



(12) 发明专利申请

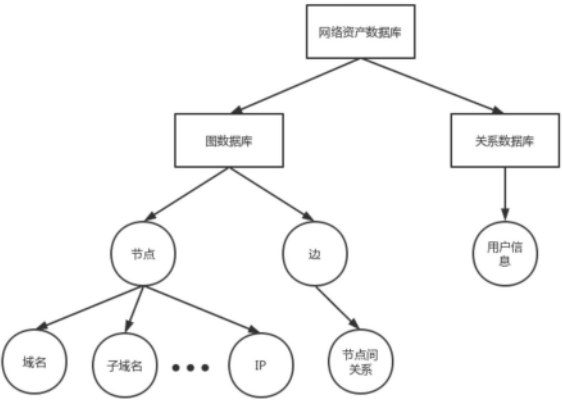
(10) 申请公布号 CN 116821155 A
(43) 申请公布日 2023. 09. 29

(21) 申请号 202310767450.2
(22) 申请日 2023.06.27
(71) 申请人 上海滕龙科技有限公司
地址 201100 上海市闵行区东川路555号乙楼1层1001室
(72) 发明人 王昊天
(74) 专利代理机构 北京维正专利代理有限公司
11508
专利代理师 邓骏杰
(51) Int.Cl.
G06F 16/242 (2019.01)
G06F 16/28 (2019.01)
G06F 16/54 (2019.01)
G06F 16/22 (2019.01)

权利要求书1页 说明书6页 附图2页

(54) 发明名称
一种网络资产数据存储和查询方法、装置及存储介质
(57) 摘要

本申请公开了一种网络资产数据存储和查询方法、装置及存储介质,包括:将包括若干个第二文档的第一文档存储至非关系型数据库中;第二文档包括时间数据及与时间数据对应的网络资产数据,时间数据为网络资产数据生成和/或修改的时间;根据属性数据建立查询索引然后建立倒排索引表,以使搜索引擎并行查询存储于若干个节点内的若干个数据组。本申请实现了自动化的网络资产管理和监控,提高管理效率和精度,具有良好的横向扩展性和分布式处理能力,可以满足大规模数据的存储和处理需求,提高数据存储和查询的速度和可扩展性。



1. 一种网络资产数据存储方法,其特征在于,包括以下步骤:

将包括若干个第二文档的第一文档存储至非关系型数据库中;所述第二文档包括时间数据及与所述时间数据对应的网络资产数据,所述时间数据为所述网络资产数据生成和/或修改的时间。

2. 根据权利要求1所述的网络资产数据存储方法,其特征在于:所述非关系型数据库为图数据库。

3. 根据权利要求2所述的网络资产数据存储方法,其特征在于:所述网络资产数据包括组织数据、域名数据、子域名数据、IP数据、端口数据和组件数据以及不同数据之间的关系;所述组织数据、域名数据、子域名数据、IP数据、端口数据和组件数据在所述图数据库中存储为节点,不同数据之间的关系在所述图数据库中存储为边向量,所述节点之间通过所述边向量连接。

4. 根据权利要求3所述的网络资产数据存储方法,其特征在于:所述节点和所述边向量还包括属性数据,所述属性数据为所述网络资产数据的属性。

5. 根据权利要求1或3所述的网络资产数据存储方法,其特征在于:还包括计算所述网络资产数据的热度值,将所述网络资产数据中热度值大于预设的热度阈值的资产热数据,对所述资产热数据优先进行预处理并缓存。

6. 根据权利要求5所述的网络资产数据存储方法,其特征在于,还包括将所述时间数据及与所述时间数据对应的网络资产数据划分为若干个数据组,并将若干个所述数据组分别存储于不同的节点内。

7. 一种网络资产数据查询方法,应用如权利要求1-6中任一项所述的网络资产数据存储方法,其特征在于,包括以下步骤:

根据属性数据建立查询索引;

