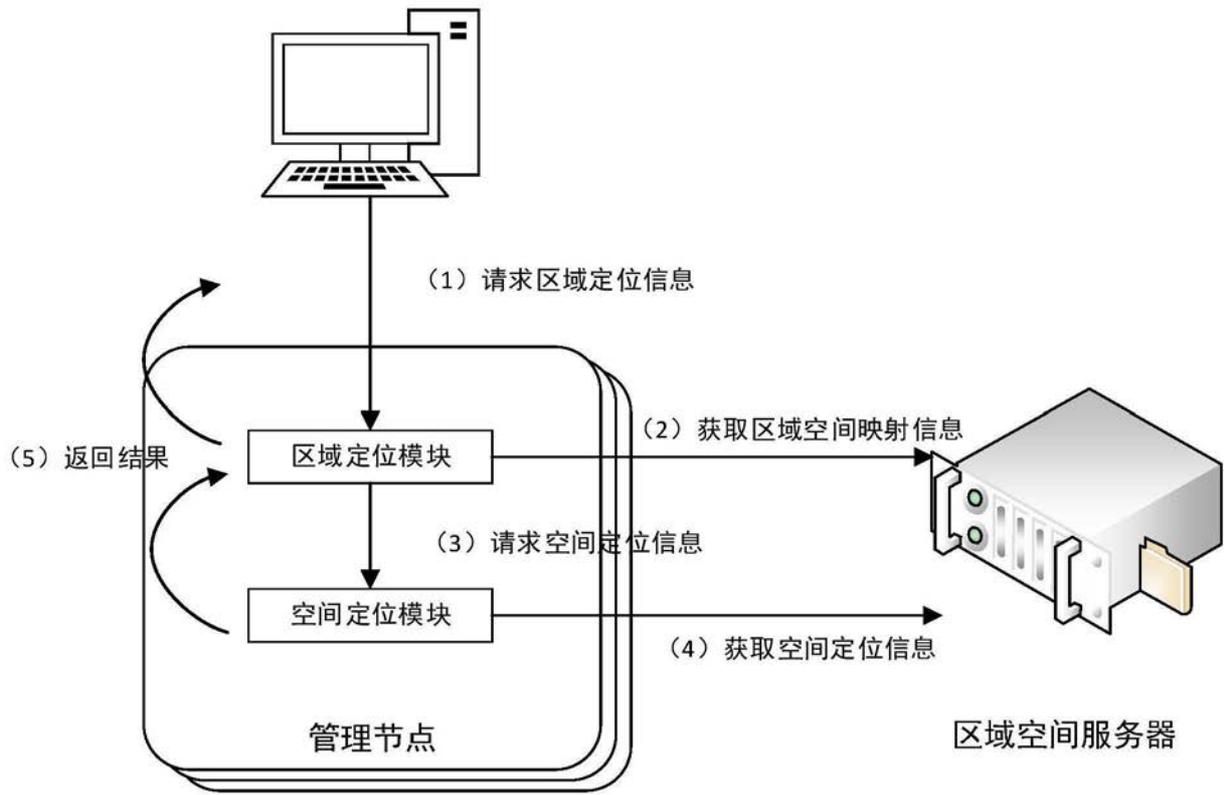


图5