



(12) 发明专利

(10) 授权公告号 CN 111600930 B

(45) 授权公告日 2022.12.09

(21) 申请号 202010272990.X

(22) 申请日 2020.04.09

(65) 同一申请的已公布的文献号
申请公布号 CN 111600930 A

(43) 申请公布日 2020.08.28

(73) 专利权人 网宿科技股份有限公司
地址 200030 上海市徐汇区斜土路2899号
甲光启文化广场A幢5楼

(72) 发明人 庄接滨 谢自强

(74) 专利代理机构 上海晨皓知识产权代理事务
所(普通合伙) 31260
专利代理师 成丽杰

US 2019319885 A1,2019.10.17

US 9716617 B1,2017.07.25

CN 109995713 A,2019.07.09

CN 110651451 A,2020.01.03

CN 110928709 A,2020.03.27

CN 110401696 A,2019.11.01

CN 110601977 A,2019.12.20

CN 108055312 A,2018.05.18

US 2018337851 A1,2018.11.22

US 2019004871 A1,2019.01.03

CN 110737425 A,2020.01.31

CN 107332744 A,2017.11.07

CN 109104368 A,2018.12.28

(51) Int.Cl.
H04L 67/63 (2022.01)
H04L 67/1001 (2022.01)
H04L 67/1004 (2022.01)
H04L 67/1034 (2022.01)

张晶等.微服务框架的设计与实现.《计算机系统应用》.2017,(第06期),
丁乙等.面向军用微服务的服务网格架构研究与设计.《指挥信息系统与技术》.2020,(第01期),

审查员 王宁

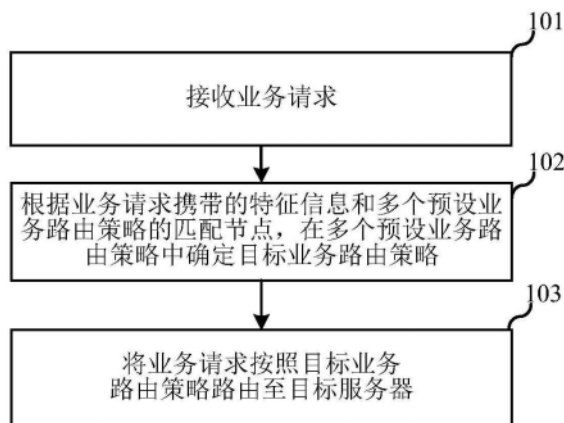
(56) 对比文件
CN 106357435 A,2017.01.25

权利要求书2页 说明书10页 附图5页

(54) 发明名称
微服务请求的流量管理方法、装置、服务器及存储介质

(57) 摘要

本发明实施例涉及微服务领域,公开了一种微服务请求的流量管理方法、装置、服务器及存储介质。本发明包括:代理服务器接收业务请求;根据业务请求携带的特征信息,在所述多个预设业务路由策略中确定目标业务路由策略;其中,预设业务路由策略在配置文件中设置;将业务请求按照目标业务路由策略路由至目标服务器。根据业务请求携带的特征信息确定路由策略,并根据路由策略将其路由至目标服务器,实现在微服务架构下对业务请求的流量的定向管理,使用户可以更加灵活的管理流量。