根据所述查询索引建立倒排索引表,以使搜索引擎并行查询存储于若干个节点内的若干个数据组。

8. 一种网络资产数据存储和查询装置,其特征在于,包括:

存储模块,用于将包括若干个第二文档的第一文档存储至非关系型数据库中;

查询模块,用于通过搜索引擎并行查询存储于若干个节点内的若干个数据组。

9. 一种计算设备,其特征在于:包括存储器、处理器以及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述处理器执行所述程序时实现如权利要求1到7任一项所述方法。

10. 一种计算机可读存储介质,其特征在于:存储有能够被处理器加载并执行如权利要求1到7中任一项所述方法的计算机程序。

一种网络资产数据存储和查询方法、装置及存储介质

技术领域

[0001] 本申请涉及网络资产数据存储技术领域,尤其是涉及一种网络资产数据存储和查询方法、装置及存储介质。

背景技术

[0002] 在网络资产数据存储和查询领域,存在的一些现有的方法,比如:

关系型数据库:传统的关系型数据库采用基于行和列的二维表存储结构,其中每行表示一个记录,每列表示一个字段。关系型数据库使用SQL作为查询语言,支持关系运算、聚合函数等多种操作。相对于其他数据库,关系型数据库在数据一致性、数据完整性、事务处理等方面具有很强的优势。

[0003] 非关系型数据库:非关系型数据库通常采用键-值、列族、文档等不同的数据模型来存储数据。非关系型数据库在横向扩展和分布式架构方面具有很强的优势,因此在大数据处理、互联网应用、实时数据处理等领域得到广泛应用。

[0004] 基于索引的存储:基于索引的存储可以通过对文本数据建立倒排索引等方式提高查询效率。在基于索引的存储中,数据被分为一个个文档,每个文档都有自己的词汇表和倒排索引。用户可以通过查询词汇表和倒排索引来快速地查找文档。这种技术通常用于全文搜索、数据挖掘等场景。

[0005] 虽然存在许多针对网络资产的查询和存储的数据库方案,但是现有的处理方式都存在一定的缺陷。针对于关系型数据库,关系型数据库在数据结构和查询方式上具有一定的局限性,无法直接表达和查询复杂的关系网络,也无法高效地处理大量的网络资产图数据。同时,在大规模分布式环境下,关系型数据库的性能和可扩展性也存在瓶颈;而对于一般的非关系型数据库,虽然非关系型数据库在分布式存储和横向扩展方面具有很强的优势,但是对于复杂的关系网络的存储和查询能力较弱,也不太适合网络资产图数据处理;而基于索引的存储可以通过建立倒排索引等方式提高查询效率,但是对于大规模的图数据处理还是存在一定的局限性。在复杂的网络资产存储和查询场景下,基于索引的存储的查询速度和效率都比较低。

发明内容

[0006] 本申请提供了一种网络资产数据存储和查询方法、装置及存储介质,能够满足大量复杂的网络资产数据的存储和快捷的查询。

[0007] 第一方面,本申请提供一种网络资产数据存储方法,采用如下的技术方案:

一种网络资产数据存储方法,包括以下步骤:

将包括若干个第二文档的第一文档存储至非关系型数据库中;所述第二文档包括时间数据及与所述时间数据对应的网络资产数据,所述时间数据为所述网络资产数据生成和/或修改的时间。

[0008] 进一步地,上述非关系型数据库为图数据库。

[0009] 进一步地,上述网络资产数据包括组织数据、域名数据、子域名数据、IP数据、端口数据和组件数据以及不同数据之间的关系;所述组织数据、域名数据、子域名数据、IP数据、端口数据和组件数据在所述图数据库中存储为节点,不同数据之间的关系在所述图数据库中存储为边向量,所述节点之间通过所述边向量连接。

