(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 110308983 B
(45)授权公告日 2022.04.05

(21)申请号 201910316501.3
(22)申请日 2019.04.19
(65)同一申请的已公布的文献号
申请公布号 CN 110308983 A
(43)申请公布日 2019.10.08
(73)专利权人 中国工商银行股份有限公司
地址 100140 北京市西城区复兴门内大街55号
(72)发明人 沈赟 袁一 张学舟 翁晓俊
(74)专利代理机构 北京三友知识产权代理有限公司 11127
代理人 王涛 任默闻
(51)Int.Cl.
G06F 9/50(2006.01)

(56)对比文件
CN 108076092 A,2018.05.25
CN 108076092 A,2018.05.25
CN 102025630 A,2011.04.20
CN 106686102 A,2017.05.17
CN 102098354 A,2011.06.15

审查员 卢双龙

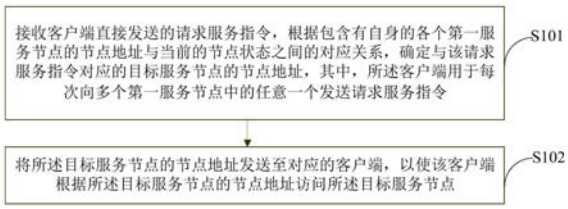
权利要求书3页 说明书17页 附图7页

(54)发明名称

资源负载均衡方法及系统、服务节点和客户端

(57)摘要

本申请提供一种资源负载均衡方法及系统、服务节点和客户端，方法包括：接收客户端直接发送的请求服务指令，根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址，其中，所述客户端用于每次向多个第一服务节点中的任意一个发送请求服务指令；将所述目标服务节点的节点地址发送至对应的客户端，以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点。本申请能够提高负载均衡的稳定性和准确性并降低运维成本。



CN 110308983 B

1.一种资源负载均衡方法,其特征在于,包括:

接收客户端直接发送的请求服务指令,根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系,确定与该请求服务指令对应的目标服务节点的节点地址,其中,所述客户端用于每次向多个第一服务节点中的任意一个发送请求服务指令;

将所述目标服务节点的节点地址发送至对应的客户端,以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点;

在所述确定与该请求服务指令对应的目标服务节点的节点地址之前,还包括:

获取各个第二服务节点的节点地址与当前的节点状态之间的对应关系;

相对应的,所述确定与该请求服务指令对应的目标服务节点的节点地址,包括:

分别根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系,以及,各个所述第二服务节点的节点地址与当前的节点状态之间的对应关系,确定与该请求服务指令对应的目标服务节点的节点地址;

其中,所述目标服务节点为各个所述第一服务节点中的一个或各个所述第二服务节点中的一个;

所述第二服务节点仅能为客户端提供服务,不能接收客户端发送的请求服务指令以及为该客户端分配服务节点。

2.根据权利要求1所述的资源负载均衡方法,其特征在于,所述分别根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系,以及,各个所述第二服务节点的节点地址与当前的节点状态之间的对应关系,确定与该请求服务指令对应的目标服务节点的节点地址,包括:

在本地存储的服务节点列表中查找与所述服务指令对应的目标服务节点的节点地址,所述服务节点列表用于存储各个第一服务节点的节点地址与当前的节点状态之间的对应关系,以及各个第二服务节点的节点地址与当前的节点状态之间的对应关系。

3.根据权利要求2所述的资源负载均衡方法,其特征在于,还包括:

每隔第一预设时间向资源管理装置发送心跳信息,以使所述资源管理装置根据接收自多个第一服务节点和第二服务节点的心跳信息更新服务节点列表,并将更新后的服务节点列表发送至发出心跳信息的多个第一服务节点;

接收所述资源管理装置发送的更新后的服务节点列表,并将原存储的服务节点列表替换为该更新后的服务节点列表。

4.一种资源负载均衡方法,其特征在于,包括:

向多个第一服务节点中的任意一个发送请求服务指令,以使接收到该请求服务指令的第一服务节点根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系,确定与该请求服务指令对应的目标服务节点的节点地址;

若在第二预设时间内接收到对应的第一服务节点发送的所述目标服务节点的节点地址,则根据所述目标服务节点的节点地址访问所述目标服务节点;

所述目标服务节点为各个所述第一服务节点中的一个或各个第二服务节点中的一个;

相对应的,确定与该请求服务指令对应的目标服务节点的节点地址,包括:

分别根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系,以及,各个所述第二服务节点的节点地址与当前的节点状态之间的对应关系,确定与

该请求服务指令对应的目标服务节点的节点地址；

其中，每个所述第一服务节点中均存储有各个第一服务节点的节点地址与当前的节点状态之间的对应关系，以及，各个所述第二服务节点的节点地址与当前的节点状态之间的对应关系；

所述第二服务节点仅能为客户端提供服务，不能接收客户端发送的请求服务指令以及为该客户端分配服务节点。

5.根据权利要求4所述的资源负载均衡方法，其特征在于，所述向多个第一服务节点中的任意一个发送请求服务指令，包括：

自存储在本地的第一服务节点列表中选取任意一个未有故障标记的第一服务节点，并向该未有故障标记的第一服务节点发送请求服务指令；

相对应的，若在第二预设时间内未接收到对应的第一服务节点发送的所述目标服务节点的节点地址，则在所述第一服务节点列表对该第一服务节点进行故障标记。

6.根据权利要求5所述的资源负载均衡方法，其特征在于，所述第一服务节点列表包括：

客户端所在区域的第一服务节点第一子列表，以及，非客户端所在区域的第一服务节点第二子列表；

相对应的，自存储在本地的第一服务节点第一子列表中选取任意一个未有故障标记的第一服务节点，并向该未有故障标记的第一服务节点发送请求服务指令；在第二预设时间内未接收到对应的第一服务节点发送的所述目标服务节点的节点地址，则在所述第一服务节点列表对该第一服务节点进行故障标记；

若所述第一服务节点第一子列表中的所有第一服务节点均被标记为故障，则从所述第一服务节点第二子列表中选取任意一个未有故障标记的第一服务节点，并向该未有故障标记的第一服务节点发送请求服务指令。

7.一种第一服务节点，其特征在于，包括：

服务节点查找模块，用于接收客户端直接发送的请求服务指令，根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址，其中，所述客户端用于每次向多个第一服务节点中的任意一个发送请求服务指令；

服务节点发送模块，用于将所述目标服务节点的节点地址发送至对应的客户端，以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点；

所述第一服务节点还包括：获取模块，用于获取各个第二服务节点的节点地址与当前的节点状态之间的对应关系；

相对应的，所述服务节点查找模块具体用于：

分别根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，以及，各个所述第二服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址；

其中，所述目标服务节点为各个所述第一服务节点中的一个或各个所述第二服务节点中的一个；

所述第二服务节点仅能为客户端提供服务，不能接收客户端发送的请求服务指令以及

为该客户端分配服务节点。

8.一种客户端，其特征在于，包括：

指令发送模块，用于向多个第一服务节点中的任意一个发送请求服务指令，以使接收到该请求服务指令的第一服务节点根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址；

访问模块，用于若在第二预设时间内接收到对应的第一服务节点发送的所述目标服务节点的节点地址，则根据所述目标服务节点的节点地址访问所述目标服务节点；

所述目标服务节点为各个所述第一服务节点中的一个或各个第二服务节点中的一个；

相对应的，所述指令发送模块具体用于：

分别根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，以及，各个所述第二服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址；

其中，每个所述第一服务节点中均存储有各个第一服务节点的节点地址与当前的节点状态之间的对应关系，以及，各个所述第二服务节点的节点地址与当前的节点状态之间的对应关系；

所述第二服务节点仅能为客户端提供服务，不能接收客户端发送的请求服务指令以及为该客户端分配服务节点。

9.一种资源负载均衡系统，其特征在于，包括：多个第一服务节点和资源管理装置；

所述第一服务节点用于实现权利要求1至3任一项所述的资源负载均衡方法；

各个所述第一服务节点均与所述客户端通信连接，且该客户端用于实现权利要求4至6任一项所述的资源负载均衡方法；

各个所述第一服务节点分别周期性向资源管理装置发送心跳信息，以使所述资源管理装置根据所述心跳信息生成所述服务节点列表并对所述服务节点列表进行周期性更新，以及将更新后的服务节点列表发送至多个第一服务节点。

10.根据权利要求9所述的资源负载均衡系统，其特征在于，还包括：多个第二服务节点；

各个所述第二服务节点均与所述客户端通信连接；

其中，各个所述第一服务节点和各个所述第二服务节点分别周期性向资源管理装置发送心跳信息，以使所述资源管理装置根据所述心跳信息生成所述服务节点列表并对所述服务节点列表进行周期性更新，以及将更新后的服务节点列表发送至多个第一服务节点。

11.一种电子设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，其特征在于，所述处理器执行所述程序时实现权利要求1至3任一项所述的资源负载均衡方法的步骤。

12.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，该计算机程序被处理器执行时实现权利要求1至3任一项所述的资源负载均衡方法的步骤。

资源负载均衡方法及系统、服务节点和客户端

技术领域

[0001] 本发明涉及计算机应用技术领域，具体涉及一种资源负载均衡方法及系统、服务节点和客户端。

背景技术

[0002] 当前，在企业内部广泛使用服务器、数据存储、网络等IT应用，伴随业务发展IT应用的内部结构也趋于复杂。业务发展到一定程度，交易量和数据处理量的增长使得服务资源遭遇瓶颈，而且还要随时应对用户访问量突然暴增的影响，例如，企业的电子商务平台因商业促销活动导致后台服务器在响应用户请求和交易订单并发处理等多个环节都会面临超负荷运转。与此同时，企业开始重新定义业务发展过程中的共性需求，例如，延长IT基础设施使用年限，提高服务器设备综合使用率以及节约软硬件设备运维、更替、扩展成本等。而负载均衡技术为企业从软件方面提供了一种可行的方案来解决上述问题。

