



(12)发明专利申请

(10)申请公布号 CN 110138837 A

(43)申请公布日 2019.08.16

(21)申请号 201910301116.1

(22)申请日 2019.04.15

(71)申请人 平安科技(深圳)有限公司

地址 518000 广东省深圳市福田区福田街
道福安社区益田路5033号平安金融中
心23楼

(72)发明人 易杉峰

(74)专利代理机构 深圳市明日今典知识产权代
理事务所(普通合伙) 44343

代理人 王杰辉

(51)Int.Cl.

H04L 29/08(2006.01)

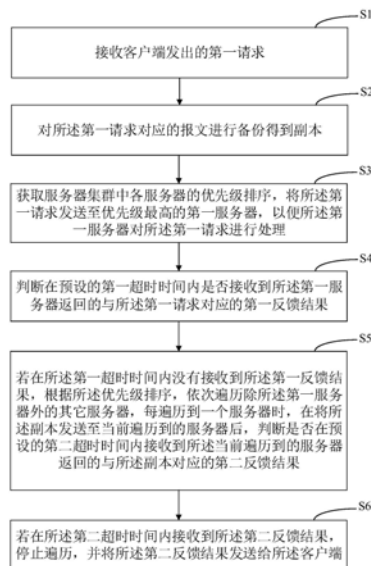
权利要求书2页 说明书17页 附图4页

(54)发明名称

请求处理方法、装置、计算机设备和存储介
质

(57)摘要

本申请揭示了一种请求处理方法、装置、计
算机设备和存储介质,其中方法包括:接收客
户端发出的第一请求;对第一请求对应的报
文进行备份得到副本;获取服务器集群中各
服务器的优先级排序,将第一请求发送至优
先级最高的第一服务器;判断在第一超时时
间内是否接收到第一服务器返回的第一反
馈结果;若否,根据优先级排序,依次遍历
除第一服务器外的其它服务器,每遍历到
一个服务器时,在将副本发送至当前遍历
到的服务器后,判断是否在预设的第二超
时时间内接收到第二反馈结果;若是,停
止遍历,并将第二反馈结果发送给客户端。
通过本申请能在第一服务器异常的情况下
调用其它的服务器来对第一请求处理,避
免出现对第一请求的处理产生异常超时。



CN 110138837 A

1. 一种请求处理方法,其特征在于,包括:

接收客户端发出的第一请求;

对所述第一请求对应的报文进行备份得到副本;

获取服务器集群中各服务器的优先级排序,将所述第一请求发送至优先级最高的第一服务器,以便所述第一服务器对所述第一请求进行处理,其中所述优先级排序为按照各所述服务器的优先级数值从小到大的顺序进行排序得到;

判断在预设的第一超时时间内是否接收到所述第一服务器返回的与所述第一请求对应的第一反馈结果;

若在所述第一超时时间内没有接收到所述第一反馈结果,根据所述优先级排序,依次遍历除所述第一服务器外的其它服务器,每遍历到一个服务器时,在将所述副本发送至当前遍历到的服务器后,判断是否在预设的第二超时时间内接收到所述当前遍历到的服务器返回的与所述副本对应的第二反馈结果;

若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端。

2. 根据权利要求1所述的请求处理方法,其特征在于,所述获取服务器集群中各服务器的优先级排序,将所述第一请求发送至优先级最高的第一服务器的步骤之前,包括:

获取各所述服务器的运行参数数据;

根据所述运行参数数据,按照预设规则计算各所述服务器的优先级数值;

将各所述服务器的优先级数值按照从小到大的顺序进行排序,得到所述优先级排序。

3. 根据权利要求2所述的请求处理方法,其特征在于,所述根据所述运行参数数据,按照预设规则计算各所述服务器的优先级数值的步骤,包括:

设置与各运行参数分别对应的权重值;

根据各所述运行参数数据与各所述权重值计算各所述服务器的运行参数和值;

将所述运行参数和值确定为所述优先级数值。

4. 根据权利要求1所述的请求处理方法,其特征在于,当前遍历到的服务器是在所述优先级排序中排序最末的服务器,所述判断是否在预设的第二超时时间内接收到所述当前遍历到的服务器返回的与所述副本对应的第二反馈结果的步骤之后,包括:

若在所述第二超时时间内没有接收到所述第二反馈结果,则向所述客户端返回请求处理失败信息;

为所有所述服务器添加异常标记;

向维修人员发出维修信息,以便所述维修人员对所有所述服务器进行维修处理。

5. 根据权利要求1所述的请求处理方法,其特征在于,所述若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端的步骤之后,包括:

筛选出没有返回与所述第一请求对应的反馈结果的指定服务器,并为所述指定服务器添加异常标记,其中所述指定服务器的数量为一个或多个;

向各所述指定服务器发送心跳信号;

若存在在所述第一预设时间内返回了心跳回复信息的第一指定服务器,则对所述第一指定服务器解除所述异常标记。

6. 根据权利要求5所述的请求处理方法,其特征在于,所述向各所述指定服务器发送心跳信号的步骤之后,包括:

若存在在所述第一预设时间内没有返回心跳回复信息的第二指定服务器,则按照预设的时间间隔向所述第二指定服务器发送心跳信息;

记录所述第二指定服务器没有返回心跳回复信息的次数;

判断所述次数是否大于预设的次数阈值;

若是,向维修人员发出维修信息,以便所述维修人员对所述第二指定服务器进行维修处理。

7. 根据权利要求6所述的请求处理方法,其特征在于,所述若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端的步骤之后,包括:

判断是否接收到新的第二请求;

若是,根据各所述服务器的标记状态,从所有所述服务器中筛选出没有添加所述异常标记的第三指定服务器;

根据所述优先级排序,从所述第三指定服务器中确定出优先级最高的第二服务器;

将所述第二请求发送至所述第二服务器,以便所述第二服务器对所述第二请求进行处理。

8. 一种请求处理装置,其特征在于,包括:

接收模块,用于接收客户端发出的第一请求;

备份模块,用于对所述第一请求对应的报文进行备份得到副本;

第一发送模块,用于获取服务器集群中各服务器的优先级排序,将所述第一请求发送至优先级最高的第一服务器,以便所述第一服务器对所述第一请求进行处理,其中所述优先级排序为按照各所述服务器的优先级数值从小到大的顺序进行排序得到;

第一判断模块,用于判断在预设的第一超时时间内是否接收到所述第一服务器返回的与所述第一请求对应的第一反馈结果;

遍历模块,用于若在所述第一超时时间内没有接收到所述第一反馈结果,根据所述优先级排序,依次遍历除所述第一服务器外的其它服务器,每遍历到一个服务器时,在将所述副本发送至当前遍历到的服务器后,判断是否在预设的第二超时时间内接收到所述当前遍历到的服务器返回的与所述副本对应的第二反馈结果;

第二发送模块,用于若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端。

9. 一种计算机设备,包括存储器和处理器,所述存储器中存储有计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至7中任一项所述方法的步骤。

10. 一种存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1至7中任一项所述的方法的步骤。

请求处理方法、装置、计算机设备和存储介质

技术领域

[0001] 本申请涉及通信技术领域,具体涉及一种请求处理方法、装置、计算机设备和存储介质。

背景技术

[0002] 当用户通过客户端向服务器集群内的任意一个服务器发送请求时,在经过请求分发服务器对该请求的分发后,如果接收到请求的服务器在对上述请求进行处理的过程中突然宕机,或者出现短暂异常的情况,例如服务器由于网络连接过多产生了OOM(Out Of Memory,内存溢出),则服务器无法向客户端返回与请求对应的反馈信息,使得客户端一直在花费时间等待服务器响应,最后却接收到了请求处理失败的结果,从而造成用户发出的请求不能得到有效的处理,用户体验不好。

发明内容

[0003] 本申请的主要目的为提供一种请求处理方法、装置、计算机设备和存储介质,现有的服务器在处理用户发出的请求后如果出现异常,会造成用户发出的请求不能得到有效的处理的技术问题。

[0004] 本申请提出一种请求处理方法,所述方法包括步骤:

[0005] 接收客户端发出的第一请求;

[0006] 对所述第一请求对应的报文进行备份得到副本;

[0007] 获取服务器集群中各服务器的优先级排序,将所述第一请求发送至优先级最高的第一服务器,以便所述第一服务器对所述第一请求进行处理,其中所述优先级排序为按照各所述服务器的优先级数值从小到大的顺序进行排序得到;

[0008] 判断在预设的第一超时时间内是否接收到所述第一服务器返回的与所述第一请求对应的第一反馈结果;