1. 一种微服务请求的流量管理方法,其特征在于,包括:
 - 代理服务器接收业务请求;
 - 根据所述业务请求携带的特征信息,在多个预设业务路由策略中确定目标业务路由策略;其中,所述预设业务路由策略在配置文件中设置;
 - 将所述业务请求按照所述目标业务路由策略路由至目标服务器;
 - 所述根据所述业务请求携带的特征信息,在多个预设业务路由策略中确定目标业务路由策略,包括:
 - 获取各个预设业务路由策略对应的匹配节点定义的信息;
 - 若预设业务路由策略的匹配节点定义的信息与所述特征信息相同,或所述预设业务路由策略的匹配节点未定义信息,则将所述预设业务路由策略作为目标业务路由策略。
2. 根据权利要求1所述的微服务请求的流量管理方法,其特征在于,所述将所述预设业务路由策略作为目标业务路由策略,包括:
 - 若所述目标业务路由策略的数量大于一,则将所述目标业务路由策略按预设优先级顺序进行排序;
 - 所述将所述业务请求按照所述目标业务路由策略路由至目标服务器,包括:
 - 按所述预设优先级顺序的目标业务路由策略确定对应的目标服务器,直至所确定的目标服务器满足路由条件;
 - 将所述业务请求路由至所述满足路由条件的目标服务器。
3. 根据权利要求1所述的微服务请求的流量管理方法,其特征在于,在确定所述目标业务路由策略之前,还包括:
 - 通过注册中心的网络接口下载配置文件;
 - 解析所述下载的配置文件。
4. 根据权利要求1所述的微服务请求的流量管理方法,其特征在于,所述将所述业务请求按照所述目标业务路由策略路由至目标服务器,包括:
 - 根据所述目标业务路由策略中包括的权重,对所述业务请求路由至所述目标服务器的流量进行管理。
5. 根据权利要求1所述的微服务请求的流量管理方法,其特征在于,所述配置文件中还包括以下任一策略或其组合:负载均衡策略、限流策略或容错策略;
 - 所述将所述业务请求按照所述路由策略路由至目标服务器,包括:
 - 按照所述路由策略,并结合所述负载均衡策略、所述限流策略或所述容错策略中任一策略或其组合确定目标服务器;
 - 将所述业务请求路由至所述目标服务器。
6. 根据权利要求1所述的微服务请求的流量管理方法,其特征在于,所述方法包括:
 - 注册中心读取页面服务设备的数据库中的流量管理数据;其中,所述流量管理数据中至少包括:多个与业务请求携带的特征信息相匹配的预设业务路由策略;
 - 根据所述读取的流量管理数据生成配置文件,以供微服务架构根据所述配置文件对接收的业务请求进行路由。
7. 根据权利要求1所述的微服务请求的流量管理方法,其特征在于,所述方法包括:
 - 页面服务设备设置多个与业务请求携带的特征信息相匹配的预设业务路由策略;

将所述多个预设业务路由策略作为流量管理数据存储于数据库,以供注册中心根据所述流量管理数据生成配置文件。

8. 一种微服务请求的流量管理装置,其特征在于,包括:接收模块,确定模块,路由模块;

所述接收模块用于接收业务请求;

所述确定模块用于根据所述业务请求携带的特征信息,在多个预设业务路由策略中确定目标业务路由策略;其中,所述预设业务路由策略在配置文件中设置;

所述路由模块用于将所述业务请求按照所述目标业务路由策略路由至目标服务器;

确定模块用于获取各个预设业务路由策略对应的匹配节点定义的信息;若预设业务路由策略的匹配节点定义的信息与特征信息相同,或预设业务路由策略的匹配节点未定义信息,则将预设业务路由策略作为目标业务路由策略。

9. 一种服务器,其特征在于,包括:

至少一个处理器;以及,

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行如权利要求1至7中任一所述的微服务请求的流量管理方法。

10. 一种计算机可读存储介质,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中任一项所述的微服务请求的流量管理方法。

微服务请求的流量管理方法、装置、服务器及存储介质

技术领域

[0001] 本发明实施例涉及微服务领域,特别涉及一种微服务请求的流量管理方法、装置、服务器及存储介质。

背景技术

[0002] 微服务架构是一种由多个松散耦合的微服务组成一个完整应用系统的架构,其中,每个微服务实现单一的职责与功能,可以独立开发、测试、部署。随着业务的发展,越来越多的企业采用微服务架构来开发应用系统,而一个复杂的应用系统可能由数十个甚至上百个微服务组成,每个微服务又会为多个用户的业务请求进行服务,这就导致了整个微服务架构下的服务器中会接收到大量的业务请求,如此大量的业务请求如果不加以管理,很容易造成用户的业务请求不能及时处理,导致业务请求的处理效率降低。

[0003] 目前对微服务请求的流量管理,通常采用同机房或同网络优先的策略,将业务请求分发至同机房或同网络的服务器进行处理,减少了业务请求在路由至服务器时耗费的时间,从而加快了业务请求的处理效率。

[0004] 发明人发现相关技术中至少存在如下问题:在通过机房或网络等信息对业务请求进行流量管理时,管理策略较为单一,无法适用于复杂的应用场景,且用户无法根据需求灵活调整流量的流向。