[0010] 进一步地,上述节点和所述边向量还包括属性数据,所述属性数据为所述网络资产数据的属性。

[0011] 进一步地,上述方法还包括计算所述网络资产数据的热度值,将所述网络资产数据中热度值大于预设的热度阈值的资产热数据,对所述资产热数据优先进行预处理并缓存。

[0012] 进一步地,上述方法还包括将所述时间数据及与所述时间数据对应的网络资产数据划分为若干个数据组,并将若干个所述数据组分别存储于不同的节点内。

[0013] 第二方面,本申请还提供了一种网络资产数据查询方法,应用上述的网络资产数据存储方法,包括以下步骤:

根据属性数据建立查询索引;

根据所述查询索引建立倒排索引表,以使搜索引擎并行查询存储于若干个节点内的若干个数据组。

[0014] 第三方面,本申请还提供了一种网络资产数据存储和查询装置,包括:

存储模块,用于将包括若干个第二文档的第一文档存储至非关系型数据库中;

查询模块,用于通过搜索引擎并行查询存储于若干个节点内的若干个数据组。

[0015] 第四方面,本申请提供一种计算机设备,采用如下的技术方案:

一种计算机设备,包括存储器、处理器以及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述处理器执行所述程序时实现如第一方面所述的方法。

[0016] 第五方面,本申请提供一种计算机可读存储介质,采用如下的技术方案:

一种计算机可读存储介质,存储有能够被处理器加载并执行如第一方面中任一种方法的计算机程序。

[0017] 综上所述,本申请包括以下至少一种有益技术效果:

本申请提供的一种网络资产数据存储和查询方法、装置及存储介质,实现了自动化的网络资产管理和监控,提高管理效率和精度,具有良好的横向扩展性和分布式处理能力,可以满足大规模数据的存储和处理需求,提高数据存储和查询的速度和可扩展性。

附图说明

[0018] 图1是本发明实施方式中网络资产数据库结构示意图。

[0019] 图2是本发明实施方式中网络资产数据结构示意图。

[0020] 图3是本发明实施方式中网络资产数据库数据同步方法流程示意图。

具体实施方式

[0021] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处所描述的具体实施例仅用以解释本申请,并不用于限定本申请。

[0022] 本申请实施方式为了满足大量的复杂的网络资产图数据的存储和快捷的查询,基于图数据库,提供了一种网络资产数据存储和查询方法、装置及存储介质,可以直接建模和查询复杂的关系网络,提高数据处理的效率和精度。图数据库是以点、边为基础存储单元,以高效存储、查询图数据为设计原理的数据管理系统,“点”表示实体,“边”表示实体间的关系。在图数据库中,数据间的关系和数据本身同样重要,它们被作为数据的一部分存储起来。这样的架构使图数据库能够快速响应复杂关联查询,因为实体间的关系已经提前存储到了数据库中。图数据库可以直观地可视化关系,是存储、查询、分析高度互联数据的最优办法,常见的图数据库包括如Neo4j、Galaxybase、GraphDB等。

[0023] 关系型数据库是指采用了关系模型来组织数据的数据库,其以行和列的形式存储数据,以便于用户理解,关系型数据库这一系列的行和列被称为表,一组表组成了数据库。用户通过查询来检索数据库中的数据,而查询是一个用于限定数据库中某些区域的执行代码。关系模型可以简单理解为二维表格模型,而一个关系型数据库就是由二维表及其之间的关系组成的一个数据组织。主流的关系型数据库包括Oracle、DB2、MySQL、Microsoft SQL Server、Microsoft Access等。

[0024] 如图2所示,网络资产数据整体上来说是具有层次关系和连接关系的图数据,例如:“域名->子域名->IP->组件”这样的数据,每个域名拥有多个子域名,每个子域名拥有对应的IP而每个IP代表的主机中拥有不同的组件,这些域名、子域名等数据都可以称之为网络资产,网络资产自建拥有层次关系,可以构建出这些网络资产之间的拓扑关系图,这些拓扑关系图数据是非常适合存储在图数据库中的,因为图数据库可以有效地表示和查询这种拓扑结构。