[0009] 若在所述第一超时时间内接收到所述第一反馈结果,根据所述优先级排序,依次遍历除所述第一服务器外的其它服务器,每遍历到一个服务器时,在将所述副本发送至当前遍历到的服务器后,判断是否在预设的第二超时时间内接收到所述当前遍历到的服务器返回的与所述副本对应的第二反馈结果;

[0010] 若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端。

[0011] 可选地,所述获取服务器集群中各服务器的优先级排序,将所述第一请求发送至优先级最高的第一服务器的步骤之前,包括:

[0012] 获取各所述服务器的运行参数数据;

[0013] 根据所述运行参数数据,按照预设规则计算各所述服务器的优先级数值;

[0014] 将各所述服务器的优先级数值按照从小到大的顺序进行排序,得到所述优先级排序。

- [0015] 可选地,所述根据所述运行参数数据,按照预设规则计算各所述服务器的优先级数值的步骤,包括:
- [0016] 设置与各运行参数分别对应的权重值;
- [0017] 根据各所述运行参数数据与各所述权重值计算各所述服务器的运行参数和值;
- [0018] 将所述运行参数和值确定为所述优先级数值。
- [0019] 可选地,当前遍历到的服务器是在所述优先级排序中排序最末的服务器,所述判断是否在预设的第二超时时间内接收到所述当前遍历到的服务器返回的与所述副本对应的第二反馈结果的步骤之后,包括:
- [0020] 若在所述第二超时时间内没有接收到所述第二反馈结果,则向所述客户端返回请求处理失败信息;
- [0021] 为所有所述服务器添加异常标记;
- [0022] 向维修人员发出维修信息,以便所述维修人员对所有所述服务器进行维修处理。
- [0023] 可选地,所述若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端的步骤之后,包括:
- [0024] 筛选出没有返回与所述第一请求对应的反馈结果的指定服务器,并为所述指定服务器添加异常标记,其中所述指定服务器的数量为一个或多个;
- [0025] 向各所述指定服务器发送心跳信号;
- [0026] 若存在在所述第一预设时间内返回了心跳回复信息的第一指定服务器,则对所述第一指定服务器解除所述异常标记。
- [0027] 可选地,所述向各所述指定服务器发送心跳信号的步骤之后,包括:
- [0028] 若存在在所述第一预设时间内没有返回心跳回复信息的第二指定服务器,则按照预设的时间间隔向所述第二指定服务器发送心跳信息;
- [0029] 记录所述第二指定服务器没有返回心跳回复信息的次数;
- [0030] 判断所述次数是否大于预设的次数阈值;
- [0031] 若是,向维修人员发出维修信息,以便所述维修人员对所述第二指定服务器进行维修处理。
- [0032] 可选地,所述若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端的步骤之后,包括:
- [0033] 判断是否接收到新的第二请求;
- [0034] 若是,根据各所述服务器的标记状态,从所有所述服务器中筛选出没有添加所述异常标记的第三指定服务器;
- [0035] 根据所述优先级排序,从所述第三指定服务器中确定出优先级最高的第二服务器;
- [0036] 将所述第二请求发送至所述第二服务器,以便所述第二服务器对所述第二请求进行处理。
- [0037] 本申请还提供一种请求处理装置,包括:
- [0038] 接收模块,用于接收客户端发出的第一请求;
- [0039] 备份模块,用于对所述第一请求对应的报文进行备份得到副本;
- [0040] 第一发送模块,用于获取服务器集群中各服务器的优先级排序,将所述第一请求

发送至优先级最高的第一服务器,以便所述第一服务器对所述第一请求进行处理,其中所述优先级排序为按照各所述第一服务器的优先级数值从小到大的顺序进行排序得到;

[0041] 第一判断模块,用于判断在预设的第一超时时间内是否接收到所述第一服务器返回的与所述第一请求对应的第一反馈结果;

[0042] 遍历模块,用于若在所述第一超时时间内没有接收到所述第一反馈结果,根据所述优先级排序,依次遍历除所述第一服务器外的其它服务器,每遍历到一个服务器时,在将所述副本发送至当前遍历到的服务器后,判断是否在预设的第二超时时间内接收到所述当前遍历到的服务器返回的与所述副本对应的第二反馈结果;

[0043] 第二发送模块,用于若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端。

[0044] 本申请还提供一种计算机设备,包括存储器和处理器,所述存储器中存储有计算机程序,所述处理器执行所述计算机程序时实现上述方法的步骤。

[0045] 本申请还提供一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现上述方法的步骤。

[0046] 本申请中提供的请求处理方法、装置、计算机设备和存储介质,具有以下有益效果:

[0047] 本申请中提供的请求处理方法、装置、计算机设备和存储介质,接收客户端发出的第一请求;对所述第一请求对应的报文进行备份得到副本;获取服务器集群中各服务器的优先级排序,将所述第一请求发送至优先级最高的第一服务器,以便所述第一服务器对所述第一请求进行处理;判断在预设的第一超时时间内是否接收到所述第一服务器返回的与所述第一请求对应的第一反馈结果;若在所述第一超时时间内接收到所述第一反馈结果,根据所述优先级排序,依次遍历除所述第一服务器外的其它服务器,每遍历到一个服务器时,在将所述副本发送至当前遍历到的服务器后,判断是否在预设的第二超时时间内接收到所述当前遍历到的服务器返回的与所述副本对应的第二反馈结果;若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端。本申请在调用优先级最高的第一服务器对接收到的第一请求进行处理时,如果没有在第一超时时间内没有接收到第一服务器返回的第一反馈结果,则会根据各服务器的优先级排序将对第一请求进行备份得到的副本转发到除第一服务器之外的其它的服务器进行处理,,从而能够在第一服务器异常的情况下快速地调用其它的服务器对该第一请求进行处理,从而避免了向用户返回请求处理失败的结果,保证了对用户发出的请求处理不会产生异常超时,保障了用户的使用体验。

附图说明

[0048] 图1是本申请一实施例的请求处理方法的流程示意图;

[0049] 图2是本申请另一实施例的请求处理方法的流程示意图;

[0050] 图3是本申请一实施例的请求处理装置的结构示意图;

[0051] 图4是本申请一实施例的计算机设备的结构示意图。

[0052] 本申请目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

具体实施方式

[0053] 应当理解,此处所描述的具体实施例仅仅用于解释本申请,并不用于限定本申请。

[0054] 需要说明,本申请实施例中所有方向性指示(诸如上、下、左、右、前、后……)仅用于解释在某一特定姿态(如附图所示)下各部件之间的相对位置关系、运动情况等,如果该特定姿态发生改变时,则该方向性指示也相应地随之改变,所述的连接可以是直接连接,也可以是间接连接。

[0055] 参照图1,本申请一实施例的请求处理方法,包括:

[0056] S1:接收客户端发出的第一请求;

[0057] S2:对所述第一请求对应的报文进行备份得到副本;

[0058] S3:获取服务器集群中各服务器的优先级排序,将所述第一请求发送至优先级最高的第一服务器,以便所述第一服务器对所述第一请求进行处理,其中所述优先级排序为按照各所述服务器的优先级数值从小到大的顺序进行排序得到;

[0059] S4:判断在预设的第一超时时间内是否接收到所述第一服务器返回的与所述第一请求对应的第一反馈结果;

[0060] S5:若在所述第一超时时间内接收到所述第一反馈结果,根据所述优先级排序,依次遍历除所述第一服务器外的其它服务器,每遍历到一个服务器时,在将所述副本发送至当前遍历到的服务器后,判断是否在预设的第二超时时间内接收到所述当前遍历到的服务器返回的与所述副本对应的第二反馈结果;

[0061] S6:若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端。

[0062] 如上述步骤S1至S6所述,本方法的执行主体为请求处理装置,具体可为请求分发服务器,上述的客户端可为手机,PDA(Personal Digital Assistant,个人数字助理)、膝上型计算机、掌上电脑等等,用户可通过该客户端来发送请求,来获取服务器的对于该请求的业务服务。上述优先级排序是指根据每个服务器的优先级数值进行排序得到的,优先级越高的服务器对应越空闲的工作状态,且处理请求的效率和速率越快,其中上述优先级数值可通过对服务器集群中各服务器的运行参数数据进行计算得到。上述的超时时间则是预先根据各服务器的优先级排序进行设置的,不同优先级的服务器可对应不同的超时时间,例如对于一个相同的请求,优先级越高的服务器对应的超时时间比优先级低的服务器对应的超时时间要小。在本实施例,在接收到用户通过客户端发送的第一请求后,首先对该第一请求的报文进行备份得到副本,并在获取了服务器集群中各服务器的优先级排序后将该第一请求发送到优先级最高的第一服务器,然后进一步判断是否在预设的第一超时时间内接收到第一服务器返回的与该第一请求对应的第一反馈结果,如果接收到了该第一反馈结果,则会将该第一反馈结果发送给客户端,以完成对该第一请求的处理,通过调用优先级最高的第一服务器对该第一请求进行处理,有利于提高对于该第一请求的处理效率,提高用户的使用体验。另外,如果在上述第一超时时间内没有接收到该第一反馈结果,则会根据上述优先级排序,依次遍历除第一服务器之外的其它服务器,每遍历到一个服务器时,在将所述副本发送至当前遍历到的服务器后,判断是否在预设的第二超时时间内接收到所述当前遍历到的服务器返回的与所述副本对应的第二反馈结果,若在第二超时时间内接收到该第二反馈结果,则停止遍历,并将第二反馈结果发送给客户端。中调用第二服务器对该副本进行