[0003] 负载均衡是利用多台服务器搭建起一个松耦合的服务器集群，通过负载均衡设备实现服务器集群以统一的整体共同对外提供服务，达到缓解服务器的单点损耗，以此延长服务器的服务年限。

[0004] 但是，采用负载均衡设备会导致成本的增加，而且负载均衡设备是单节点配置，由单一的节点来接收和分配客户端的请求，这种单点配置方式容易导致当单节点故障时整个服务系统陷入瘫痪的问题。

发明内容

[0005] 针对现有技术中的问题，本发明提供一种资源负载均衡方法及系统、服务节点和客户端，能够提高负载均衡的稳定性和准确性并降低运维成本。

[0006] 为解决上述技术问题，本发明提供以下技术方案：

[0007] 第一方面，本发明提供一种资源负载均衡方法，包括：

[0008] 接收客户端直接发送的请求服务指令，根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址，其中，所述客户端用于每次向多个第一服务节点中的任意一个发送请求服务指令；

[0009] 将所述目标服务节点的节点地址发送至对应的客户端，以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点。

[0010] 进一步地，在所述确定与该请求服务指令对应的目标服务节点的节点地址之前，还包括：

[0011] 获取各个第二服务节点的节点地址与当前的节点状态之间的对应关系；

[0012] 相对应的，所述确定与该请求服务指令对应的目标服务节点的节点地址，包括：

[0013] 分别根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，以及，各个所述第二服务节点的节点地址与当前的节点状态之间的对应关系，确

定与该请求服务指令对应的目标服务节点的节点地址；

[0014] 其中，所述目标服务节点为各个所述第一服务节点中的一个或各个所述第二服务节点中的一个。

[0015] 进一步地，所述分别根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，以及，各个所述第二服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址，包括：

[0016] 在本地存储的服务节点列表中查找与所述服务指令对应的目标服务节点的节点地址，所述服务节点列表用于存储各个第一服务节点的节点地址与当前的节点状态之间的对应关系，以及各个第二服务节点的节点地址与当前的节点状态之间的对应关系。

[0017] 进一步地，还包括：

[0018] 每隔第一预设时间向资源管理装置发送心跳信息，以使所述资源管理装置根据接收自多个第一服务节点和第二服务节点的心跳信息更新服务节点列表，并将更新后的服务节点列表发送至发出心跳信息的多个第一服务节点；

[0019] 接收所述资源管理装置发送的更新后的服务节点列表，并将原存储的服务节点列表替换为该更新后的服务节点列表。

[0020] 第二方面，本发明提供另一种资源负载均衡方法，包括：

[0021] 向多个第一服务节点中的任意一个发送请求服务指令，以使接收到该请求服务指令的第一服务节点根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址；

[0022] 若在第二预设时间内接收到对应的第一服务节点发送的所述目标服务节点的节点地址，则根据所述目标服务节点的节点地址访问所述目标服务节点。

[0023] 进一步地，所述目标服务节点为各个所述第一服务节点中的一个或各个所述第二服务节点中的一个；

[0024] 相对应的，确定与该请求服务指令对应的目标服务节点的节点地址，包括：

[0025] 分别根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，以及，各个所述第二服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址；

[0026] 其中，每个所述第一服务节点中均存储有各个第一服务节点的节点地址与当前的节点状态之间的对应关系，以及，各个所述第二服务节点的节点地址与当前的节点状态之间的对应关系。

[0027] 进一步地，所述向多个第一服务节点中的任意一个发送请求服务指令，包括：

[0028] 自存储在本地的第一服务节点列表中选取任意一个未有故障标记的第一服务节点，并向该未有故障标记的第一服务节点发送请求服务指令；

[0029] 相对应的，若在第二预设时间内未接收到对应的第一服务节点发送的所述目标服务节点的节点地址，则在所述第一服务节点列表对该第一服务节点进行故障标记。

[0030] 进一步地，所述第一服务节点列表包括：

[0031] 客户端所在区域的第一服务节点第一子列表，以及，非客户端所在区域的第一服务节点第二子列表；

[0032] 相对应的，自存储在本地的第一服务节点第一子列表中选取任意一个未有故障标

记的第一服务节点，并向该未有故障标记的第一服务节点发送请求服务指令；在第二预设时间内未接收到对应的第一服务节点发送的所述目标服务节点的节点地址，则在所述第一服务节点列表对该第一服务节点进行故障标记；

[0033] 若所述第一服务节点第一子列表中的所有第一服务节点均被标记为故障，则从所述第一服务节点第二子列表中选取任意一个未有故障标记的第一服务节点，并向该未有故障标记的第一服务节点发送请求服务指令。

[0034] 第三方面，本发明提供另一种第一服务节点，包括：

[0035] 服务节点查找模块，用于接收客户端直接发送的请求服务指令，根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址，其中，所述客户端用于每次向多个第一服务节点中的任意一个发送请求服务指令；

[0036] 服务节点发送模块，用于将所述目标服务节点的节点地址发送至对应的客户端，以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点。

[0037] 第四方面，本发明提供另一种客户端，包括：

[0038] 指令发送模块，用于向多个第一服务节点中的任意一个发送请求服务指令，以使接收到该请求服务指令的第一服务节点根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址；

[0039] 访问模块，用于若在第二预设时间内接收到对应的第一服务节点发送的所述目标服务节点的节点地址，则根据所述目标服务节点的节点地址访问所述目标服务节点。

[0040] 第五方面，本发明提供另一种资源负载均衡系统，包括：多个第一服务节点和资源管理装置；

[0041] 所述第一服务节点用于实现上述第一方面所述的资源负载均衡方法；

[0042] 各个所述第一服务节点均与所述客户端通信连接，且该客户端用于实现上述第二方面所述的资源负载均衡方法；

[0043] 各个所述第一服务节点分别周期性向资源管理装置发送心跳信息，以使所述资源管理装置根据所述心跳信息生成所述服务节点列表并对所述服务节点列表进行周期性更新，以及将更新后的服务节点列表发送至多个第一服务节点。

[0044] 进一步地，还包括：多个第二服务节点；

[0045] 各个所述第二服务节点均与所述客户端通信连接；

[0046] 其中，各个所述第一服务节点和各个所述第二服务节点分别周期性向资源管理装置发送心跳信息，以使所述资源管理装置根据所述心跳信息生成所述服务节点列表并对所述服务节点列表进行周期性更新，以及将更新后的服务节点列表发送至多个第一服务节点。

[0047] 第六方面，本发明提供一种电子设备，包括存储器、处理器及存储在存储器上并可在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上述第一方面所述的资源负载均衡方法的步骤。

[0048] 第七方面，本发明提供一种计算机可读存储介质，其上存储有计算机程序，该计算机程序被处理器执行时实现上述第一方面所述的资源负载均衡方法的步骤。

[0049] 由上述技术方案可知，本发明提供一种资源负载均衡方法及系统、服务节点和客户端，通过接收客户端直接发送的请求服务指令，根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址，其中，所述客户端用于每次向多个第一服务节点中的任意一个发送请求服务指令；实现由多个第一服务节点接收客户端的请求，避免了单节点故障的负面影响，提高负载均衡的稳定性和准确性；将所述目标服务节点的节点地址发送至对应的客户端，以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点，实现由服务节点分配客户端的请求，无需另设置负载均衡设备降低运维成本。

附图说明

[0050] 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。

[0051] 图1为现有技术中资源负载均衡方法的流程图；

[0052] 图2为本发明实施例提供的第一种资源负载均衡方法的流程示意图；

[0053] 图3为本发明实施例提供的第二种资源负载均衡方法的流程示意图；

[0054] 图4为本发明实施例提供的第三种资源负载均衡方法的流程图；

[0055] 图5为本发明实施例提供的为客户端提供资源负载均衡服务的流程图；

[0056] 图6为本发明实施例提供的为客户端提供资源负载均衡服务中获取服务节点列表的流程图；

[0057] 图7为本发明实施例提供的资源负载均衡系统的结构示意图；

[0058] 图8为本发明实施例提供的资源负载均衡系统中第一服务节点的结构示意图；

[0059] 图9为本发明实施例提供的资源负载均衡系统跨集群服务的结构示意图；

[0060] 图10为本发明实施例提供的第一服务节点的结构示意图；

[0061] 图11为本发明实施例提供的客户端的结构示意图；

[0062] 图12为本发明实施例中的电子设备的结构示意图。

具体实施方式

[0063] 为使本发明实施例的目的、技术方案和优点更加清楚，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、完整的描述，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做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。

[0064] 图1为现有的资源负载均衡方法的流程图，在现有的资源负载均衡方法中，由单节点(转发节点)接收多个客户端的服务请求，并在多个服务节点中选取一个或多个服务节点为发出服务请求的客户端提供服务，在这过程中，若是转发节点出现故障，会导致整个服务系统陷入瘫痪的问题，而且，采用转发节点来接收和分配客户端的请求会导致成本的增加。考虑到现有的资源负载均衡方法中存在的问题，本发明提供一种资源负载均衡方法及系统、服务节点和客户端，通过接收客户端直接发送的请求服务指令，根据包含有自身的各个

第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址，其中，所述客户端用于每次向多个第一服务节点中的任意一个发送请求服务指令；实现由多个第一服务节点接收客户端的请求，避免了单节点故障的负面影响，提高负载均衡的稳定性和准确性；将所述目标服务节点的节点地址发送至对应的客户端，以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点，实现由服务节点分配客户端的请求，无需另设置负载均衡设备降低运维成本。