发明内容

[0005] 本发明实施方式的目的在于提供一种微服务请求的流量管理方法、装置、服务器及存储介质,根据业务请求携带的特征信息确定其所路由的目标服务器,实现在微服务构架下对业务请求的流量的定向管理,使用户可以更加灵活的管理流量。

[0006] 为解决上述技术问题,本发明的实施方式提供了一种微服务请求的流量管理方法,包括:代理服务器接收业务请求;根据业务请求携带的特征信息,在多个预设业务路由策略中确定目标业务路由策略;其中,预设业务路由策略在配置文件中设置;将业务请求按照目标业务路由策略路由至目标服务器。

[0007] 本发明的实施方式还提供了一种微服务请求的流量管理装置,包括:接收模块,确定模块,路由模块;接收模块用于接收业务请求;确定模块用于根据业务请求携带的特征信息,在多个预设业务路由策略中确定目标业务路由策略;其中,预设路由策略在配置文件中设置;路由模块用于将业务请求按照目标业务路由策略路由至目标服务器。

[0008] 本发明的实施方式还提供了一种微服务请求的流量管理系统,包括:页面服务设备,注册中心和代理服务器;页面服务设备用于设置多个与业务请求携带的特征信息相匹配的预设业务路由策略,并将多个预设业务路由策略作为流量管理数据存储至数据库中;注册中心用于获取流量管理数据,并根据流量管理数据生成配置文件;代理服务器用于在接收到业务请求之后,根据业务请求携带的特征信息,在注册中心下载的配置文件中的多个预设业务路由策略中确定目标业务路由策略;并将业务请求按照目标业务路由策略路由

至目标服务器。

[0009] 本发明的实施方式还提供了一种服务器,包括:至少一个处理器;以及,与至少一个处理器通信连接的存储器;其中,存储器存储有可被至少一个处理器执行的指令,指令被至少一个处理器执行,以使至少一个处理器能够执行微服务请求的流量管理方法。

[0010] 本发明的实施方式还提供了一种存储介质,存储有计算机程序,计算机程序被处理器执行时上述的微服务请求的流量管理方法。

[0011] 本发明实施方式相对于现有技术而言,在接收到用户的业务请求时,获取业务请求中携带的特征信息,特征信息为某一个业务请求或某一类业务请求的标识信息,通过该特征信息即可确定业务请求的类型,从而区分不同类型的业务请求。在获取业务请求的特征信息之后,获取预先配置的该类型的业务请求的目标业务路由策略,业务路由策略由管理员根据各个类型的业务请求的实际需求进行预先配置,在业务请求的实际需求发生变化时可以对业务路由策略的配置项进行修改,从而可以灵活的实现业务请求的流量的管理。在根据预先配置的配置文件中确定目标业务路由策略之后,根据目标业务路由策略将业务请求路由至目标服务器,由于目标业务路由策略中规定了业务请求所路由的目标服务器,因此实现了对业务请求流量的定向管理,使用户可以灵活的对流量进行管理。

[0012] 另外,根据业务请求携带的特征信息,在多个预设业务路由策略中确定业务路由策略,包括:获取各个预设业务路由策略对应的匹配节点定义的信息;若预设业务路由策略的匹配节点定义的信息与特征信息相同,或预设业务路由策略的匹配节点未定义信息,则将预设业务路由策略作为目标业务路由策略。定义在配置文件中,若某一路由规则未定义匹配节点,则说明该路由规则适用于任一业务请求,所以与特征信息相匹配的匹配节点,不仅包括定义的信息与特征信息相同的匹配节点,还可以包括未定义信息的匹配节点。

[0013] 另外,将预设业务路由策略作为目标业务路由策略,包括:若目标业务路由策略的数量大于一,则将目标业务路由策略按预设优先级顺序进行排序;将业务请求按照目标业务路由策略路由至目标服务器,包括:按预设优先级顺序的目标业务路由策略确定对应的目标服务器,直至所确定的目标服务器满足路由条件;将业务请求路由至满足路由条件的目标服务器。在接收业务请求之后,为该业务请求配置多个业务路由策略,每个业务路由策略分别对应不同的目标服务器,以便在业务请求所需路由的服务器处于无法正常工作情况下,将业务请求路由至其他满足条件的目标服务器,保证业务请求可以及时被处理。