处理。当第一服务器无法在第一超时时间内返回第一反馈结果时,可认为第一服务器当前处于异常状态,且无法对当前的第一请求进行处理,此时根据各服务器的优先级排序将通过第一请求进行备份得到的副本转发到除第一服务器之外的其它的服务器进行处理,从而能够在第一服务器异常的情况下快速地调用其它的服务器对该第一请求进行处理,从而避免了向用户返回请求处理失败的结果,保证了对用户发出的请求处理不会产生异常超时,保障了用户的使用体验,另外上述其它的服务器是根据优先级排序对除第一服务器之外的其它的所有服务器进行筛选得出的特殊服务器,且该特殊服务器的优先级相对最高,通过调用该特殊服务器对第一请求进行处理,在保证了对该第一请求的正常处理的情况下,也能有效的保证对该第一请求的处理效率。

[0063] 参照图2,进一步地,本申请一实施例中,上述步骤S3之前,包括:

[0064] S300:获取各所述服务器的运行参数数据;

[0065] S301:根据所述运行参数数据,按照预设规则计算各所述服务器的优先级数值;

[0066] S302:将各所述服务器的优先级数值按照从小到大的顺序进行排序,得到所述优先级排序。

[0067] 如上述步骤S300至S302所述,在对第一请求进行处理之前,为了实现从所有的服务器中选出处理效率最高的一个服务器来对该第一请求进行处理,需要计算出每一个服务器的优先级数值,并根据上述优先级数值来对所有的服务器进行优先级排序。具体地,首先获取每一个服务器的运行参数数据,其中上述运行参数数据包括但不限于CPU使用率、内存占用率、网络使用率,CPU使用率是服务器中运行的程序占用的CPU资源,表示服务器在某个时间点的运行程序的情况,CPU使用率越高,说明该服务器在当前时间运行了较多程序,反之较少,内存占用率是指进程所开销的内存,网络使用率是指网卡带宽的利用率。在获得了每个服务器的运行参数数据后,按照预设规则来计算出每一个服务器的优先级数值,最后将每一个服务器的优先级数值按照从小到大的顺序进行排序,来得到各服务器的优先级排序。其中,上述优先级数值用于表征服务器的负载状态,优先级数值越小表明服务器的负载状态越空闲,则服务器的优先级越高,优先级数值越大表明服务器的负载状态越忙碌,则服务器的优先级越低。另外,对于上述预设规则的具体内容不作限定,例如预设规则是指根据各服务器的运行参数数据以及预设的与各运行参数数据分别对应的权重值,进而来计算各服务器的运行参数和值,即上述的优先级数值。或者预设规则还可以是指将服务器的每一个运行参数数据直接进行求和处理得到该服务器的运行参数总和,即上述的优先级数值。之后,当接收到用户通过客户端输入的第一请求后,便可以根据上述优先级排序来从所有的服务器中确定出用于处理该第一请求的服务器,并将该第一请求转发到该确定的服务器,以使得该服务器对该第一请求进行处理。在本实施例,根据各所有服务器的优先级排序来从所有的服务器中确定出用于处理该第一请求的服务器,从而实现将第一请求分发到负载较小的服务器上进行处理,提高了对第一请求的处理效率。

[0068] 进一步地,本申请一实施例中,上述步骤S302,包括:

[0069] S3020:设置与各所述运行参数数据分别对应的权重值;

[0070] S3021:根据各所述运行参数数据与各所述权重值计算各所述服务器的运行参数和值;

[0071] S3022:将所述运行参数和值确定为所述优先级数值。

[0072] 如上述步骤S3020至S3022所述,上述根据所述运行参数数据,按照预设规则计算各所述服务器的优先级数值的步骤具体包括:首先对每个运行参数设置对应的权重值,然后根据各上述运行参数数据与各上述权重值计算出每个服务器的运行参数和值,并将该运行参数和值确定为优先级数值,以使得通过将计算得出的各服务器的优先级数值按照从小到大的顺序进行排序,得到上述优先级排序。其中,上述与运行参数对应的权重值用于标识每个运行参数在衡量服务器的当前负载状态时所占的比重。运行参数的权重值越大,说明该运行参数所占比重越大。同理,运行指标的预设权重越小,说明该运行参数所占比重越小。例如当服务器的运行参数包括CPU使用率、内存占用率、及网络使用率时,可为CPU使用率设置预设权重为50%,为内存占用率设置预设权重为30%,为网络使用率设置预设权重为20%。可以理解地,各个运行参数对应的权重值的具体数值不作限定,可以由根据具体应用环境进行变更设置。另外,上述运行参数和值,即上述优先级数值用于标识服务器的运行状况,上述优先级数值与服务器的优先级呈负相关关系,且运行参数和值越小表示该服务器当前处理的访问请求越少,即该服务器处于相对空闲的工作状态,该服务器的优先级越高;运行参数和值越大表示该服务器当前处理的访问请求越多,即该物理服务器处于相对忙碌的工作状态,该服务器的优先级越低。举例地,服务器A当前的运行参数数据中的CPU使用率为20%、内存占用率为30%及网络使用率为30%,则服务器A的运行参数和值,即优先级数值为 $M=20\% \times 50\%+30\% \times 30\%+30\% \times 20\%=0.25$ 。此时,服务器B当前的运行参数数据中的CPU使用率为50%、内存占用率为40%及网络使用率为50%,则物理服务器B的运行参数和值,即优先级数值为 $M=50\% \times 50\%+40\% \times 30\%+50\% \times 20\%=0.47$ 。服务器C当前的运行参数数据中的CPU使用率为80%、内存占用率为80%、及网络使用率为70%,则物理服务器C的运行参数和值,即优先级数值为 $M=80\% \times 50\%+70\% \times 30\%+70\% \times 20\%=0.75$,因此服务器A、B、C的按照优先级数值从小到大的顺序进行排序得到的优先级排序为A、B、C。

[0073] 进一步地,本申请一实施例中,当前遍历到的服务器是在所述优先级排序中排序最末的服务器,上述步骤S5之后,包括:

[0074] S500:若在所述第二超时时间内没有接收到所述第二反馈结果,则向所述客户端返回请求处理失败信息;

[0075] S501:为所有所述服务器添加异常标记;

[0076] S502:向维修人员发出维修信息,以便所述维修人员对所有所述服务器进行维修处理。

[0077] 如上述步骤S500至S502所述,假如当前遍历到的服务器是在所述优先级排序中排序最末的服务器,且在第二超时时间内没有接收到所述第二反馈结果,则表明在遍历完上述所有的其它服务器之后,都不存在能够在对应的超时时间内返回与第一请求对应的反馈信息的服务器,即上述所有的服务器都处于异常状态,因此当前无法对用户发出的第一请求进行处理,进而会向用户的客户端返回请求处理失败信息,向用户表明当前所有的服务器都无法对第一请求进行处理,由于所有的服务器暂时无法提供服务,则可发出故障提醒信息来提醒用户不用再等待,同时会为所有的服务器加上异常标记,以表明所有的服务器均处于异常的状态,并向维修人员发出维修信息,以使得维修人员对所有的服务器进行维修,具体的发出维修信息的形式可为邮件形式或者发出报警信号。进一步地,在维护人员

对所有的服务器进行维修检查,并完全修复好各服务器后,维护人员会返回维修完成信息,在接收到用户上述维修完成信息后,便会解除所有服务器的上述异常标记,从而可通过维修完成的服务器继续为用户提供服务。

[0078] 进一步地,本申请一实施例中,上述步骤S6之后,包括:

[0079] S600:筛选出没有返回与所述第一请求对应的反馈结果的指定服务器,并为所述指定服务器添加异常标记,其中所述指定服务器的数量为一个或多个;

[0080] S601:向各所述指定服务器发送心跳信号;