[0025] 如图1所示,本申请实施例公开了一种网络资产数据存储方法,包括以下步骤:

将包括若干个第二文档的第一文档存储至非关系型数据库中;所述第二文档包括时间数据及与所述时间数据对应的网络资产数据,所述时间数据为所述网络资产数据生成和/或修改的时间,上述非关系型数据库为图数据库。

[0026] 在本实施方式中,由于关系型数据库在数据结构和查询方式上具有一定的局限性,无法直接表达和查询复杂的关系网络,也无法高效地处理大量的网络资产图数据,因此,本申请提供了一种基于图数据库的网络资产数据存储方法,首先,使用本申请的技术方案采集企业所有的网络资产数据,包括域名、IP地址、设备信息等,并将这些数据存储到图数据库中,其中,第一文档可以包括若干个第二文档,第二文档可以是将网络资产数据抽象为节点和有向边之后的一类网络资产数据,以及这些网络资产数据生成或存储的时间数据。其中,一类网络资产数据可以是一种网络资产数据在不同时间点的不同版本,或一个时间点下的各种网络资产数据,此处不设限制;

在本发明的其中一种实施方式中,网络资产数据包括组织数据、域名数据、子域名数据、IP数据、端口数据和组件数据以及不同数据之间的关系;所述组织数据、域名数据、子域名数据、IP数据、端口数据和组件数据在所述图数据库中存储为节点,不同数据之间的关系在所述图数据库中存储为边向量,所述节点之间通过所述边向量连接,节点和边向量还包括属性数据,所述属性数据为所述网络资产数据的属性。

[0027] 在本实施方式中,对于网络资产数据,可以将其抽象为节点和边向量,节点表示网络资产,边向量表示不同网络资产之间的关系,如域名到子域名、子域名到IP地址等。在

我们的图数据库存储引擎中,每个节点和边向量都有自己的属性列表,可以为节点和边添加任意数量的属性,基于这些属性数据,当存储大量节点和边时,图数据库存储引擎可以使用索引和缓存等优化技术来加速节点和边向量之间的查询和导航。

[0028] 在本申请的其中一种实施方式中,上述方法还包括计算所述网络资产数据的热度值,将所述网络资产数据中热度值大于预设的热度阈值的资产热数据,对所述资产热数据优先进行预处理并缓存。

[0029] 在本实施方式中,本申请的技术方案首先采集企业所有的网络资产数据,包括域名、IP地址、设备信息等,并将这些数据存储到图数据库中。可以采用预处理和缓存的方式,将常用的查询结果预先计算并缓存,以便快速响应用户的查询请求。

[0030] 在本申请的其中一种实施方式中,上述方法还包括将所述时间数据及与所述时间数据对应的网络资产数据划分为若干个数据组,并将若干个所述数据组分别存储于不同的节点内。

[0031] 如图3所示,在本实施方式中,为了支持大规模数据的存储、同步、处理和查询,可以对网络资产数据进行分布式部署,将网络资产数据存储到多个节点上,使得数据库具有高可用性、可靠性和扩展性,为了保证多个数据库之间数据一致性和可靠性,本申请的技术方案使用了数据库同步工具,数据库同步工具使用消息队列技术,将数据源中的数据采集到一个消息队列中,并将该消息队列中的数据推送到目标数据库中,实现不同节点数据库信息的同步。

[0032] 如图1所示,为了满足更多的实际需求,本申请的技术方案还结合了关系型数据库的存储能力,用于存储一些不需要存储到图数据库中的一些少量信息,如用户信息等。这些信息可以采用关系型数据库的方式进行存储和查询,同时与图数据库中的节点和边建立关联关系,从而实现全面的数据管理和查询功能。

[0033] 本申请实施例还公开了一种网络资产数据查询方法,应用上述的网络资产数据存储方法,包括以下步骤:

根据属性数据建立查询索引;

根据所述查询索引建立倒排索引表,以使搜索引擎并行查询存储于若干个节点内的若干个数据组。

[0034] 在本实施方式中,采用的图数据库存储引擎支持高效的节点和边查询,方便用户根据条件查询特定的网络资产,具体来说,本图数据库存储引擎使用基于Lucene的全文搜索索引技术来为节点和边的属性建立索引。这种索引技术可以将节点和边的属性值进行分词,并建立一个倒排索引表,用于记录每个属性值所属的节点和边的信息。通过使用这种倒排索引表,查询时可以根据用户提供的查询条件快速地定位到包含这些条件的节点和边。同时本专利还使用分片存储和并行查询的方式,将数据分成多个分片存储在不同的节点上,并行查询多个分片,从而提高查询效率。也采用预处理和缓存的方式,将常用的查询结果预先计算并缓存,以便快速响应用户的查询请求。

[0035] 本申请实施例还公开一种网络资产数据存储和查询装置,包括:

存储模块,用于将包括若干个第二文档的第一文档存储至非关系型数据库中;

查询模块,用于通过搜索引擎并行查询存储于若干个节点内的若干个数据组。

[0036] 本申请实施例的网络资产数据存储和查询装置能够实现上述网络资产数据存储

方法和查询方法的任一种方法,且网络资产数据存储和查询装置中各个模块的具体工作过程可参考上述方法实施例中的对应过程。

[0037] 在本申请所提供的几个实施例中,应该理解到,所提供的方法和系统,可以通过其它的方式实现。例如,以上所描述的系统实施例仅仅是示意性的;例如,某个模块的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个模块可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。

[0038] 本申请实施例还公开一种计算机设备。

[0039] 计算机设备,包括存储器、处理器以及存储在存储器上并可在处理器上运行的计算机程序,处理器执行计算机程序时实现如上述的xx方法。

[0040] 本申请实施例还公开一种计算机可读存储介质。

[0041] 计算机可读存储介质,存储有能够被处理器加载并执行如上述的xx方法中任一种方法的计算机程序。

[0042] 其中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用;计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0043] 实施例:

若某个用户或者企业拥有大量的网络资产,需要对这些资产进行管理和监控,以保障网络的安全性和稳定性。传统的管理方式需要手动录入和管理,效率低下且容易出错。而采用本申请的技术方案可以实现自动化的网络资产管理和监控,提高管理效率和精度,具体步骤如下:

S101. 数据采集:首先,使用本申请的技术方案采集企业所有的网络资产数据,包括域名、IP地址、设备信息等,并将这些数据存储到图数据库中。可以采用预处理和缓存的方式,将常用的查询结果预先计算并缓存,以便快速响应用户的查询请求。

[0044] S102. 数据建模:将采集到的网络资产数据抽象为节点和边的形式,使用图数据库将这些节点和边存储为图形结构。节点表示网络资产,边表示不同网络资产之间的关系,如域名到子域名、子域名到 IP 地址等。这种图形化建模方式可以更好地支持网络资产数据的查询和关系分析。

[0045] S103. 数据管理:同时将不需要存储到图数据库中的一些少量信息,如用户信息等,存储到关系型数据库中,并与图数据库中的节点和边建立关联关系,从而实现全面的数据管理和查询功能。

[0046] S104. 数据查询:用户可以更具需求快速查询出想要的网络资产数据,比如某个域名相关的子域名,子域名下的IP等一系列的关系数据。

[0047] S105. 数据安全:采用数据加密和身份验证等安全策略,保障数据的安全性。同时,使用多数据库同步功能,将数据同步到多个数据库中,并进行备份和恢复操作,从而提高数据的可靠性和可用性。