[0065] 为了提高负载均衡的稳定性和准确性并降低运维成本，本发明提供一种资源负载均衡方法的实施例，参见图2，本实施例所提供的资源负载均衡方法适用于服务端，用于接收客端发送的请求服务指令并反馈响应的结果，具体包含如下内容：

[0066] S101：接收客户端直接发送的请求服务指令，根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址，其中，所述客户端用于每次向多个第一服务节点中的任意一个发送请求服务指令；

[0067] 在本步骤中，第一服务节点直接接收客户端发送的请求服务指令，并且为该客户端分配一个目标服务节点的节点地址，实现由服务节点直接进行接收和分配客户端的请求服务指令。而且客户端可以多次向多个第一服务节点发送请求服务指令，能够避免采用单一转发节点会导致整个服务系统陷入瘫痪的问题。

[0068] 需要说明的是，第一服务节点能够实现直接接收客户端发送的请求服务指令，并为该客户端分配服务节点，实现为客户端提供服务。

[0069] S102：将所述目标服务节点的节点地址发送至对应的客户端，以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点。

[0070] 在本步骤中，第一服务节点能够将分配的目标服务节点的节点地址发送至对应的客户端，以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点，实现为客户端提供响应的服务。

[0071] 从上述描述可知，本发明实施例提供一种资源负载均衡方法，通过多个第一服务节点接收客户端的请求并由服务节点为该客户端分配一个服务节点，避免了单节点故障的负面影响，提高负载均衡的稳定性和准确性，通过第一服务节点直接接收客户端发送的请求服务指令，无需另设置负载均衡设备降低运维成本。

[0072] 在保证资源负载均衡的稳定性和准确性的情况下，减少处理的负载程度和设备损耗，在本发明的一实施例中，参见图3，在所述确定与该请求服务指令对应的目标服务节点的节点地址之前，还包括：

[0073] S100：获取各个第二服务节点的节点地址与当前的节点状态之间的对应关系；

[0074] 在本步骤中，第一服务节点还能够获取各个第二服务节点的节点地址与当前的节点状态之间的对应关系。因此，第一服务节点在确定与该请求服务指令对应的目标服务节点的节点地址时，分别根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，以及，各个所述第二服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址。在这一过程中所确定的目标服务节点为各个所述第一服务节点中的一个或各个所述第二服务节点中的一个。由于第一服务节点的节点地址与当前的节点状态之间的对应关系中包含有接收请求服务指令的

第一服务节点，因此，接收请求服务指令的第一服务节点也存在为客户端提供服务的可行性。

[0075] 需要说明的是，第二服务节点不同于第一服务节点，该第二服务节点仅能为客户端提供服务，不能接收客户端发送的请求服务指令以及为该客户端分配服务节点。通过第一服务节点接收和分配客户端的请求服务指令，第一服务节点和第二服务节点共同为客户端提供服务，能够在保证资源负载均衡的稳定性和准确性的情况下，减少处理的负载程度和设备损耗，实现节约成本。

[0076] 进一步的，第一服务节点在确定与该请求服务指令对应的目标服务节点的节点地址时，在本地存储的服务节点列表中查找与所述服务指令对应的目标服务节点的节点地址，其中，所述服务节点列表用于存储各个第一服务节点的节点地址与当前的节点状态之间的对应关系，以及各个第二服务节点的节点地址与当前的节点状态之间的对应关系。

[0077] 第一服务节点通过查找本地存储的服务节点列表，避免与外部进行通信，能够提高第一服务节点的处理效率，实现提高均衡过程的整体效率。

[0078] 为提高第一服务节点的处理效率，实现提高均衡过程的整体效率，第一服务节点存储的服务节点列表，具体的实现过程如下：

[0079] 第一服务节点每隔第一预设时间向资源管理装置发送心跳信息，以使所述资源管理装置根据接收自多个第一服务节点的心跳信息更新服务节点列表，并将更新后的服务节点列表发送至发出心跳信息的多个第一服务节点；

[0080] 需要说明的是，如果设置有第二服务节点，则第二服务节点每隔第一预设时间向资源管理装置发送心跳信息，以使所述资源管理装置根据接收自多个第一服务节点和第二服务节点的心跳信息更新服务节点列表，并将更新后的服务节点列表发送至发出心跳信息的多个第一服务节点；

[0081] 为了提高存储器的利用率并降低存储器的使用空间，第一服务节点在接收所述资源管理装置发送的更新后的服务节点列表，并将原存储的服务节点列表替换为该更新后的服务节点列表。

[0082] 为了提高负载均衡的稳定性和准确性并降低运维成本，本发明提供另一种资源负载均衡方法的实施例，参见图4，所述资源负载均衡方法适用于客户端，用于向服务端发送请求服务指令并接收服务端反馈响应的结果，具体包含有如下内容：

[0083] S201：向多个第一服务节点中的任意一个发送请求服务指令，以使接收到该请求服务指令的第一服务节点根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址；

[0084] 在本步骤中，客户端内部存储有第一服务节点列表，该第一服务节点列表中存储有多个第一服务节点，客户端向多个第一服务节点中的任意一个发送请求服务指令，以使接收到该请求服务指令的第一服务节点根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址；

[0085] 客户端可以多次向多个第一服务节点发送请求服务指令，避免采用单一转发节点会导致整个服务系统陷入瘫痪的问题。

[0086] S202：若在第二预设时间内接收到对应的第一服务节点发送的所述目标服务节点

的节点地址，则根据所述目标服务节点的节点地址访问所述目标服务节点。

[0087] 在本步骤中，客户端直接向第一服务节点发送请求服务指令，并接收第一服务节点发送的目标服务节点的节点地址，客户端根据所述目标服务节点的节点地址访问所述目标服务节点，提高负载均衡的稳定性和准确性。

[0088] 从上述描述可知，本发明实施例提供一种资源负载均衡方法，由客户端向多个第一服务节点发送请求服务指令，并由接收请求服务指令的服务节点为该客户端分配一个目标服务节点，避免了单节点故障的负面影响，提高负载均衡的稳定性和准确性，而且客户端直接向服务节点发送请求服务指令，无需另设置负载均衡设备降低运维成本。

[0089] 在上述实施例中，客户端接收的目标服务节点的节点地址，该目标服务节点为各个第一服务节点中的一个或各个第二服务节点中的一个；

[0090] 需要说明的是，第一服务节点能够实现直接接收客户端发送的请求服务指令，并为该客户端分配服务节点，实现为客户端提供服务。第二服务节点不同于第一服务节点，该第二服务节点仅能为客户端提供服务，不能接收客户端发送的请求服务指令以及为该客户端分配服务节点。在保证资源负载均衡的稳定性和准确性的情况下，服务端为了减少处理的负载程度和设备损耗，实现节约成本，服务端提供的服务节点可以是多个第一服务节点和多个第二服务节点的组合。在该组合中，由第一服务节点接收和分配客户端的请求服务指令，第一服务节点和第二服务节点共同为客户端提供服务，能够在保证资源负载均衡的稳定性和准确性的情况下，减少处理的负载程度和设备损耗，避免全部使用第一服务节点而实现节约成本。

[0091] 相对应的，第一服务节点在确定与该请求服务指令对应的目标服务节点的节点地址时：分别根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，以及，各个所述第二服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址；

[0092] 其中，每个所述第一服务节点中均存储有各个第一服务节点的节点地址与当前的节点状态之间的对应关系，以及，各个所述第二服务节点的节点地址与当前的节点状态之间的对应关系。

[0093] 在上述实施例中，客户端若在第二预设时间内未接收到对应的第一服务节点发送的所述目标服务节点的节点地址，则在所述第一服务节点列表对该第一服务节点进行故障标记。客户端在第三预设时间内不向被标记为故障的第一服务节点发送请求服务指令。

[0094] 进一步的，为提高客户端的通信效率，客户端自存储在本地的第一服务节点列表中选取任意一个未有故障标记的第一服务节点，并向该未有故障标记的第一服务节点发送请求服务指令。

[0095] 若在第二预设时间内接收到对应的第一服务节点发送的所述目标服务节点的节点地址，则根据所述目标服务节点的节点地址访问所述目标服务节点；

[0096] 若在第二预设时间内未接收到对应的第一服务节点发送的所述目标服务节点的节点地址，则在所述第一服务节点列表对该第一服务节点进行故障标记。

[0097] 进一步的，为提高客户端具备灾备性质的跨区域请求服务的稳定性，客户端存储在本地的第一服务节点列表包括不同区域的第一服务节点子列表，其中将不同区域划分为：客户端所在区域和非客户端所在区域，则客户端内存储的第一服务节点列表包括：客户

端所在区域的第一服务节点第一子列表，以及，非客户端所在区域的第一服务节点第二子列表。

[0098] 客户端向第一服务节点第一子列表中选取任意一个未有故障标记的第一服务节点，并向该未有故障标记的第一服务节点发送请求服务指令；在第二预设时间内未接收到对应的第一服务节点发送的所述目标服务节点的节点地址，则在所述第一服务节点第一子列表中对该第一服务节点进行故障标记；

[0099] 若所述第一服务节点第一子列表中的所有第一服务节点均被标记为故障，客户端则在所述第一服务节点第二子列表中选取任意一个未有故障标记的第一服务节点，并向该未有故障标记的第一服务节点发送请求服务指令。

[0100] 从上述描述可知，客户端所在区域可以正常提供服务的前提下，该区域的客户端的服务在本区域内部处理，这样的工作效率最高；当客户端所在区域发生集群瘫痪的异常时，客户端通过另一个区域的集群来进行处理，保证客户端请求得到及时处理。