[0081] S602:若存在在所述第一预设时间内返回了心跳回复信息的第一指定服务器,则对所述第一指定服务器解除所述异常标记。

[0082] 如上述步骤S600至S602所述,上述指定服务器是指没有返回与所述第一请求对应的反馈结果的服务器,对于具有异常标记的指定服务器,这些指定服务器可能是出现了宕机情况,或者只是出现了短暂的异常,例如由于处于繁忙状态而无法在对应的超时时间内返回与第一请求对应的反馈结果。在本实施例,通过分别向每一个指定服务器发送心跳信号来检测每一个指定服务器是否处于正常工作状态还是处于异常状态。具体地,预先设置一个心跳响应超时时间,即上述的第一预设时间,例如可设为5s,如果某一个指定服务器在接收到心跳信号的5s内没有向系统返回心跳回复信息,则表明该某一个指定服务器当前处于异常状态;而如果该某一个指定服务器在接收到心跳信号的5s内成功向系统返回了心跳回复信息,则表明该某一个指定服务器当前处于正常工作的状态,则会对该某一个指定服务器解除异常标记。在本实施例,通过向具有异常标记的指定服务器发送心跳信号,来检测指定服务器是否真的处于异常状态,避免出现将只是处于暂时异常的服务器误认为出现了宕机而不能工作的情况,并且可筛选出在第一预设时间内返回了心跳回复信息的第一指定服务器,并对该第一指定服务器解除上述异常标记,从而在接收到新的请求时,可以调用解除了异常标记的第一指定服务器进行处理,有利于从所有的没有异常标记的服务器中选择优先级更高的服务器对新的请求进行处理,保证了对于新的请求的处理效率。

[0083] 进一步地,本申请一实施例中,上述步骤S601之后,包括:

[0084] S6010:若存在在所述第一预设时间内没有返回心跳回复信息的第二指定服务器,则按照预设的时间间隔向所述第二指定服务器发送心跳信息;

[0085] S6011:记录所述第二指定服务器没有返回心跳回复信息的次数;

[0086] S6012:判断所述次数是否大于预设的次数阈值;

[0087] S6013:若是,向维修人员发出维修信息,以便所述维修人员对所述第二指定服务器进行维修处理。

[0088] 如上述步骤S6010和S6013所述,上述第二指定服务器为在第一预设时间内没有返回心跳回复信息的服务器,如果在预设的时间间隔向第二指定服务器发送了多次心跳信息,而第二指定服务器都没有在上述第一预设时间内对该多次心跳信息进行对应的心跳回复信息的回复,则判定该第二指定服务器确实处于异常状态。具体地,预先设置一个时间间隔,例如每隔一小时一次,以及预先设置一个次数阈值,例如可设为3次,如果第二指定服务器超过3次没有在第一预设时间内返回对应于心跳信息的心跳回复信息,则可以确定该第二指定服务器处于异常状态,例如出现了机器故障或者宕机的状况,此时通过指定形式向维修人员发出维修信息,以使得维修人员对该第二指定服务器检查后进行修复或更换处

理,上述指定形式可为邮件形式、短信形式或者报警语音形式,等等,在此不作具体的限定。在本实施例,通过向具有异常标记的第二指定服务器定时发送心跳信号,进而获得第二指定服务器的真实工作状态,若第二指定服务器没有返回心跳回复信息的次数大于上述次数阈值,则判定该第二指定服务器处于异常状态,并可及时地通知维修人员对第二指定服务器进行维修,以避免由于不对出现异常的第二指定服务器进行及时处理而导致更大的损伤,保证整个服务器系统的正常运作。

[0089] 进一步地,本申请一实施例中,上述步骤S6之后,包括:

[0090] S610:判断是否接收到新的第二请求;

[0091] S611:若是,根据各所述服务器的标记状态,从所有所述服务器中筛选出没有添加所述异常标记的第三指定服务器;

[0092] S612:根据所述优先级排序,从所述第三指定服务器中确定出优先级最高的第二服务器;

[0093] S613:将所述第二请求发送至所述第二服务器,以便所述第二服务器对所述第二请求进行处理。

[0094] 如上述步骤S610至S613所述,当接收到新的第二请求时,首先获取每一个服务器的标记状态,具有异常标记的服务器的状态对应为异常状态,不具有异常标记的服务器的状态对应为正常状态,然后根据上述标记状态从所有服务器中筛选出没有添加异常标记的第三指定服务器,其中上述第三指定服务器为处于正常状态的服务器,只有处于正常状态的服务器才有对该新的第二请求进行处理的权限,即只会从处于正常状态的服务器中挑选出对该第二请求进行处理的服务器。然后根据上述优先级排序,即根据处于正常状态的每一个第三指定服务器在优先级排序中的排序位置信息,从所有的第三指定服务器中选择优先级最高的第二服务器,并将上述第二请求发送到该第二服务器,以通过该第二服务器进行处理并返回对应新的第二请求的第三反馈结果。在本实施例,根据优先级排序,并只从处于正常状态(即没有添加异常标记)的所有服务器中选择优先级最高的第二服务器来对新的第二请求进行处理,避免出现选择了处于异常状态(即添加有异常标记)的服务器来对新的请求进行处理而得不到反馈结果的情况,保证了对新的第二请求的有效处理,另外在所有的处于正常状态的服务器中选择优先级最高的第二服务器来对新接收的第二请求进行处理,有利于提高对于该新的第二请求的处理效率,提高用户的使用体验。

[0095] 参照图3,本申请一实施例中还提供了一种请求处理装置,包括:

[0096] 接收模块1,用于接收客户端发出的第一请求;

[0097] 备份模块2,用于对所述第一请求对应的报文进行备份得到副本;

[0098] 第一发送模块3,用于获取服务器集群中各服务器的优先级排序,将所述第一请求发送至优先级最高的第一服务器,以便所述第一服务器对所述第一请求进行处理,其中所述优先级排序为按照各所述服务器的优先级数值从小到大的顺序进行排序得到;

[0099] 第一判断模块4,用于判断在预设的第一超时时间内是否接收到所述第一服务器返回的与所述第一请求对应的第一反馈结果;

[0100] 遍历模块5,用于若在所述第一超时时间内没有接收到所述第一反馈结果,根据所述优先级排序,依次遍历除所述第一服务器外的其它服务器,每遍历到一个服务器时,在将所述副本发送至当前遍历到的服务器后,判断是否在预设的第二超时时间内接收到所述当

前遍历到的服务器返回的与所述副本对应的第二反馈结果；

[0101] 第二发送模块6,用于若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端。

[0102] 本实施例中,本实施例的执行主体为请求处理装置,具体可为请求分发服务器,上述的客户端可为手机,PDA(Personal Digital Assistant,个人数字助理)、膝上型计算机、掌上电脑等等,用户可通过该客户端来发送请求,来获取服务器的对于该请求的业务服务。上述优先级排序是指根据每个服务器的优先级数值进行排序得到的,优先级越高的服务器对应越空闲的工作状态,且处理请求的效率和速率越快,其中上述优先级数值可通过对服务器集群中各服务器的运行参数数据进行计算得到。上述的超时时间则是预先根据各服务器的优先级排序进行设置的,不同优先级的服务器可对应不同的超时时间,例如对于一个相同的请求,优先级越高的服务器对应的超时时间比优先级低的服务器对应的超时时间要小。在本实施例,在接收到用户通过客户端发送的第一请求后,首先对该第一请求的报文进行备份得到副本,并在获取了服务器集群中各服务器的优先级排序后将该第一请求发送到优先级最高的第一服务器,然后进一步判断是否在预设的第一超时时间内接收到第一服务器返回的与该第一请求对应的第一反馈结果,如果接收到了该第一反馈结果,则会将该第一反馈结果发送给客户端,以完成对该第一请求的处理,通过调用优先级最高的第一服务器对该第一请求进行处理,有利于提高对于该第一请求的处理效率,提高用户的使用体验。另外,如果在上述第一超时时间内没有接收到该第一反馈结果,则会根据上述优先级排序,依次遍历除第一服务器之外的其它服务器,每遍历到一个服务器时,在将所述副本发送至当前遍历到的服务器后,判断是否在预设的第二超时时间内接收到所述当前遍历到的服务器返回的与所述副本对应的第二反馈结果,若在第二超时时间内接收到该第二反馈结果,则停止遍历,并将第二反馈结果发送给客户端。中调用第二服务器对该副本进行处理。当第一服务器无法在第一超时时间内返回第一反馈结果时,可认为第一服务器当前处于异常状态,且无法对当前的第一请求进行处理,此时根据各服务器的优先级排序将通过对第一请求进行备份得到的副本转发到除第一服务器之外的其它的服务器进行处理,从而能够在第一服务器异常的情况下快速地调用其它的服务器对该第一请求进行处理,从而避免了向用户返回请求处理失败的结果,保证了对用户发出的请求处理不会产生异常超时,保障了用户的使用体验,另外上述其它的服务器是根据优先级排序对除第一服务器之外的其它的所有服务器进行筛选得出的特殊服务器,且该特殊服务器的优先级相对最高,通过调用该特殊服务器对第一请求进行处理,在保证了对该第一请求的正常处理的情况下,也能有效的保证对该第一请求的处理效率。