[0048] 通过以上处理步骤,可以实现自动化的网络资产管理和监控,提高管理效率和精度。

[0049] 由上可见,本申请提供了一种网络资产数据存储和查询方法、装置及存储介质一

方面采用图数据库作为核心存储引擎,将网络资产数据抽象为节点和边,并以图形结构存储,从而能够更快速地进行查询和关系分析,更好的支持网络资产数据的图形化建模和查询;另一方面,使用了基于Lucene的全文搜索索引技术来为节点和边的属性建立索引,方便用户查询时可以根据用户提供的查询条件快速地定位到包含这些条件的节点和边。采用了分片存储和并行查询的方式,对大规模网络资产数据进行存储和查询,可以提高查询效率。同时,采用预处理和缓存的方式,将常用的查询结果预先计算并缓存,以便快速响应用户的查询请求。提高存储和查询的效率;除了采用图数据库存储节点和边之外,还结合了关系型数据库的存储能力,用于存储一些不需要存储到图数据库中的一些少量信息。这些信息可以采用关系型数据库的方式进行存储和查询,与图数据库中的节点和边建立关联关系,从而实现全面的数据管理和查询功能;而以上的图数据库和关系型数据库采用多数据库同步功能,可以将数据同步到多个数据库中,并进行备份和恢复操作,从而提高数据的可靠性和可用性。采用数据加密和身份验证等安全策略,保障数据的安全性,提高数据的安全性和可用性。

[0050] 需要说明的是,在上述实施例中,对各个实施例的描述各有侧重,某个实施例中沒有详述的部分,可以参见其他实施例的相关描述。

[0051] 以上均为本申请的较佳实施例,并非依此限制本申请的保护范围,本说明书(包括摘要和附图)中公开的任一特征,除非特别叙述,均可被其他等效或者具有类似目的的替代特征加以替换。即,除非特别叙述,每个特征只是一系列等效或类似特征中的一个例子而已。