[0101] 基于上述内容，本发明还给出一种结合服务节点和客户端的资源负载均衡方法的具体应用实施例，参见图5，具体包含有如下内容：

[0102] S501：客户端随机选择一个第一服务节点，向其发送Request_Address的指令后，进入等待状态。

[0103] 每个客户端内部维护有第一服务节点的地址列表，列表内容包括第一服务节点的全量信息，第一服务节点负责客户端请求的调度。若客户端需要向第一服务节点请求服务，则随机从地址列表中选定一个第一服务节点，向选定的第一服务节点发送Request_Address指令。

[0104] S502：客户端判断选定的第一服务节点是否在预期答复时间内向其提供目标服务节点的节点地址，如是执行S503，如否执行S501重新连接并随机从地址列表中再选定一个第一服务节点。

[0105] 第一服务节点将收到的Request_Address指令封装成一个请求项，并将该请求项加入到其服务请求队列的尾部。第一服务节点从服务请求队列首部获取一个请求项，从本地的地址列表中随机选择一个地址反馈给请求项中的客户端。

[0106] 其中，当第一服务节点无法及时反馈客户端请求时，在第一服务节点内部维护了一个数据队列，来缓存客户端Request_Address指令信息。

[0107] 具体地，每个第一服务节点内部维护着一个服务节点列表。该服务节点列表的信息定期更新，由第一服务节点每间隔一段指定时间向资源管理装置获取而得。第一服务节点每次从服务请求队列中获取一个请求项后，再从本地的服务节点列表中随机选择一个目标服务节点的节点地址，选中的目标服务节点可能是第一服务节点或者第二服务节点反馈给客户端。

[0108] 为了避免客户端因向发生故障的第一服务节点发送请求长时间未回应而阻塞，设置预期答复时间。如第一服务节点未能及时答复，则认为当前连接的第一服务节点可能发生故障，需要重新选择节点进行连接。同时，客户端会将该故障地址从第一服务节点列表中去除，将该故障地址加入至故障地址列表。在下一次选择新的节点时，可排除该故障地址选项。对于故障地址列表，客户端每隔一定的时间遍历且测试每个故障地址是否恢复，若恢复则重新将故障地址加入至服务节点列表。

[0109] S503:客户端接收到第一服务节点提供的目标服务节点的节点地址,随机生成一段服务时长,将该时长信息置入请求指令中,发送至目标服务节点。

[0110] 由于客户端需要服务的时间是未知的,客户端可能需要一次或者多次请求才能完成所需业务处理。客户端在每次请求服务时,都需先进行一次寻求地址的操作,随后进入请求服务的状态。将每次服务连接的时长设为不定长,如此可以错开不同客户端寻址以及请求服务连接的时刻。因此,客户端的每次请求服务时长是不定长的,由客户端在服务连接前取指定时间范围内的随机设置。

[0111] S504:目标服务节点将收到的服务请求封装成一个作业,放入作业队列尾部。

[0112] 目标服务节点在接收到客户端的服务请求后,将服务请求封装成一个作业,若无法及时处理该作业,可暂存于作业队列,按照“先进先出”规则先处理较早作业请求。目标服务节点记录其作业队列的长度来用于实时负载评价。

[0113] S505:目标服务节点从作业队列取出最早的作业,解析并执行作业内容。

[0114] 目标服务节点从作业队列取出最早的作业并从作业中解析出客户端地址、所需服务和参数信息、服务时长等内容,按照作业设定的服务需求和参数信息执行相应任务。在作业设定的服务时长内,提供服务的目标服务节点不间断地进行相关业务处理。服务处理时长超过设定的有效期限后,目标服务节点停止当前任务处理。

[0115] S506:目标服务节点将执行结果发送至请求该服务的客户端地址。

[0116] S507:客户端收到执行结果,若还需继续请求服务,则转向S501,重新开始服务节点寻址以及请求服务连接;若客户端不需要继续请求服务,则停止。

[0117] 基于上述实施例,本发明实施例提供一种第一服务节点每间隔一段指定时间向资源管理装置获取服务节点列表的方法,参见图6,具体包含有如下内容:

[0118] S601:服务节点采集本机运行状态,获取负载指标。

[0119] 需要说明的是,服务节点包括第一服务节点和第二服务节点。服务节点采集其在时刻T1至当前时刻Tn之间的运行状态。

[0120] 为了提高收集服务节点运行状态的准确性,将时刻T1至时刻Tn之间的时间区间分割成若干时间段,每隔一时间段采集当前的负载参数。已知ΔT为时刻T1至时刻Tn之间的时间区间,ΔT=Tn-T1,将ΔT切割为n份,每份时长为ΔT/n,则采集负载参数的时刻分别是T1、T1+ΔT/n、T1+2·ΔT/n、……、Tn,分别用T1、T2、T3、……、Tn代替表述。在T1至Tn的每一时刻,服务节点采集自身的负载参数,包括:CPU使用率p1、内存使用率p2、硬盘使用率p3以及当前服务节点作业队列长度p4,其中,服务节点作业长度在S504可采集获得。已知pi(i=1,2,3,4)为某一项指标(CPU、内存、硬盘、所在节点作业队列长度)的负载参数,pi(Ti)即某一项指标在Ti时刻提取的负载状态,则,在T1、T2、T3至Tn的值为pi(T1)、pi(T2)、pi(T3)、……、pi(Tn)。取均值$\overline{pi}=\frac{\sum pi(Ti)}{n}$代表当前某负载参数的平均值。最后将所有指标的负载参数汇总可求得负载指标$\sum \alpha i\cdot \overline{pi}$,这里,αi为对应负载均值的权重。负载指标既能评价服务节点的实时运行状态,也是资源管理装置选择服务节点列表的标准。由于第一服务节点在为客户端服务的同时负担请求调度的任务,因此在安排工作时第一服务节点为客户端服务的工作量要小于第二服务节点的工作量。如上所述第一服务节点在设置服务节点

作业队列长度对应权重时，其值要大于第二服务节点，例如设置为第二服务节点的1.5倍。

[0121] S602：服务节点定期向资源管理装置发送心跳。

[0122] 每间隔一定时间，服务节点将当前时间戳以及当前工作负载指标随同本机地址包装成一条心跳信息，并向资源管理装置发送心跳。

[0123] S603：资源管理装置维护着全部服务节点的负载信息。

[0124] 资源管理装置将获取的心跳信息进行解析，从中提取服务节点地址、心跳时间戳和服务节点负载指标。再分别以[服务节点地址，心跳时间戳]和[服务节点地址，负载指标]两组信息添加至服务节点心跳时间集与服务节点负载集，若集合中已经含有上述服务节点地址的旧信息内容，则以新数据代替旧数据。

[0125] S604：第一服务节点向资源管理装置请求服务节点列表。

[0126] 第一服务节点由于其调度客户端请求的需要，定期向资源管理装置请求服务节点列表。如此，第一服务节点在调度客户端发送的服务请求时，能够反馈给客户端目标节点地址，均衡整个服务集群的工作任务量。

[0127] S605：资源管理装置计算并反馈请求结果，并及时删除心跳过期的服务节点。

[0128] 资源管理装置在收到第一服务节点的请求，找到全体服务节点的资源信息。首先，资源管理装置从服务节点心跳时间集筛选出心跳时间戳为有效时间的服务节点，删除心跳过期的服务节点，并且将删除的服务节点汇总至故障服务节点列表。然后，将服务节点负载集与故障服务节点列表进行比对，删除服务节点负载集中包含故障服务节点的信息。筛选得到的服务节点负载集为服务节点考察集，服务节点考察集包含了当前时间段可正常为客户端提供服务的服务节点以及其负载状态。资源管理装置从服务节点考察集中提取小于集群平均负载的多个服务节点，包装成一条数据包反馈给提出请求的第一服务节点。其中，提取的多个服务节点即为第一服务节点存储的服务节点列表。

[0129] S606：第一服务节点及时更新服务节点列表。

[0130] 第一服务节点收到数据包，解析出数据包中的若干服务节点地址，添加到地址列表中，同时删除地址列表中的过期信息。

[0131] 为了提高负载均衡的稳定性和准确性并降低运维成本，本发明提供一种资源负载均衡系统的实施例，参见图7，具体包含如下内容：

[0132] 多个第一服务节点10和资源管理装置30；

[0133] 所述第一服务节点10用于实现适用于服务端的资源负载均衡方法；

[0134] 各个所述第一服务节点10均与所述客户端通信连接，且该客户端用于实现适用于客户端的资源负载均衡方法；

[0135] 各个所述第一服务节点10分别周期性向资源管理装置30发送心跳信息，以使所述资源管理装置30根据所述心跳信息生成所述服务节点列表并对所述服务节点列表进行周期性更新，以及将更新后的服务节点列表发送至多个第一服务节点10。

[0136] 进一步的，还包括：多个第二服务节点20；

[0137] 各个所述第二服务节点20均与所述客户端通信连接；

[0138] 其中，各个所述第一服务节点10和各个所述第二服务节点20分别周期性向资源管理装置30发送心跳信息，以使所述资源管理装置30根据所述心跳信息生成所述服务节点列表并对所述服务节点列表进行周期性更新，以及将更新后的服务节点列表发送至多个第一

服务节点10。

[0139] 在本实施例中，资源负载均衡系统包括：多个第一服务节点10、多个第二服务节点20和资源管理装置30；

[0140] 其中，第一服务节点10和第二服务节点20均可对外提供服务，第二服务节点20仅为客户端提供服务的普通节点，第一服务节点10在为客户端提供服务的同时，还承担负责路由转发的职责。资源管理装置30用于维护和管理所有第一服务节点10和第二服务节点20的工作负载状态，所获取的第一服务节点10和第二服务节点20的负载信息由第一服务节点10和第二服务节点20的周期性心跳收集汇总而成，所述第一服务节点10和第二服务节点20的部署运行于服务器之上，可以包括但不限于是一个程序。