[0103] 进一步地,本申请一实施例中,上述请求处理装置,包括:

[0104] 获取模块,用于获取各所述服务器的运行参数数据;

[0105] 计算模块,用于根据所述运行参数数据,按照预设规则计算各所述服务器的优先级数值;

[0106] 排序模块,用于将各所述服务器的优先级数值按照从小到大的顺序进行排序,得到所述优先级排序。

[0107] 本实施例中,在对第一请求进行处理之前,为了实现从所有的服务器中选出处理效率最高的一个服务器来对该第一请求进行处理,需要计算出每一个服务器的优先级数

值,并根据上述优先级数值来对所有的服务器进行优先级排序。具体地,首先获取每一个服务器的运行参数数据,其中上述运行参数数据包括但不限于CPU使用率、内存占用率、网络使用率,CPU使用率是服务器中运行的程序占用的CPU资源,表示服务器在某个时间点的运行程序的情况,CPU使用率越高,说明该服务器在当前时间运行了较多程序,反之较少,内存占用率是指进程所开销的内存,网络使用率是指网卡带宽的利用率。在获得了每个服务器的运行参数数据后,按照预设规则来计算出每一个服务器的优先级数值,最后将每一个服务器的优先级数值按照从小到大的顺序进行排序,来得到各服务器的优先级排序。其中,上述优先级数值用于表征服务器的负载状态,优先级数值越小表明服务器的负载状态越空闲,则服务器的优先级越高,优先级数值越大表明服务器的负载状态越忙碌,则服务器的优先级越低。另外,对于上述预设规则的具体内容不作限定,例如预设规则是指根据各服务器的运行参数数据以及预设的与各运行参数数据分别对应的权重值,进而计算出各服务器的运行参数和值,即上述的优先级数值。或者预设规则还可以是指将服务器的每一个运行参数数据直接进行求和处理得到该服务器的运行参数总和,即上述的优先级数值。之后,当接收到用户通过客户端输入的第一请求后,便可以根据上述优先级排序来从所有的服务器中确定出用于处理该第一请求的服务器,并将该第一请求转发到该确定的服务器,以使得该服务器对该第一请求进行处理。在本实施例,根据各所有服务器的优先级排序来从所有的服务器中确定出用于处理该第一请求的服务器,从而实现将第一请求分发到负载较小的服务器上进行处理,提高了对第一请求的处理效率。

[0108] 进一步地,本申请一实施例中,上述计算模块,包括:

[0109] 设置单元,用于根据设置与各运行参数分别对应的权重值;

[0110] 计算单元,用于根据各所述运行参数数据与各所述权重值计算各所述服务器的运行参数和值;

[0111] 确定单元,用于将所述运行参数和值确定为所述优先级数值。

[0112] 本实施例中,上述根据所述运行参数数据,按照预设规则计算各所述服务器的优先级数值的步骤具体包括:首先对每个运行参数设置对应的权重值,然后根据各上述运行参数数据与各上述权重值计算出每个服务器的运行参数和值,并将该运行参数和值确定为优先级数值,以使得通过将计算得出的各服务器的优先级数值按照从小到大的顺序进行排序,得到上述优先级排序。其中,上述与运行参数对应的权重值用于标识每个运行参数在衡量服务器的当前负载状态时所占的比重。运行参数的权重值越大,说明该运行参数所占比重越大。同理,运行指标的预设权重越小,说明该运行参数所占比重越小。例如当服务器的运行参数包括CPU使用率、内存占用率、及网络使用率时,可为CPU使用率设置预设权重为50%,为内存占用率设置预设权重为30%,为网络使用率设置预设权重为20%。可以理解地,各个运行参数对应的权重值的具体数值不作限定,可以由根据具体应用环境进行变更设置。另外,上述运行参数和值,即上述优先级数值用于标识服务器的运行状况,上述优先级数值与服务器的优先级呈负相关关系,且运行参数和值越小表示该服务器当前处理的访问请求越少,即该服务器处于相对空闲的工作状态,该服务器的优先级越高;运行参数和值越大表示该服务器当前处理的访问请求越多,即该物理服务器处于相对忙碌的工作状态,该服务器的优先级越低。举例地,服务器A当前的运行参数数据中的CPU使用率为20%、内存占用率为30%及网络使用率为30%,则服务器A的运行参数和值,即优先级数值为 $M=20\%$

$\times 50\% + 30\% \times 30\% + 30\% \times 20\% = 0.25$ 。此时,服务器B当前的运行参数数据中的CPU使用率为50%、内存占用率为40%及网络使用率为50%,则物理服务器B的运行参数和值,即优先级数值为 $M = 50\% \times 50\% + 40\% \times 30\% + 50\% \times 20\% = 0.47$ 。服务器C当前的运行参数数据中的CPU使用率为80%、内存占用率为80%、及网络使用率为70%,则物理服务器C的运行参数和值,即优先级数值为 $M = 80\% \times 50\% + 70\% \times 30\% + 70\% \times 20\% = 0.75$,因此服务器A、B、C的按照优先级数值从小到大的顺序进行排序得到的优先级排序为A、B、C。

[0113] 进一步地,本申请一实施例中,当前遍历到的服务器是在所述优先级排序中排序最末的服务器,上述请求处理装置,还包括:

[0114] 返回模块,用于若在所述第二超时时间内没有接收到所述第二反馈结果,则向所述客户端返回请求处理失败信息;

[0115] 添加模块,用于为所有所述服务器添加异常标记;

[0116] 第一维修模块,用于向维修人员发出维修信息,以便所述维修人员对所有所述服务器进行维修处理。

[0117] 本实施例中,假如当前遍历到的服务器是在所述优先级排序中排序最末的服务器,且在第二超时时间内没有接收到所述第二反馈结果,则表明在遍历完上述所有的其它服务器之后,都不存在能够在对应的超时时间内返回与第一请求对应的反馈信息的服务器,即上述所有的服务器都处于异常状态,因此当前无法对用户发出的第一请求进行处理,进而会向用户的客户端返回请求处理失败信息,向用户表明当前所有的服务器都无法对第一请求进行处理,由于所有的服务器暂时无法提供服务,则可发出故障提醒信息来提醒用户不用再等待,同时会为所有的服务器加上异常标记,以表明所有的服务器均处于异常的状态,并会向维修人员发出维修信息,以使得维修人员对所有的服务器进行维修,具体的发出维修信息的形式可为邮件形式或者发出报警信号。进一步地,在维护人员对所有的服务器进行维修检查,并完全修复好各服务器后,维护人员会返回维修完成信息,在接收到用户上述维修完成信息后,便会解除所有服务器的上述异常标记,从而可通过维修完成的服务器继续为用户提供服务。

[0118] 进一步地,本申请一实施例中,上述请求处理装置,还包括:

[0119] 第一筛选模块,用于筛选出没有返回与所述第一请求对应的反馈结果的指定服务器,并为所述指定服务器添加异常标记,其中所述指定服务器的数量为一个或多个;

[0120] 第三发送模块,用于向各所述指定服务器发送心跳信号;

[0121] 解除模块,用于若存在在所述第一预设时间内返回了心跳回复信息的第一指定服务器,则对所述第一指定服务器解除所述异常标记。

[0122] 本实施例中,上述指定服务器是指没有返回与所述第一请求对应的反馈结果的服务器,对于具有异常标记的指定服务器,这些指定服务器可能是出现了宕机情况,或者只是出现了短暂的异常,例如由于处于繁忙状态而无法在对应的超时时间内返回与第一请求对应的反馈结果。在本实施例,通过分别向每一个指定服务器发送心跳信号来检测每一个指定服务器是否处于正常工作状态还是处于异常状态。具体地,预先设置一个心跳响应超时时间,即上述的第一预设时间,例如可设为5s,如果某一个指定服务器在接收到心跳信号的5s内没有向系统返回心跳回复信息,则表明该某一个指定服务器当前处于异常状态;而如果该某一个指定服务器在接收到心跳信号的5s内成功向系统返回了心跳回复信息,则表明

该某一个指定服务器当前处于正常工作的状态,则会对该某一个指定服务器解除异常标记。在本实施例,通过向具有异常标记的指定服务器发送心跳信号,来检测指定服务器是否真的处于异常状态,避免出现将只是处于暂时异常的服务器误认为出现了宕机而不能工作的情况,并且可筛选出在第一预设时间内返回了心跳回复信息的第一指定服务器,并对该第一指定服务器解除上述异常标记,从而在接收到新的请求时,可以调用解除了异常标记的第一指定服务器进行处理,有利于从所有的没有异常标记的服务器中选择优先级更高的服务器对新的请求进行处理,保证了对于新的请求的处理效率。