[0014] 另外,在确定业务路由策略之前,还包括:通过注册中心的网络接口下载配置文件;解析下载的配置文件。这样可以利用配置文件中路由规则的匹配节点实现不同特征信息业务请求路由策略的匹配,同时通过修改配置文件的匹配节点信息即可调整各类型业务请求所匹配的路由策略,调整方式简单灵活,便于实现。

[0015] 另外,路由策略还包括:将业务请求路由至目标服务器的权重;将业务请求按照目标业务路由策略路由至目标服务器,包括:根据目标业务路由策略中包括的权重,对业务请求路由至目标服务器的流量进行管理。通过设置路由的权重可以对业务请求进行分流,控制了流向目标服务器的流量比例,管理员可以设定不同的权重,轻松实现灰度升级和测试引流等功能。

[0016] 另外,配置文件中还包括以下任一策略或其组合:负载均衡策略,限流策略或容错策略;将业务请求按照路由策略路由至目标服务器,包括:按照路由策略,并结合负载均衡

策略,限流策略或容错策略确定目标服务器;将业务请求路由至目标服务器。通过负载均衡策略可以对处理业务请求的各个服务器的负载情况进行均衡,防止服务器过载。通过容错重试策略可以对异常路由的业务请求进行重新路由或调整路由的目标服务器,从而保证业务请求的正常路由,避免路由异常情况下无法完成业务请求的正常处理。通过限流策略可以防止在某一业务请求类型的流量过大时,导致微服务的服务器被大量的业务请求冲垮,在业务请求的并发数或访问频率超过预设阈值时满足限流策略,不再对该业务请求进行分发,保证服务器的正常运行。

附图说明

[0017] 一个或多个实施例通过与之对应的附图中的图片进行示例性说明,这些示例性说明并不构成对实施例的限定,附图中具有相同参考数字标号的元件表示为类似的元件,除非有特别申明,附图中的图不构成比例限制。

[0018] 图1是根据本发明第一实施例中的微服务请求的流量管理方法的流程图;

[0019] 图2是根据本发明第一实施例中的微服务管理系统的结构示意图;

[0020] 图3是根据本发明第一实施例中的配置文件的配置过程示意图;

[0021] 图4是根据本发明第二实施例中的微服务请求的流量管理方法的流程图;

[0022] 图5是根据本发明第二实施例中根据多个流量管理策略将业务请求路由至目标服务器的过程的流程图;

[0023] 图6是根据本发明第三实施例中的微服务请求的流量管理装置的结构示意图;

[0024] 图7是根据本发明第四实施例中的微服务请求的流量管理系统的结构示意图;

[0025] 图8是根据本发明第五实施例中的服务器的结构示意图。

具体实施方式

[0026] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合附图对本发明的各实施方式进行详细的阐述。然而,本领域的普通技术人员可以理解,在本发明各实施方式中,为了使读者更好地理解本申请而提出了许多技术细节。但是,即使没有这些技术细节和基于以下各实施方式的种种变化和修改,也可以实现本申请所要求保护的技术方案。

[0027] 以下各个实施例的划分是为了描述方便,不应对本发明的具体实现方式构成任何限定,各个实施例在不矛盾的前提下可以相互结合相互引用。

[0028] 本发明的第一实施例涉及一种微服务请求的流量管理方法,包括:代理服务器接收业务请求;根据业务请求携带的特征信息,在多个预设业务路由策略中确定目标业务路由策略;其中,预设业务路由策略在配置文件中设置;将业务请求按照目标业务路由策略路由至目标服务器。实现在微服务架构下对业务请求的流量的定向管理,使用户可以更加灵活的管理流量。下面对本实施例的微服务请求的流量管理方法的实现细节进行具体的说明,以下内容仅为方便理解提供的实现细节,并非实施本方案的必须。

[0029] 具体流程如图1所示,第一实施例涉及一种微服务请求的流量管理方法,包括:

[0030] 步骤101,接收业务请求。具体地说,在用户需要调用微服务时,会向代理服务器发送业务请求,代理服务器接收用户发送的业务请求,并根据业务请求中所请求的微服务将该业务请求转发至相应的微服务的服务器进行处理。在实际应用中,微服务之间可以通过