[0141] 第一服务节点10的结构如图8所示，包括：负载评价单元101、工作单元102以及请求调度单元103。第一服务节点10兼顾客户端的请求调度和为客户端提供服务的双重职责。实现避免单节点故障以及提高客户端请求调度的并发处理能力，资源负载均衡系统设有多个第一服务节点10，第一服务节点10之间相互独立工作。由上述可知，某第一服务节点10发生故障时其余第一服务节点10依旧可承担请求调度任务，保障了资源负载均衡系统不会因单节点故障而发生中断。

[0142] 其中，负载评价单元101用于定期采集本机运行数据，并在此基础上计算出负载指标后，将该负载指标封装成一个心跳信息，写入资源管理装置30。本机运行数据包括：cpu使用率、内存使用率、硬盘使用率以及以及当前服务节点作业队列长度，按不同权重比计算求和得出一个负载指标，负载指标作为统一标准，用于综合评价当前节点的负载情况。

[0143] 具体的，采集到本机器运行数据后，按照公式“cpu使用率·权重1+内存使用率·权重2+硬盘使用率·权重3+当前服务节点作业队列长度·权重4”可算得负载指标。负载指标的数值越高则说明该服务节点的负载越重，在分配新的客户请求时将请求指派给负载小于负载平均值的服务节点。负载评价单元101将获取的负载指标、当前机器时间以及本机器的节点地址及时写入资源管理装置30。

[0144] 工作单元102用于为上游客户端提供服务的核心单元。工作单元102可为多个客户端进行计算、数据传送等服务。工作单元102和客户端建立一段随机时长的连接，在该连接时长内，工作单元102为客户端服务，服务时长到期后，连接断开，服务停止。

[0145] 请求调度单元103用于为客户端的服务请求分配一个目标服务节点，向其反馈该目标服务节点的节点地址。当客户端需要服务器集群为其提供服务时，首先需要向第一服务器节点发起Request_Address的请求，服务节点地址的数目可事先指定。具体地，请求调度单元103存储了服务节点列表，表结构如表1所示。该列表由请求调度单元103定期向资源管理装置30定期获取所得。当请求调度单元103收到客户端Request_Address请求时，首先将该请求信息放入服务请求队列尾部，请求调度单元103以“先进先出”原则为客户端服务。请求调度单元103每次从服务请求队列头部中获取一个请求项，提取客户端的IP地址，随后从优选地址列表中根据随机策略选择一个服务地址，将随机选定的地址反馈给该客户端IP。通过随机分配，尽可能将客户端平均安排给不同的服务节点。

[0146] 表1：服务节点列表

[0147]

序号	服务节点地址
1	IP地址1

2	IP地址2
3	IP地址3

[0148] 以10000台客户端以5-10分钟一次的频率异步向资源负载均衡系统调用服务请求为例，第一服务节点10的集群部署3台服务器进行路由调度，根据公式“客户端数目/(服务器数目•客户端请求频率•60秒)”可知一台服务器每秒接收的请求在5.5至11之间，这样的调度频率可被一般计算机所接受。因此，资源负载均衡系统正常工作时，第一服务节点10本身承担的请求调度的压力较轻，双重职能可更高效地利用服务器资源。

[0149] 需要说明的是，当客户端数量或服务请求频率数量级增长时，需适当增设第一服务节点，以确保资源负载均衡系统提供稳定的负载均衡服务。

[0150] 第二服务节点20包括：负载评价单元和工作单元。为了提高任务完成的并发性以及减少客户端等待时间，设有多个第二服务节点20，且第二服务节点的数目可由当前资源负载均衡系统的负载进行动态增减。

[0151] 由于第二服务节点20同第一服务节点10内部结构相似，第二服务节点20的负载评价单元和工作单元的工作内容与第一服务节点10的负载评价单元101和工作单元102相同，因此关于第二服务节点20的负载评价单元和工作单元的详细说明请参考第一服务节点10中的负载评价单元101和工作单元102，在此不再赘述。

[0152] 资源管理装置30采用以“键-值”为存储方式的内存数据库技术，资源管理装置30借助Redis技术对第一服务节点10和第二服务节点20进行运行状态的管理和维护。其中，每隔预设的一定时间，各第一服务节点10和各第二服务节点20均向资源管理装置30发送心跳信息。心跳信息具体包括：第一服务节点10和第二服务节点20的发送心跳的时间戳以及心跳时期的工作负载指标。资源管理装置30在收集汇总心跳信息后转化为第一服务节点10和第二服务节点20的心跳时间集与服务节点负载指标集，这两个集合均采用了Redis内存数据库内部提供的顺序集SortedSet维护数据信息，它能保证加入集合的数据按照指定的规则有序排列，可快速获取一定区间内的数据。具体地，在接收第一服务节点10和第二服务节点20的心跳信息时，从中提取出对应服务节点地址、心跳时间戳和负载指标，分别以[服务节点地址，心跳时间戳]和[服务节点地址，负载指标]两组信息添加至服务节点心跳时间集与服务节点负载集，若集合中已经含有上述服务节点地址的旧信息内容，则以新数据代替旧数据。每隔一段指定时间，资源管理装置30通过服务节点心跳时间集来考察是否有故障服务节点，若某个节点的最新时间戳并不在指定时间戳和当前时间戳的范围中，即证明该节点未及时发送心跳，因此有极大几率发生故障。资源管理装置30若发现故障节点，将故障节点汇总至故障服务节点列表，同时将该列表同服务节点负载集比对，从服务节点负载集中排除所有故障节点。处理后的服务节点负载集也称为服务节点考察集，服务节点考察集包含了当前时间段可正常为客户端提供服务的服务节点以及其负载状态。

[0153] 由上述可知，资源管理装置30不仅能够获取正常运行的第一服务节点10和第二服务节点20的心跳信息，而且可为第一服务节点10的客户端请求指派工作提供合适的优选节点列表，从而均衡第一服务节点10和第二服务节点20的工作量。对于预设时间段内未发送心跳的服务节点，管理装置认为该节点发生故障，在确定提供服务的节点时不会选择该节点。当故障节点恢复工作后会再次向管理装置发送心跳，管理装置一旦收到故障服务节点重新发送的心跳信息，则将该节点列入集群节点的考察对象集中，该节点可再次对外提供

服务。

[0154] 从上述描述可知，本发明实施例提供的一种资源负载均衡系统，通过将客户端的服务请求以系统集群分布式协作的方式得到均衡的分配和处理，实现提高工作效率。

[0155] 需要说明的是，本实施例提供的资源负载均衡系统可服务于具有灾备性质的跨区域服务集群，即，资源负载均衡系统可部署在多个区域的服务节点集群中，若一个区域遭遇不可抗拒因素而导致该区域内的集群集体瘫痪时，其他区域仍旧正常工作并承担起原本属于瘫痪区域的工作量。由于存在多个跨区域集群，所以客户端向集群请求服务时必须清楚其所在区域，以及区分位于不同区域的所有第一服务节点，具体包括如下内容：

[0156] 每个客户端内部维护有所有区域的第一服务节点信息表，如表2所示。其中，每个客户端内部的第一服务节点信息表的内容均相同。

[0157] 表2:第一服务节点信息表

[0158]

区域ID	第一服务节点地址
1	IP地址1
1	IP地址2
2	IP地址3
2	IP地址4
……	……

[0159] 在不同区域的资源管理装置中各配置一张客户端信息表。客户端信息表用于记录不同区域(用区域ID标记)和区域内客户端IP地址范围，如表3所示。

[0160] 表3:客户端信息表

[0161]

区域ID	客户端地址范围
1	IP地址A～IP地址B
2	IP地址C～IP地址D
……	……

[0162] 客户端首次提交申请时，随机从本机的所有第一服务节点列表中选择其一，向其发送LOCATION_REQUEST指令。第一服务节点列表收到该客户端LOCATION_REQUEST指令，提取该客户端的IP地址信息，查询本机缓存的客户端信息表可知该客户端所在区域。随后第一服务节点将客户端所属区域ID转发给该客户端。经过上述流程，客户端即可获知本机所处区域，并可区分本区域以及非本区域的第一节点信息。

[0163] 当客户端所在区域内的系统集群正常运行时，客户端会优先向本区域内部集群请求服务。具体地，在每次请求服务时先随机选择一台与本机所属区域ID相同的第一服务节点发送ADDRESS_REQUEST请求，在获取到服务地址后同该地址建立服务连接。图9为区域集群发生瘫痪时该区域客户端重新寻址的示意图。如图9所示，区域1集群瘫痪，客户端向本区域内第一服务器节点进行ADDRESS_REQUEST请求时，发现请求超时，则说明该第一服务节点发生故障无法提供服务。客户端连续指定次数(本示例设置为2次)向不同的第一服务节点发起ADDRESS_REQUEST请求皆失败，可判定所属区域发生异常。由此，客户端选择跨区域请求服务。在之后的所有ADDRESS_REQUEST请求时，客户端随机从非本区域ID的第一服务节点列表中选择其一，即本示例中区域2的第一服务节点。选中的第一服务节点收到请求后反馈给该客户端本集群中的优选服务节点，后续由该区域的服务节点为瘫痪区域的客户端进行

服务。

[0164] 上述设计方案的意义在于，在本区域内集群可以正常工作的前提下，本区内的客户端服务在本区域内部处理，这样的工作效率最高；当本区域内发生集群瘫痪的异常时，本区内的客户端服务通过另一个区域的集群来进行处理，保证客户端请求得到及时处理。需要说明的是，当瘫痪区域的服务功能恢复后，同一区域的客户端服务又自动切换到本区域内部进行处理。