[0123] 进一步地,本申请一实施例中,上述请求处理装置,还包括:

[0124] 第四发送模块,用于若存在在所述第一预设时间内没有返回心跳回复信息的第二指定服务器,则按照预设的时间间隔向所述第二指定服务器发送心跳信息;

[0125] 记录模块,用于记录所述第二指定服务器没有返回心跳回复信息的次数;

[0126] 第二判断模块,用于判断所述次数是否大于预设的次数阈值;

[0127] 第二维修模块,用于若是,向维修人员发出维修信息,以便所述维修人员对所述第二指定服务器进行维修处理。

[0128] 本实施例中,上述第二指定服务器为在第一预设时间内没有返回心跳回复信息的服务器,如果在预设的时间间隔向第二指定服务器发送了多次心跳信息,而第二指定服务器都没有在上述第一预设时间内对该多次心跳信息进行对应的心跳回复信息的回复,则判定该第二指定服务器确实处于异常状态。具体地,预先设置一个时间间隔,例如每隔一小时一次,以及预先设置一个次数阈值,例如可设为3次,如果第二指定服务器超过3次没有在第一预设时间内返回对应于心跳信息的心跳回复信息,则可以确定该第二指定服务器处于异常状态,例如出现了机器故障或者宕机的状况,此时通过指定形式向维修人员发出维修信息,以使得维修人员对该第二指定服务器检查后进行修复或更换处理,上述指定形式可为邮件形式、短信形式或者报警语音形式,等等,在此不作具体的限定。在本实施例,通过向具有异常标记的第二指定服务器定时发送心跳信号,进而获得第二指定服务器的真实工作状态,若第二指定服务器没有返回心跳回复信息的次数大于上述次数阈值,则判定该第二指定服务器处于异常状态,并可及时地通知维修人员对第二指定服务器进行维修,以避免由于不对出现异常的第二指定服务器进行及时处理而导致更大的损伤,保证整个服务器系统的正常运作。

[0129] 进一步地,本申请一实施例中,上述请求处理装置,还包括:

[0130] 第三判断模块,用于判断是否接收到新的第二请求;

[0131] 第二筛选模块,用于若是,根据各所述服务器的标记状态,从所有所述服务器中筛选出没有添加所述异常标记的第三指定服务器;

[0132] 确定模块,用于根据所述优先级排序,从所述第三指定服务器中确定出优先级最高的第二服务器;

[0133] 第五发送模块,用于将所述第二请求发送至所述第二服务器,以便所述第二服务器对所述第二请求进行处理。

[0134] 本实施例中,当接收到新的第二请求时,首先获取每一个服务器的标记状态,具有异常标记的服务器的状态对应为异常状态,不具有异常标记的服务器的状态对应为正常状态,然后根据上述标记状态从所有服务器中筛选出没有添加异常标记的第三指定服务器,

其中上述第三指定服务器为处于正常状态的服务器,只有处于正常状态的服务器才有对该新的第二请求进行处理的权限,即只会从处于正常状态的服务器中挑选出对该第二请求进行处理的服务器。然后根据上述优先级排序,即根据处于正常状态的每一个第三指定服务器在优先级排序中的排序位置信息,从所有的第三指定服务器中选择优先级最高的第二服务器,并将上述第二请求发送到该第二服务器,以通过该第二服务器进行处理并返回对应新的第二请求的第三反馈结果。在本实施例,根据优先级排序,并只从处于正常状态(即没有添加异常标记)的所有服务器中选择优先级最高的第二服务器来对新的第二请求进行处理,避免出现选择了处于异常状态(即添加有异常标记)的服务器来对新的请求进行处理而得不到反馈结果的情况,保证了对新的第二请求的有效处理,另外在所有的处于正常状态的服务器中选择优先级最高的第二服务器来对新接收的第二请求进行处理,有利于提高对于该新的第二请求的处理效率,提高用户的使用体验。

[0135] 参照图4,本申请实施例中还提供一种计算机设备,该计算机设备可以是服务器,其内部结构可以如图4所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中,该计算机设备设计的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储请求以及服务器的优先级排序等数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种请求处理方法。

[0136] 上述处理器执行上述请求处理方法的步骤:

[0137] 接收客户端发出的第一请求;

[0138] 对所述第一请求对应的报文进行备份得到副本;

[0139] 获取服务器集群中各服务器的优先级排序,将所述第一请求发送至优先级最高的第一服务器,以便所述第一服务器对所述第一请求进行处理,其中所述优先级排序为按照各所述服务器的优先级数值从小到大的顺序进行排序得到;

[0140] 判断在预设的第一超时时间内是否接收到所述第一服务器返回的与所述第一请求对应的第一反馈结果;

[0141] 若在所述第一超时时间内接收到所述第一反馈结果,根据所述优先级排序,依次遍历除所述第一服务器外的其它服务器,每遍历到一个服务器时,在将所述副本发送至当前遍历到的服务器后,判断是否在预设的第二超时时间内接收到所述当前遍历到的服务器返回的与所述副本对应的第二反馈结果;

[0142] 若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端。

[0143] 在一实施例,所述获取多个服务器的优先级排序,将所述第一请求发送至优先级最高的第一服务器的步骤之前,包括:

[0144] 获取各所述服务器的运行参数数据;

[0145] 根据所述运行参数数据,按照预设规则计算各所述服务器的优先级数值;

[0146] 将各所述服务器的优先级数值按照从小到大的顺序进行排序,得到所述优先级排序。

[0147] 在一实施例,所述根据所述运行参数数据,按照预设规则计算各所述服务器的优先级数值的步骤,包括:

[0148] 设置与各运行参数分别对应的权重值;

[0149] 根据各所述运行参数数据与各所述权重值计算各所述服务器的运行参数和值;

[0150] 将所述运行参数和值确定为所述优先级数值。

[0151] 在一实施例,当前遍历到的服务器是在所述优先级排序中排序最末的服务器,所述判断是否在预设的第二超时时间内接收到所述当前遍历到的服务器返回的与所述副本对应的第二反馈结果的步骤之后,包括:

[0152] 若在所述第二超时时间内没有接收到所述第二反馈结果,则向所述客户端返回请求处理失败信息;

[0153] 为所有所述服务器添加异常标记;

[0154] 向维修人员发出维修信息,以便所述维修人员对所有所述服务器进行维修处理。

[0155] 在一实施例,所述若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端的步骤之后,包括:

[0156] 筛选出没有返回与所述第一请求对应的反馈结果的指定服务器,并为所述指定服务器添加异常标记,其中所述指定服务器的数量为一个或多个;

[0157] 向各所述指定服务器发送心跳信号;

[0158] 若存在在所述第一预设时间内返回了心跳回复信息的第一指定服务器,则对所述第一指定服务器解除所述异常标记。

[0159] 在一实施例,所述向各所述指定服务器发送心跳信号的步骤之后,包括:

[0160] 若存在在所述第一预设时间内没有返回心跳回复信息的第二指定服务器,则按照预设的时间间隔向所述第二指定服务器发送心跳信息;

[0161] 记录所述第二指定服务器没有返回心跳回复信息的次数;

[0162] 判断所述次数是否大于预设的次数阈值;

[0163] 若是,向维修人员发出维修信息,以便所述维修人员对所述第二指定服务器进行维修处理。

[0164] 在一实施例,所述若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端的步骤之后,包括:

[0165] 判断是否接收到新的第二请求;

[0166] 若是,根据各所述服务器的标记状态,从所有所述服务器中筛选出没有添加所述异常标记的第三指定服务器;

[0167] 根据所述优先级排序,从所述第三指定服务器中确定出优先级最高的第二服务器;

[0168] 将所述第二请求发送至所述第二服务器,以便所述第二服务器对所述第二请求进行处理。

[0169] 本领域技术人员可以理解,图4中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的装置、计算机设备的限定。

[0170] 本申请一实施例还提供一种计算机可读存储介质,其上存储有计算机程序,计算机程序被处理器执行时实现一种请求处理方法,具体为:

- [0171] 接收客户端发出的第一请求；
- [0172] 对所述第一请求对应的报文进行备份得到副本；
- [0173] 获取服务器集群中各服务器的优先级排序,将所述第一请求发送至优先级最高的第一服务器,以便所述第一服务器对所述第一请求进行处理,其中所述优先级排序为按照各所述服务器的优先级数值从小到大的顺序进行排序得到；
- [0174] 判断在预设的第一超时时间内是否接收到所述第一服务器返回的与所述第一请求对应的第一反馈结果；
- [0175] 若在所述第一超时时间内接收到所述第一反馈结果,根据所述优先级排序,依次遍历除所述第一服务器外的其它服务器,每遍历到一个服务器时,在将所述副本发送至当前遍历到的服务器后,判断是否在预设的第二超时时间内接收到所述当前遍历到的服务器返回的与所述副本对应的第二反馈结果；
- [0176] 若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端。
- [0177] 在一实施例,所述获取多个服务器的优先级排序,将所述第一请求发送至优先级最高的第一服务器的步骤之前,包括:
- [0178] 获取各所述服务器的运行参数数据；
- [0179] 根据所述运行参数数据,按照预设规则计算各所述服务器的优先级数值；
- [0180] 将各所述服务器的优先级数值按照从小到大的顺序进行排序,得到所述优先级排序。
- [0181] 在一实施例,所述根据所述运行参数数据,按照预设规则计算各所述服务器的优先级数值的步骤,包括:
- [0182] 设置与各运行参数分别对应的权重值；
- [0183] 根据各所述运行参数数据与各所述权重值计算各所述服务器的运行参数和值；
- [0184] 将所述运行参数和值确定为所述优先级数值。
- [0185] 在一实施例,当前遍历到的服务器是在所述优先级排序中排序最末的服务器,所述判断是否在预设的第二超时时间内接收到所述当前遍历到的服务器返回的与所述副本对应的第二反馈结果的步骤之后,包括:
- [0186] 若在所述第二超时时间内没有接收到所述第二反馈结果,则向所述客户端返回请求处理失败信息；
- [0187] 为所有所述服务器添加异常标记；
- [0188] 向维修人员发出维修信息,以便所述维修人员对所有所述服务器进行维修处理。
- [0189] 在一实施例,所述若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端的步骤之后,包括:
- [0190] 筛选出没有返回与所述第一请求对应的反馈结果的指定服务器,并为所述指定服务器添加异常标记,其中所述指定服务器的数量为一个或多个；
- [0191] 向各所述指定服务器发送心跳信号；
- [0192] 若存在在所述第一预设时间内返回了心跳回复信息的第一指定服务器,则对所述第一指定服务器解除所述异常标记。
- [0193] 在一实施例,所述向各所述指定服务器发送心跳信号的步骤之后,包括:

[0194] 若存在在所述第一预设时间内没有返回心跳回复信息的第二指定服务器,则按照预设的时间间隔向所述第二指定服务器发送心跳信息;

[0195] 记录所述第二指定服务器没有返回心跳回复信息的次数;

[0196] 判断所述次数是否大于预设的次数阈值;

[0197] 若是,向维修人员发出维修信息,以便所述维修人员对所述第二指定服务器进行维修处理。

[0198] 在一实施例,所述若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端的步骤之后,包括:

[0199] 判断是否接收到新的第二请求;

[0200] 若是,根据各所述服务器的标记状态,从所有所述服务器中筛选出没有添加所述异常标记的第三指定服务器;

[0201] 根据所述优先级排序,从所述第三指定服务器中确定出优先级最高的第二服务器;

[0202] 将所述第二请求发送至所述第二服务器,以便所述第二服务器对所述第二请求进行处理。

[0203] 综上所述,本申请实施例中提供的请求处理方法、装置、计算机设备和存储介质,接收客户端发出的第一请求;对所述第一请求对应的报文进行备份得到副本;获取服务器集群中各服务器的优先级排序,将所述第一请求发送至优先级最高的第一服务器,以便所述第一服务器对所述第一请求进行处理;判断在预设的第一超时时间内是否接收到所述第一服务器返回的与所述第一请求对应的第一反馈结果;若在所述第一超时时间内接收到所述第一反馈结果,根据所述优先级排序,依次遍历除所述第一服务器外的其它服务器,每遍历到一个服务器时,在将所述副本发送至当前遍历到的服务器后,判断是否在预设的第二超时时间内接收到所述当前遍历到的服务器返回的与所述副本对应的第二反馈结果;若在所述第二超时时间内接收到所述第二反馈结果,停止遍历,并将所述第二反馈结果发送给所述客户端。本申请在调用优先级最高的第一服务器对接收到的第一请求进行处理时,如果没有在第一超时时间内没有接收到第一服务器返回的第一反馈结果,则会根据各服务器的优先级排序将通过所述第一请求进行备份得到的副本转发到除第一服务器之外的其它的服务器进行处理,从而能够在第一服务器异常的情况下快速地调用其它的服务器对该第一请求进行处理,从而避免了向用户返回请求处理失败的结果,保证了对用户发出的请求处理不会产生异常超时,保障了用户的使用体验。

[0204] 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,所述的计算机程序可存储于一非易失性计算机可读存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的和实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括非易失性和/或易失性存储器。非易失性存储器可以包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限,RAM通过多种形式可得,诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双速据率SDRAM(SSRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink)DRAM(SLDRAM)、存储器总线(Rambus)直接RAM

(RDRAM)、直接存储器总线动态RAM (DRDRAM)、以及存储器总线动态RAM (RDRAM) 等。

[0205] 需要说明的是,在本文中,术语“包括”、“包含”或者其任何其它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、装置、物品或者方法不仅包括那些要素,而且还包括没有明确列出的其它要素,或者是还包括为这种过程、装置、物品或者方法所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、装置、物品或者方法中还存在另外的相同要素。

[0206] 以上所述仅为本申请的优选实施例,并非因此限制本申请的专利范围,凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其他相关的技术领域,均同理包括在本申请的专利保护范围内。