每个微服务对应的代理服务器实现互相调用,当有大量的微服务相互调用时,微服务之间的调用关系将会形成服务网格,而本实施例中的代理服务器可以适用于服务网格的任意微服务的调用过程。在服务网格中存在多个上述代理服务器,每个代理服务器均可以看作是服务网格的一个网格节点,各个网格节点之间可以相互通信,在代理服务器(网格节点)确定业务请求所请求的微服务对应的服务器时,通过服务网格的网格节点将该业务请求传输至这一服务器对应的服务网格的网格节点,并由这一对应的网格节点将业务请求转发至服务器进行处理,从而可以通过服务网格实现将接收的业务请求的进行管理及转发。

[0031] 步骤102,根据业务请求携带的特征信息,在多个预设业务路由策略中确定目标业务路由策略。

[0032] 具体地说,服务网格在接收业务请求之后,根据业务请求携带的特征信息,对该业务请求的流量进行管理。业务请求携带的特征信息可以从该业务请求的头部信息中获取,特征信息可以是该业务请求的来源信息,所请求的目标服务信息,或者是预先规定的标识信息,如预先规定VIP代表会员信息等。服务网格可以根据业务请求携带的特征信息确定该业务请求的种类,并从预先配置的配置文件中获取为该种类业务请求预先配置的流量管理策略,配置的流量管理策略中至少包括目标业务路由策略,目标业务路由策略规定了该种类业务请求的流量所流向的目标服务器。

[0033] 在确定业务请求配置的目标业务路由策略时,将业务请求携带的特征信息与预先下载的配置文件中的匹配节点相匹配,由于匹配文件中的每个匹配节点均对应设置一个预设业务路由策略,所以通过匹配节点确定业务请求的目标业务路由策略,即根据业务请求携带的特征信息确定相匹配的匹配节点,并将该匹配的匹配节点对应的预设业务路由策略作为目标业务路由策略,从而在配置文件中设置的多个预设业务路由策略中确定该业务请求的目标业务路由策略。

[0034] 配置文件通过注册中心的HTTP网络接口进行下载,服务网格下载配置文件之后,解析配置文件中的业务路由规则,并对业务路由规则进行缓存,从而在接收到业务请求时可以通过缓存的业务路由规则快速匹配到业务请求对应的目标业务路由规则,提高了业务请求的路由效率。另外,服务网格可以通过定时任务定时从注册中心下载最新的配置文件,及时对配置文件进行更新。

[0035] 在通过配置文件中的匹配节点匹配业务请求的特征信息时,若匹配节点中设置了一条匹配信息,则将该匹配信息与业务请求的特征信息进行比对,若业务请求的特征信息包含该匹配信息,则说明业务请求携带的特征信息与业务路由策略的匹配节点相匹配。例如,配置文件中业务路由策略的匹配节点中的匹配信息为VIP会员信息,若业务请求的头部信息同样携带有VIP特征信息,则代表该业务请求与这一业务路由策略的匹配节点相匹配。另外,若匹配节点中设置了多条匹配信息,则可以通过以下几种匹配方式进行匹配:精确匹配,正则匹配,不等于匹配,大于等于匹配,小于等于匹配,大于匹配或小于匹配。精确匹配表示业务请求的特征信息必须与匹配节点定义的信息完全一致,比如,匹配节点定义的信息为VIP信息,业务请求的来源信息A和所请求的目标服务信息B,业务请求携带的特征信息也必须是VIP信息,业务请求的来源信息A和所请求的目标服务信息B才满足精确匹配;正则匹配是将每个特征信息以特定字符表示,比如将1表示VIP信息,2表示业务请求的来源信息A,3表示所请求的目标服务信息B。若匹配节点定义的信息为123,业务请求携带的特征信息

是VIP信息,业务请求的来源信息A和所请求的目标服务信息B则满足正则匹配;相应的,不等于匹配表示业务请求携带的特征信息不等于匹配节点定义的信息则满足匹配;大于等于匹配表示业务请求携带的特征信息不小于匹配节点定义的信息则满足匹配;小于等于匹配表示业务请求携带的特征信息不大于匹配节点定义的信息则满足匹配;大于匹配表示业务请求携带的特征信息大于匹配节点定义的信息则满足匹配;小于匹配表示业务请求携带的特征信息小于匹配节点定义的信息则满足匹配。将配置文件中满足匹配的匹配节点对应的业务路由策略作为携带特征信息的业务请求的目标业务路由策略,并根据目标业务路由策略对业务请求的流量进行管理。