[0165] 从上述描述可知，本发明实施例提供的资源负载均衡系统，设置了多个负责调度客户端请求的第一服务节点以及多个可为客户端提供服务的节点，成功避免了单点故障的负面影响，提高资源负载均衡系统的可靠性；并且由第一服务节点直接接收客户端发送的服务请求，降低了资源负载均衡系统的运维成本；在为客户端提供服务时选择负载小于平均负载的服务节点，提高资源负载均衡系统的工作效率；而且本发明实施例提供的系统运维成本低资源负载均衡系统具备灾备能力，确保客户端的服务在服务集群瘫痪等极端恶劣情况下依旧可以得到正常处理。

[0166] 本发明提供一种第一服务节点，参见图10，具体包含如下内容：

[0167] 服务节点查找模块1001，用于接收客户端直接发送的请求服务指令，根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址，其中，所述客户端用于每次向多个第一服务节点中的任意一个发送请求服务指令；

[0168] 服务节点发送模块1002，用于将所述目标服务节点的节点地址发送至对应的客户端，以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点。

[0169] 本发明提供的第一服务节点的实施例具体可以用于执行上述实施例中的适用于服务端的资源负载均衡方法的实施例的处理流程，其功能在此不再赘述，可以参照上述适用于服务端的资源负载均衡方法的实施例的详细描述。

[0170] 本发明提供一种客户端，参见图11，具体包含如下内容：

[0171] 指令发送模块1101，用于向多个第一服务节点中的任意一个发送请求服务指令，以使接收到该请求服务指令的第一服务节点根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址；

[0172] 访问模块1102，用于若在第二预设时间内接收到对应的第一服务节点发送的所述目标服务节点的节点地址，则根据所述目标服务节点的节点地址访问所述目标服务节点。

[0173] 本发明提供的客户端的实施例具体可以用于执行上述实施例中的适用于客户端的资源负载均衡方法的实施例的处理流程，其功能在此不再赘述，可以参照上述适用于客户端的资源负载均衡方法的实施例的详细描述。

[0174] 本发明的实施例还提供能够实现上述实施例中的适用于服务端的资源负载均衡方法中全部步骤的一种电子设备的具体实施方式，参见图12，所述电子设备具体包括如下内容：

[0175] 处理器(processor)601、存储器(memory)602、通信接口(Communications Interface)603和总线604；

[0176] 其中，所述处理器601、存储器602、通信接口603通过所述总线604完成相互间的通

信;所述处理器601用于调用所述存储器602中的计算机程序,所述处理器执行所述计算机程序时实现上述实施例中的适用于服务端的资源负载均衡方法中的全部步骤,例如,所述处理器执行所述计算机程序时实现下述步骤:接收客户端直接发送的请求服务指令,根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系,确定与该请求服务指令对应的目标服务节点的节点地址,其中,所述客户端用于每次向多个第一服务节点中的任意一个发送请求服务指令;将所述目标服务节点的节点地址发送至对应的客户端,以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点。

[0177] 从上述描述可知,本发明实施例提供的电子设备,通过接收客户端直接发送的请求服务指令,根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系,确定与该请求服务指令对应的目标服务节点的节点地址,其中,所述客户端用于每次向多个第一服务节点中的任意一个发送请求服务指令;实现由多个第一服务节点接收客户端的请求,避免了单节点故障的负面影响,提高负载均衡的稳定性和准确性;将所述目标服务节点的节点地址发送至对应的客户端,以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点,实现由服务节点分配客户端的请求,无需另设置负载均衡设备降低运维成本。

[0178] 本发明的实施例还提供能够实现上述实施例中的适用于服务端的资源负载均衡方法中全部步骤的一种计算机可读存储介质,所述计算机可读存储介质上存储有计算机程序,该计算机程序被处理器执行时实现上述实施例中的适用于服务端的资源负载均衡方法的全部步骤,例如,所述处理器执行所述计算机程序时实现下述步骤:接收客户端直接发送的请求服务指令,根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系,确定与该请求服务指令对应的目标服务节点的节点地址,其中,所述客户端用于每次向多个第一服务节点中的任意一个发送请求服务指令;将所述目标服务节点的节点地址发送至对应的客户端,以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点。

[0179] 从上述描述可知,本发明实施例提供的计算机可读存储介质,通过接收客户端直接发送的请求服务指令,根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系,确定与该请求服务指令对应的目标服务节点的节点地址,其中,所述客户端用于每次向多个第一服务节点中的任意一个发送请求服务指令;实现由多个第一服务节点接收客户端的请求,避免了单节点故障的负面影响,提高负载均衡的稳定性和准确性;将所述目标服务节点的节点地址发送至对应的客户端,以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点,实现由服务节点分配客户端的请求,无需另设置负载均衡设备降低运维成本。

[0180] 虽然本发明提供了如实施例或流程图所述的方法操作步骤,但基于常规或者无创造性的劳动可以包括更多或者更少的操作步骤。实施例中列举的步骤顺序仅仅为众多步骤执行顺序中的一种方式,不代表唯一的执行顺序。在实际中的装置或客户端产品执行时,可以按照实施例或者附图所示的方法顺序执行或者并行执行(例如并行处理器或者多线程处理的环境)。

[0181] 上述实施例阐明的系统、装置、模块或单元,具体可以由计算机芯片或实体实现,或者由具有某种功能的产品来实现。一种典型的实现设备为计算机。具体的,计算机例如可

以为个人计算机、膝上型计算机、车载人机交互设备、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任何设备的组合。

[0182] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0183] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0184] 在一个典型的配置中，计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0185] 内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器(RAM)和/或非易失性内存等形式，如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

[0186] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括暂存电脑可读媒体(transitory media)，如调制的数据信号和载波。

[0187] 本领域技术人员应明白，本说明书的实施例可提供为方法、系统或计算机程序产品。因此，本说明书实施例可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。

[0188] 本说明书实施例可以在由计算机执行的计算机可执行指令的一般上下文中描述，例如程序模块。一般地，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本说明书实施例，在这些分布式计算环境中，由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中，程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[0189] 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。在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且

还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。对于本领域的普通技术人员而言，可以根据具体情况理解上述术语在本发明中的具体含义。

[0190] 本发明的说明书中，说明了大量具体细节。然而能够理解的是，本发明的实施例可以在没有这些具体细节的情况下实践。在一些实例中，并未详细示出公知的方法、结构和技术，以便不模糊对本说明书的理解。类似地，应当理解，为了精简本发明公开并帮助理解各个发明方面中的一个或多个，在上面对本发明的示例性实施例的描述中，本发明的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而，并不应将该公开的方法解释呈反映如下意图：即所要求保护的本发明要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说，如权利要求书所反映的那样，发明方面在于少于前面公开的单个实施例的所有特征。因此，遵循具体实施方式的权利要求书由此明确地并入该具体实施方式，其中每个权利要求本身都作为本发明的单独实施例。需要说明的是，在不冲突的情况下，本发明中的实施例及实施例中的特征可以相互组合。本发明并不局限于任何单一的方面，也不局限于任何单一的实施例，也不局限于这些方面和/或实施例的任意组合和/或置换。而且，可以单独使用本发明的每个方面和/或实施例或者与一个或更多其他方面和/或其实施例结合使用。

[0191] 最后应说明的是：以上各实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述各实施例对本发明进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分或者全部技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本发明各实施例技术方案的范围，其均应涵盖在本发明的权利要求和说明书的范围当中。

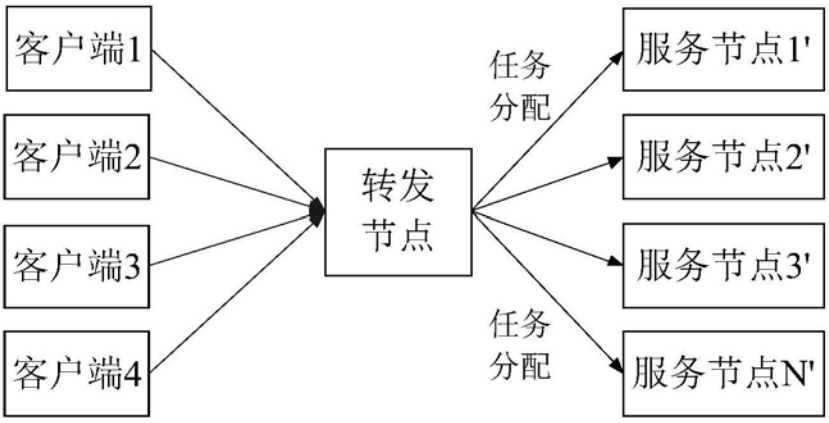

客户端1
客户端2
客户端3
客户端4
转发节点
任务分配
任务分配
服务节点1'
服务节点2'
服务节点3'
服务节点N'

图1

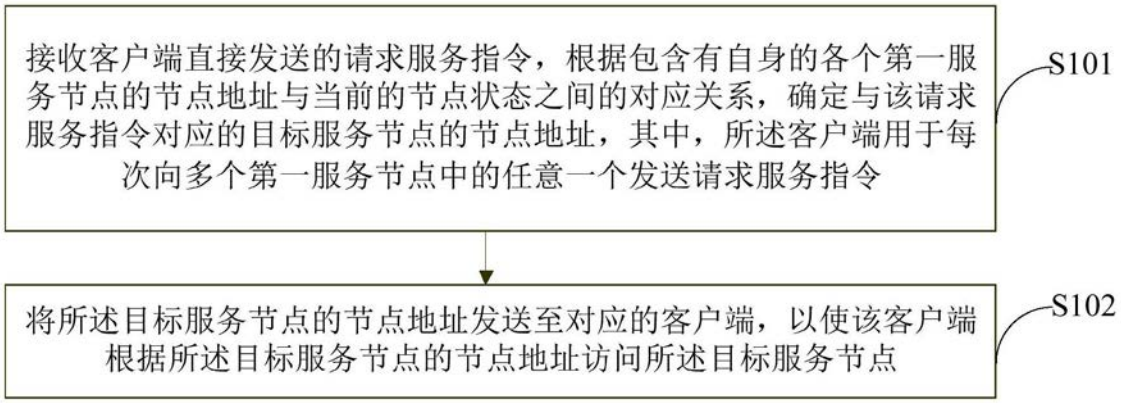

接收客户端直接发送的请求服务指令，根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址，其中，所述客户端用于每次向多个第一服务节点中的任意一个发送请求服务指令
S101
将所述目标服务节点的节点地址发送至对应的客户端，以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点
S102