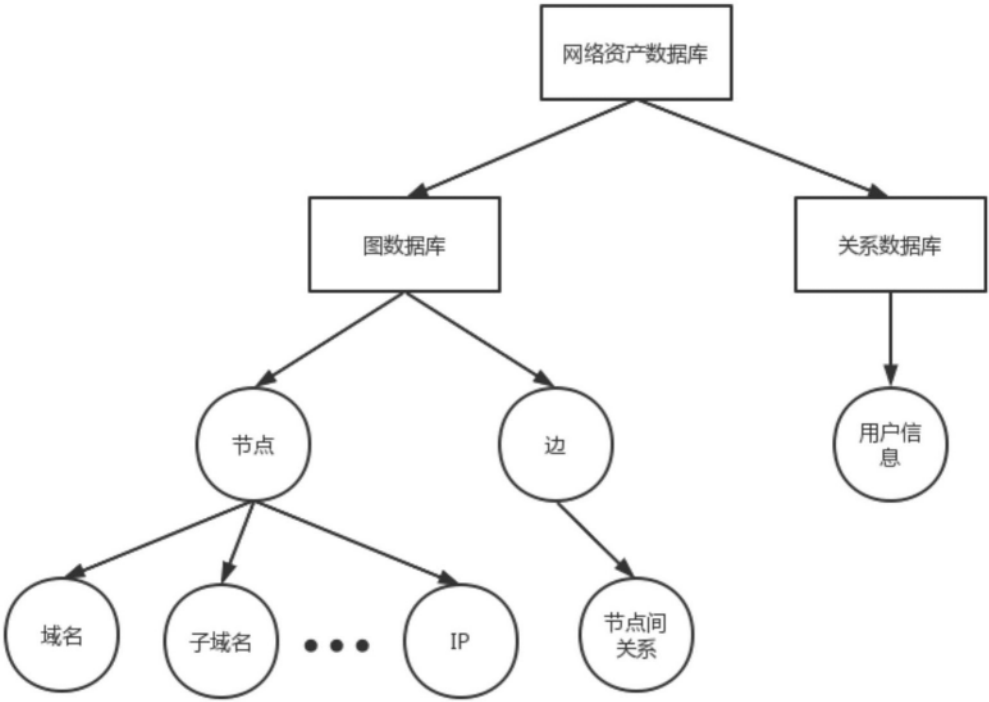


图1

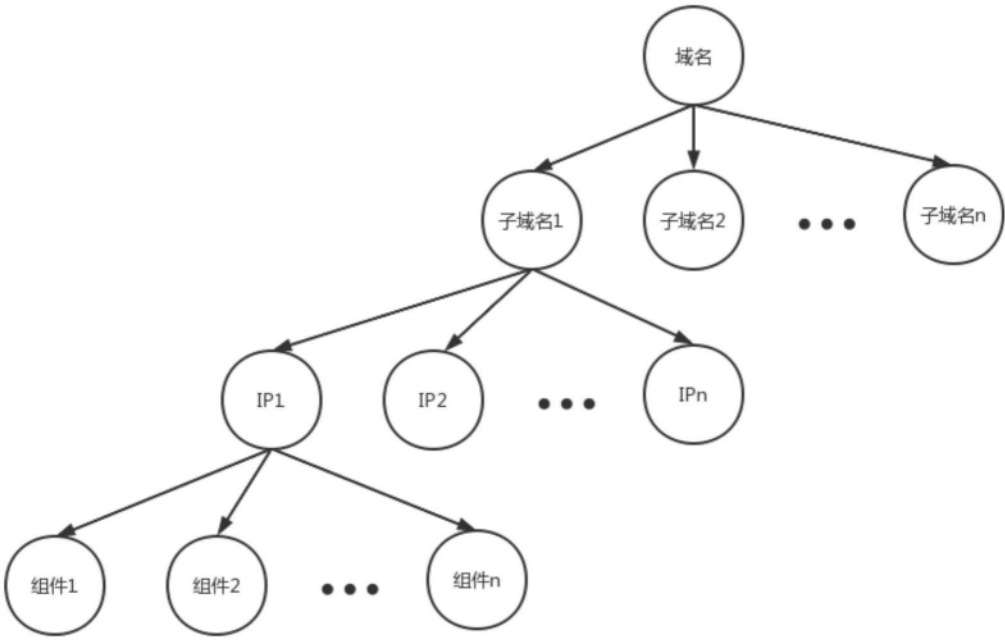


图2

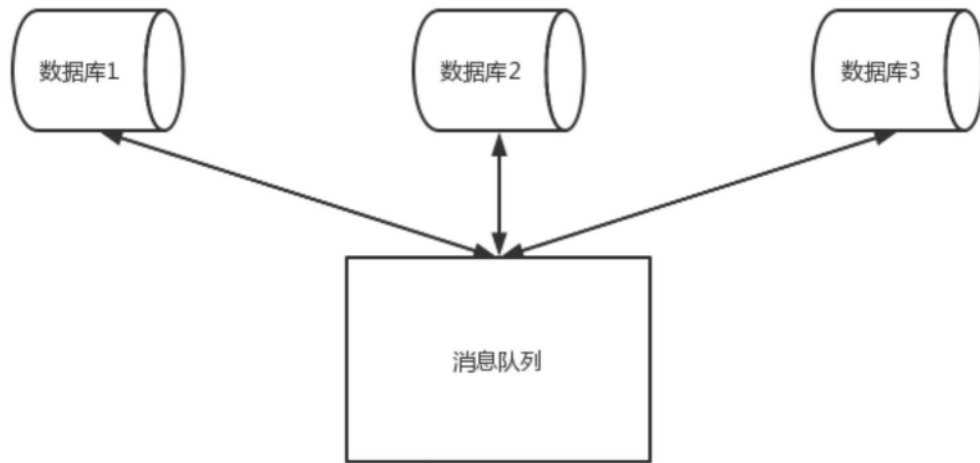


图3