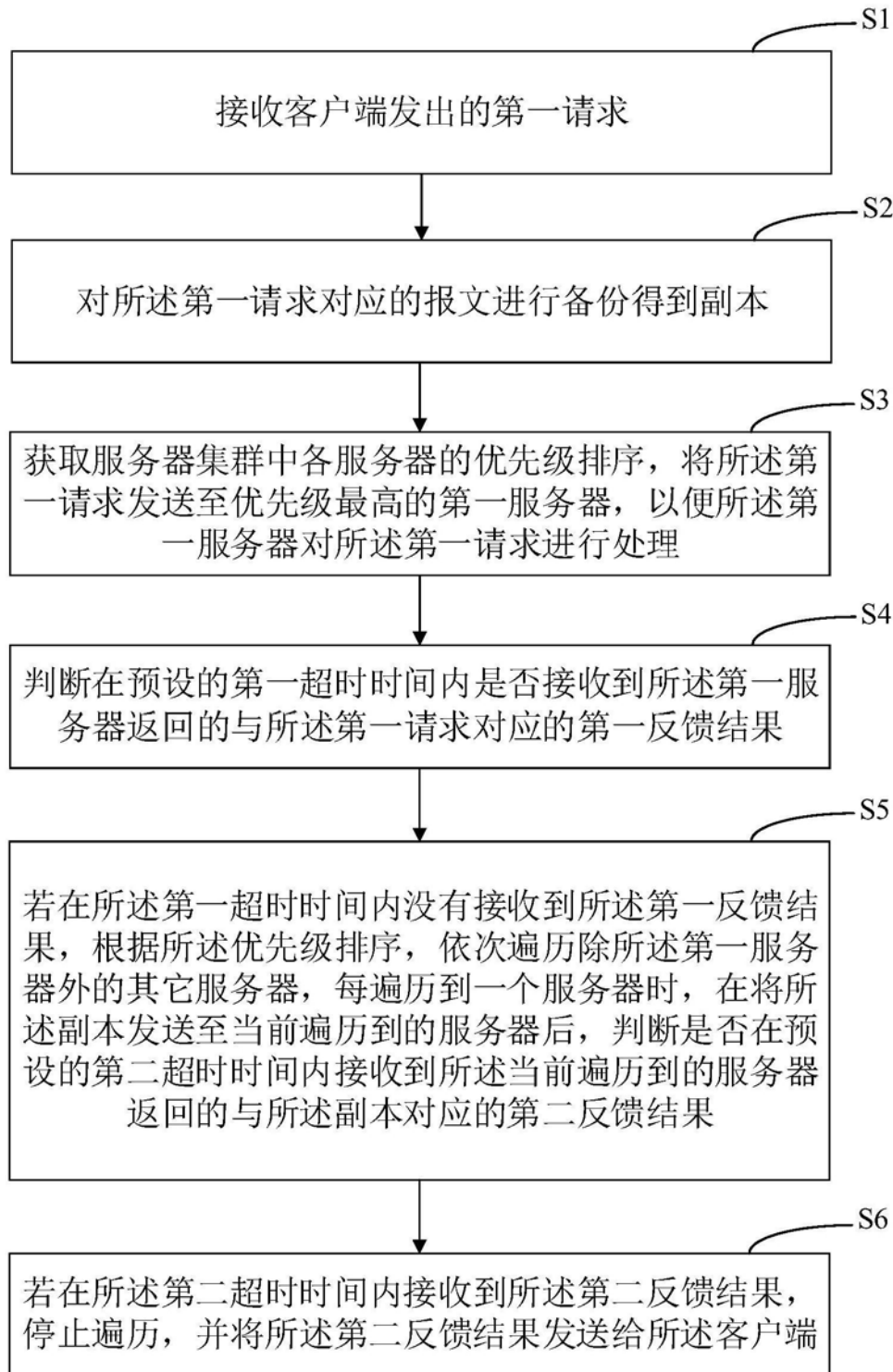


图1

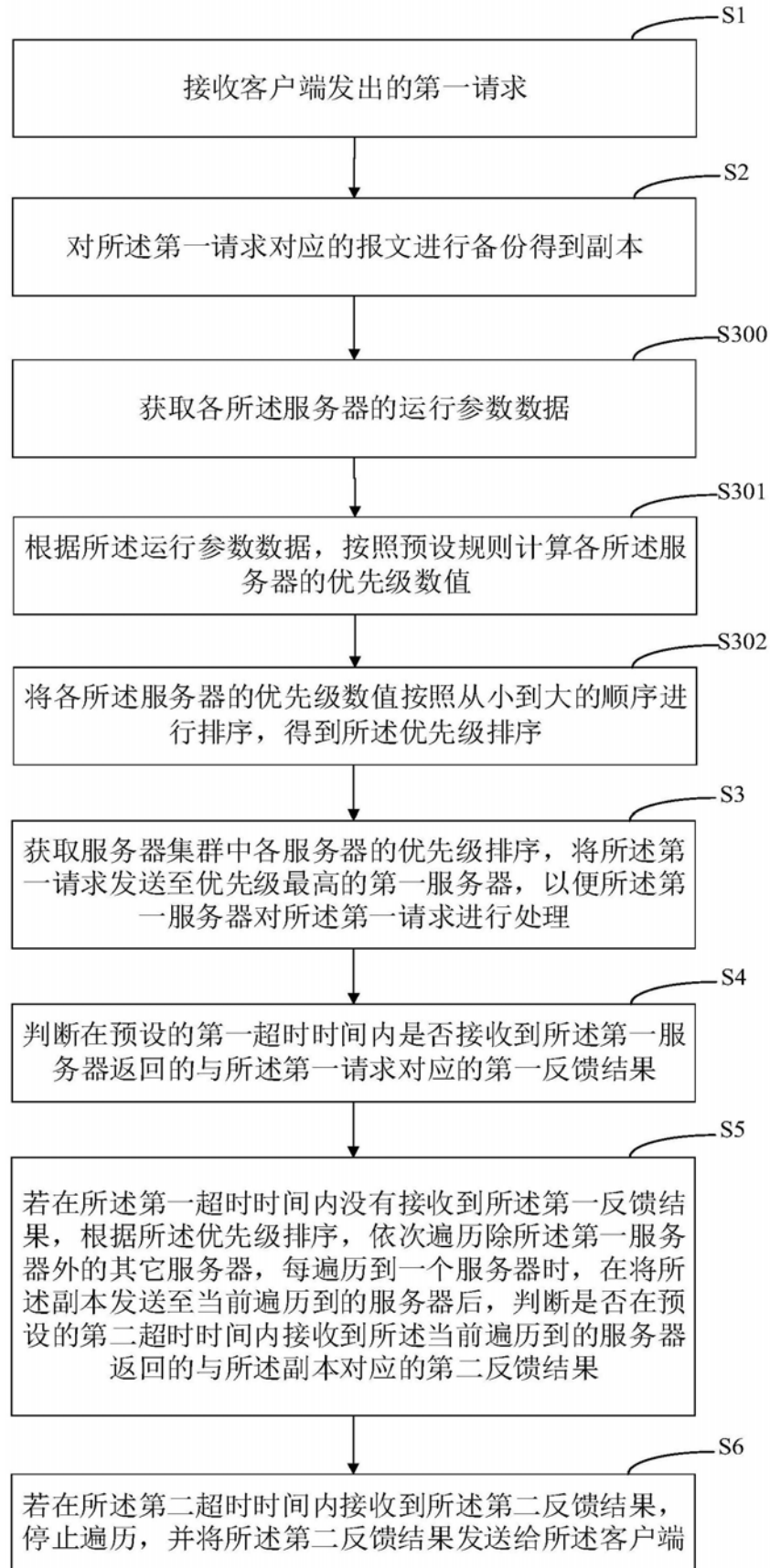


图2

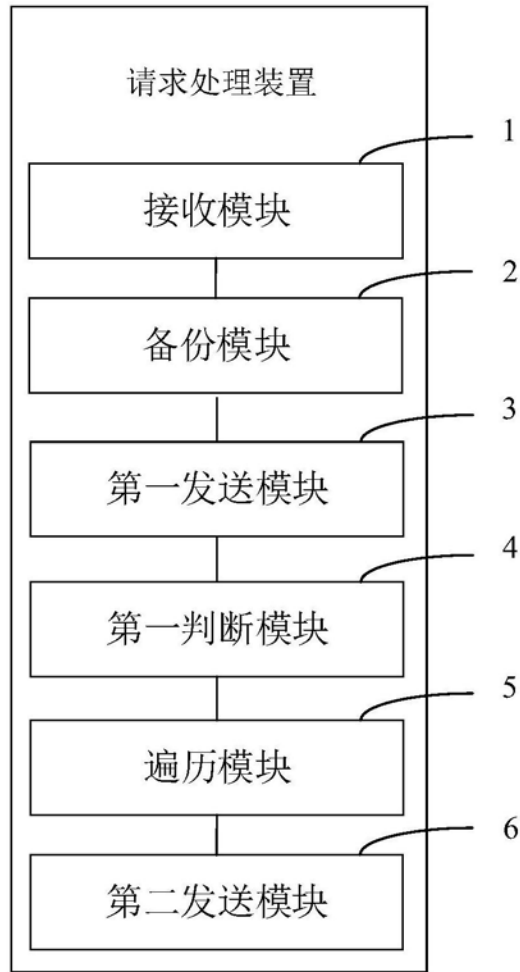


图3

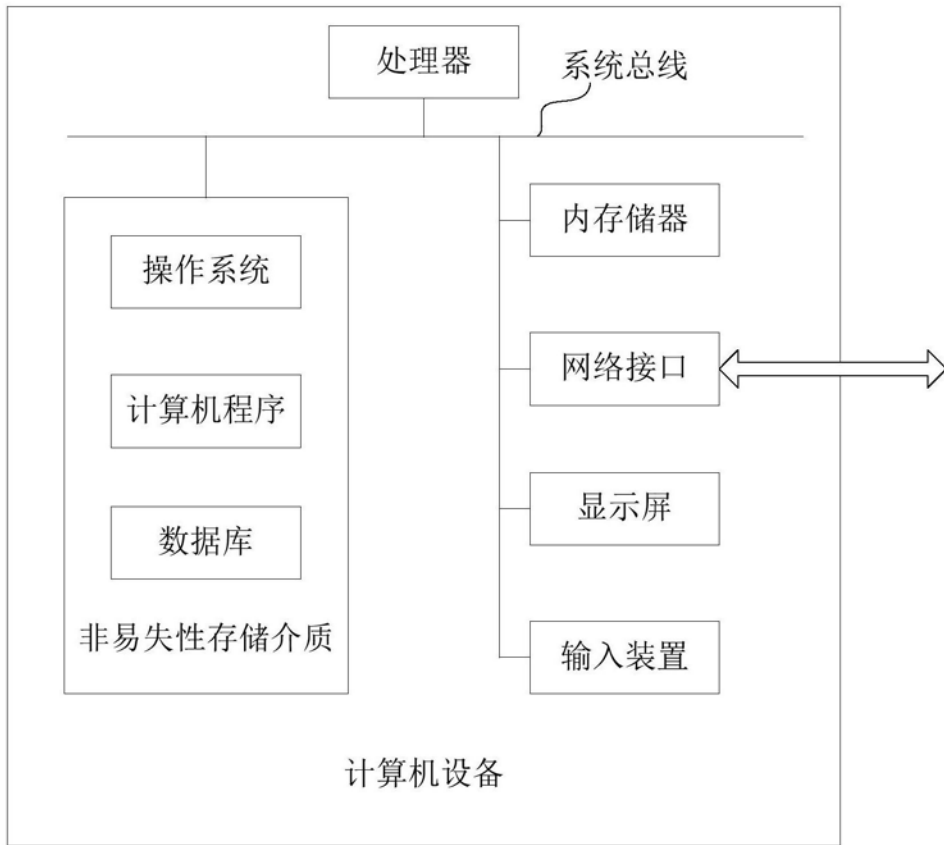


图4