图2

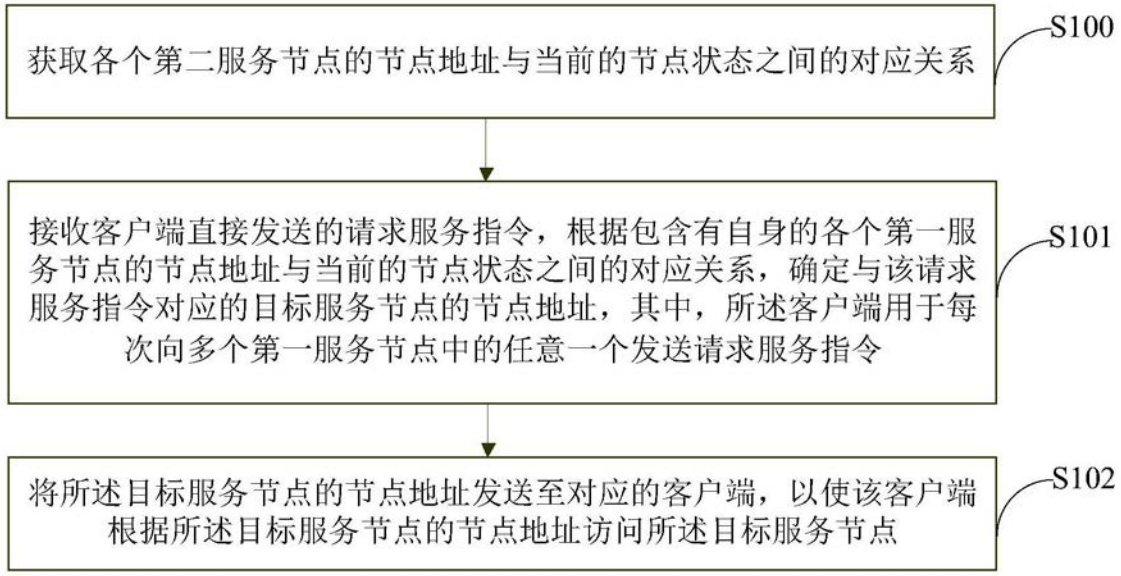

获取各个第二服务节点的节点地址与当前的节点状态之间的对应关系
S100
接收客户端直接发送的请求服务指令，根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址，其中，所述客户端用于每次向多个第一服务节点中的任意一个发送请求服务指令
S101
将所述目标服务节点的节点地址发送至对应的客户端，以使该客户端根据所述目标服务节点的节点地址访问所述目标服务节点
S102

图3

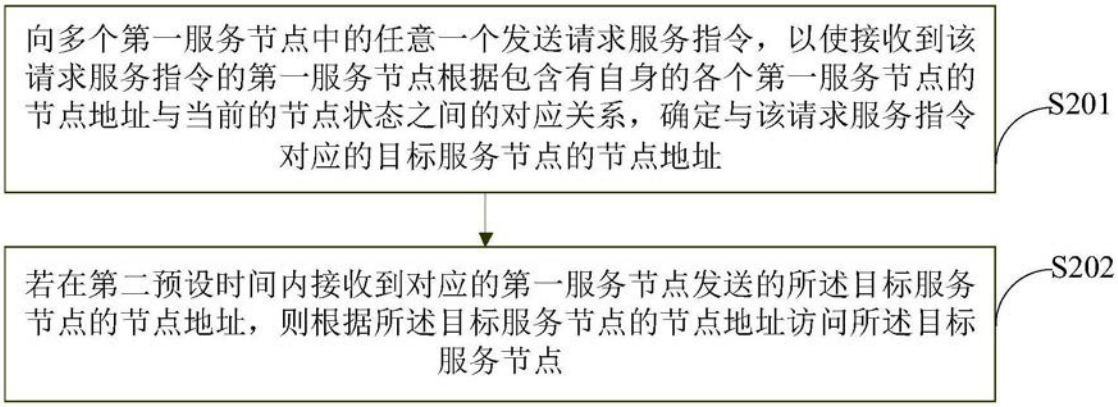

向多个第一服务节点中的任意一个发送请求服务指令，以使接收到该请求服务指令的第一服务节点根据包含有自身的各个第一服务节点的节点地址与当前的节点状态之间的对应关系，确定与该请求服务指令对应的目标服务节点的节点地址
S201
若在第二预设时间内接收到对应的第一服务节点发送的所述目标服务节点的节点地址，则根据所述目标服务节点的节点地址访问所述目标服务节点
S202

图4

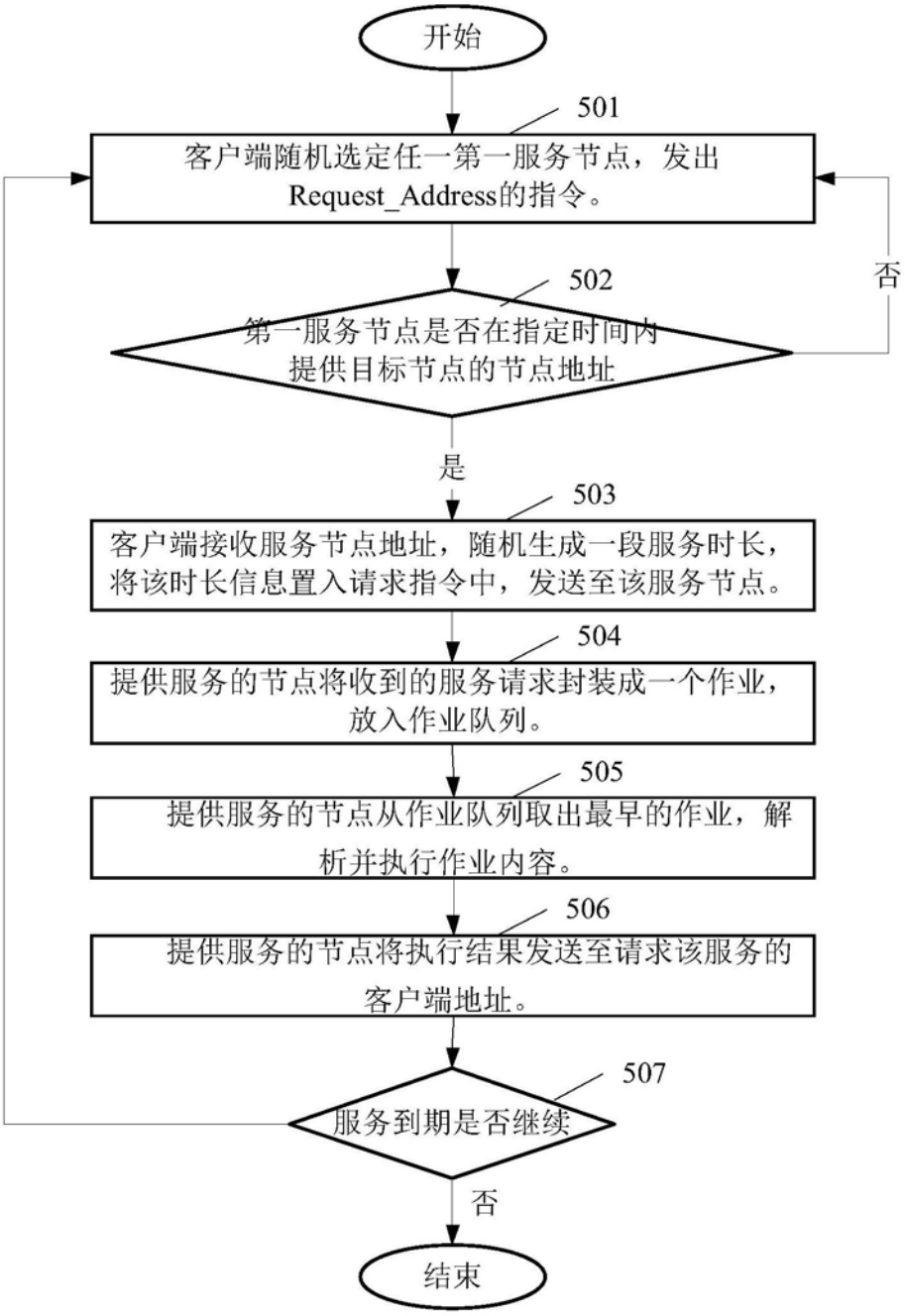

开始
501
客户端随机选定任一第一服务节点，发出Request_Address的指令。
502
第一服务节点是否在指定时间内提供目标节点的节点地址
否
是
503
客户端接收服务节点地址，随机生成一段服务时长，将该时长信息置入请求指令中，发送至该服务节点。
504
提供服务的节点将收到的服务请求封装成一个作业，放入作业队列。
505
提供服务的节点从作业队列取出最早的作业，解析并执行作业内容。
506
提供服务的节点将执行结果发送至请求该服务的客户端地址。
507
服务到期是否继续
否
结束

图5

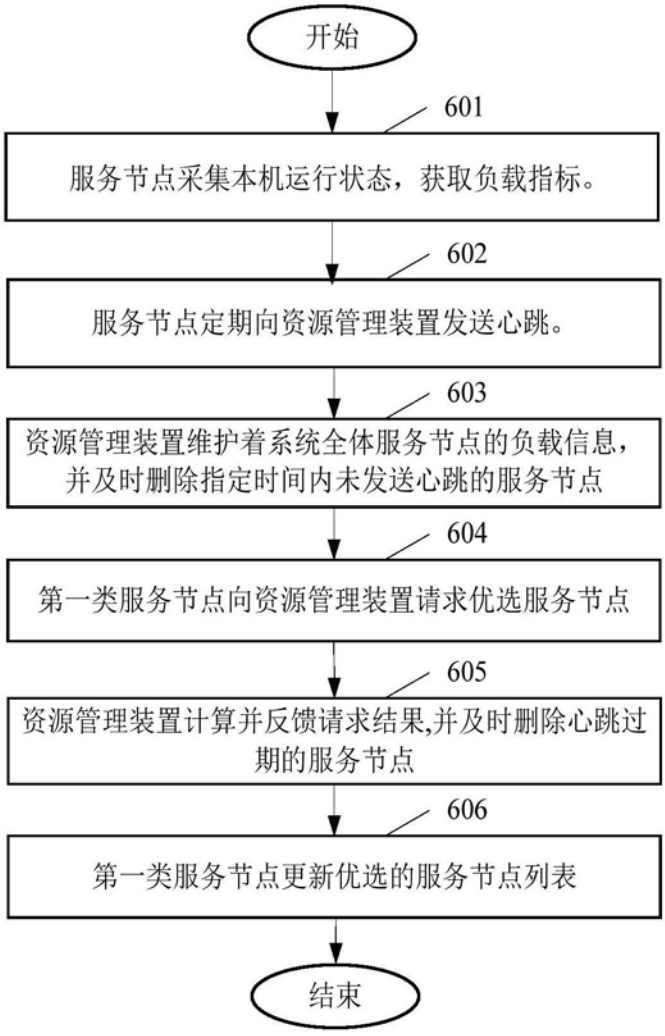

开始
601
服务节点采集本机运行状态，获取负载指标。
602
服务节点定期向资源管理装置发送心跳。
603
资源管理装置维护着系统全体服务节点的负载信息，并及时删除指定时间内未发送心跳的服务节点
604
第一类服务节点向资源管理装置请求优选服务节点
605
资源管理装置计算并反馈请求结果,并及时删除心跳过期的服务节点
606
第一类服务节点更新优选的服务节点列表
结束

图6

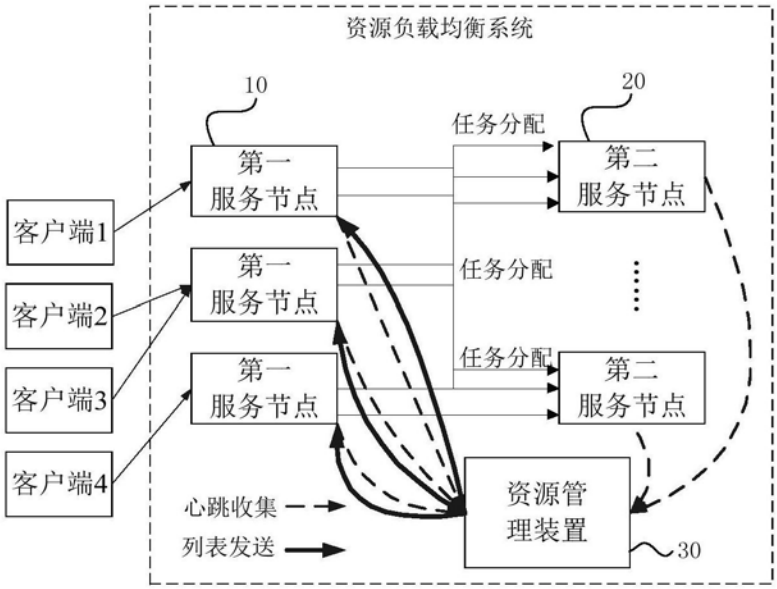

资源负载均衡系统
10
20
任务分配
第一
服务节点
第二
服务节点
客户端1
客户端2
客户端3
客户端4
第一
服务节点
任务分配
第一
服务节点
任务分配
第二
服务节点
资源管
理装置
30
心跳收集
列表发送

图7

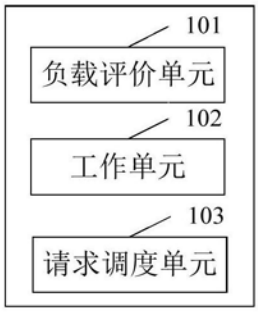

101
负载评价单元
102
工作单元
103
请求调度单元

图8

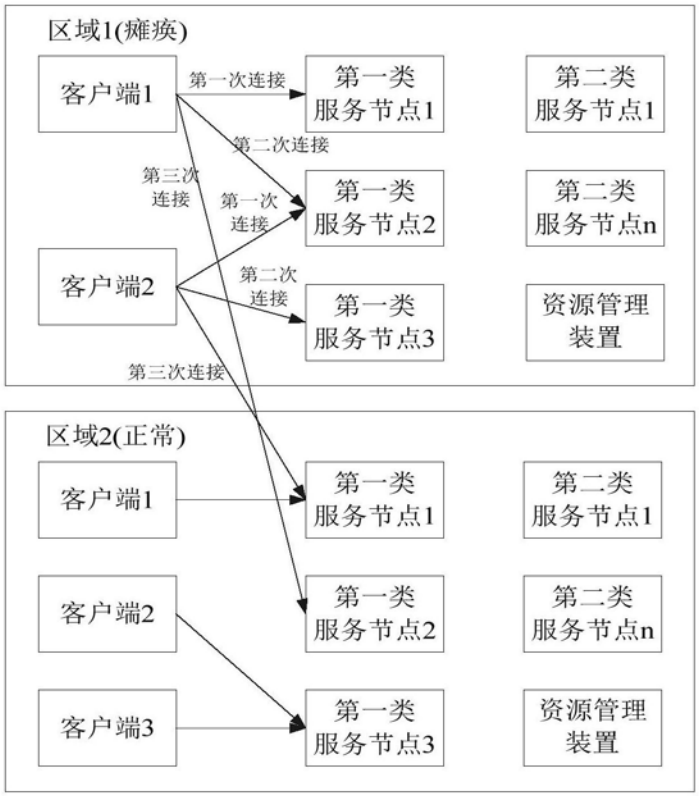

区域1(瘫痪)
客户端1
客户端2
第一次连接
第二次连接
第三次连接
第一次连接
第二次连接
第三次连接
第一类服务节点1
第一类服务节点2
第一类服务节点3
第二类服务节点1
第二类服务节点n
资源管理装置
区域2(正常)
客户端1
客户端2
客户端3
第一类服务节点1
第一类服务节点2
第一类服务节点3
第二类服务节点1
第二类服务节点n
资源管理装置

图9

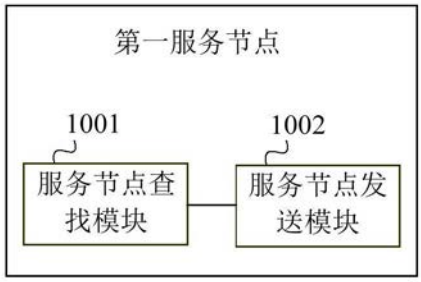

第一服务节点
1001
1002
服务节点查找模块
服务节点发送模块

图10

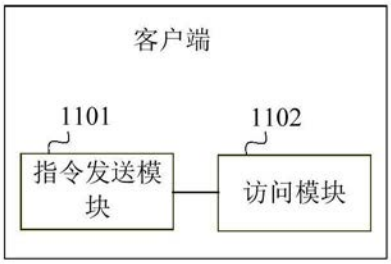

客户端
1101
1102
指令发送模块
访问模块

图11

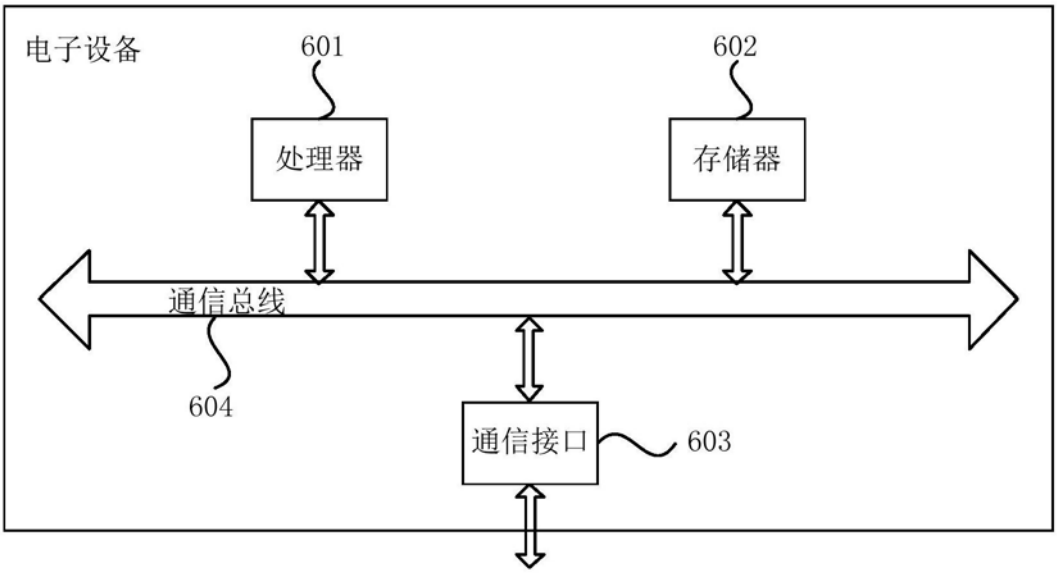


图12