[0036] 在实际应用中,配置文件中的业务路由策略等流量管理数据通过如图2所示中的配置层的页面服务设备生成。页面服务设备为用户及管理员提供网页界面服务,用户及管理员可以直接在网页界面进行配置文件的流量管理数据的配置,并在配置之后将流量管理数据上传至数据库,在流量管理数据上传之后,由注册中心根据上传的流量管理数据生成可扩展标记语言(xml)格式的配置文件,注册中心在检测是否存在最新上传的流量管理数据时,可以根据预先设置的定时任务,定时从数据库中获取页面服务中定义的所有配置,从而可以及时的对配置文件进行更新。用户在配置流量管理数据之后不需要重启服务直接将数据上传至数据库,注册中心可以直接在数据库中获取流量管理数据并生成配置文件,使所配置的数据可以实时生效。注册中心所生成的配置文件可由应用层的服务网格通过注册中心的HTTP网络接口进行下载,从而实现服务网格根据配置文件对业务请求进行流量管理。

[0037] 页面服务设备对流量管理数据的配置过程如图3所示,在网页界面中包括:标签定义,路由规则定义,路由方案定义,服务路由管理、服务标签管理五个部分,其中,用户通过标签定义部分定义需要使用的标签,定义的标签在路由规则中使用,每一个路由规则对应一个定义的标签,通过标签即可找到路由规则中规定的目标服务器;路由规则定义,用于定义路由的规则,路由规则通过yaml格式进行匹配,路由规则的配置中主要包括match和route两个节点,match代表匹配节点,通过match匹配节点定义匹配信息,若match节点并未定义任何信息,则说明这一匹配节点可以与携带任意特征信息的业务请求相匹配;route代表分发节点,通过route分发节点定义路由至的目标服务器,route分发节点定义的目标服务器的数量可以为一个服务器也可以为一组多个服务器组成的服务器组。在对业务请求携带的特征信息匹配相应的路由规则时,确定与业务请求相匹配的匹配节点,并将匹配节点对应的分发节点中定义的目标服务器作为该业务请求的路由目标,从而实现业务请求的定向管理。另外,route分发节点还可以定义路由至目标服务器的权重,从而管理业务请求路由之目标服务器的流量。路由方案定义部分用于定义路由的方案,路由方案由多个路由规则组成,方案内多个路由规则以优先级从高到低的顺序排列,路由方案可以在网页界面上直接拖拉需要的路由规则进行配置,操作简单。服务路由管理部分是针对应用或服务进行路由方案的配置,可以针对某一应用或服务配置多个路由方案,将多个路由方案对应的目标服务器作为一个路由分组,在接收到这一应用或服务的业务请求时,将业务请求路由至路由分组,在路由至路由分组后根据最终确定的路由方案确定目标服务器。若在确定的路由方案的目标服务器无法为业务请求提供服务时,可以执行预先设定的动作,设定的动作可以是all,break,default。动作all表示从路由分组中随机选择一个服务器作为目标服务

器;动作break表示返回没有提供者错误,结束调用服务;动作default表示从默认服务器分组中选择一个服务器作为目标服务器。以上路由规则定义,路由方案定义,服务路由管理部分所定义的路由规则,路由方案及服务路由方案均为业务请求的路由策略,是针对不同的需求提供不同的路由策略,从而适应不同的应用场景。最后一个服务标签管理部分是管理服务器的分组标签,将多个同一服务的服务器作为一个分组,并为该分组分配一个标签,在服务标签管理部分用户可以在页面查看或修改服务器分组的标签。标签修改之后无需重启,可以实时生效。综上所述,在配置流量管理数据时,可以通过网页界面中的以上模块依次对配置文件中的流量管理数据进行设定,从而灵活的实现对配置文件中的流量管理规则进行配置,从而更好的实现业务请求的流量管理。

[0038] 步骤103,将业务请求按照目标业务路由策略路由至目标服务器。

[0039] 具体地说,根据目标业务路由策略中定义的目标服务器,服务网格可以确定业务请求路由的路由目标,服务网格根据确定的路由目标将业务请求转发至对应的网格节点,并由对应的网格节点发送至目标服务器,最后由目标服务器对业务请求进行处理。

[0040] 在实际应用中,若配置文件中并未配置业务请求的业务路由策略,可以通过预先设置的质量路由策略对该业务请求进行路由。质量路由规则为通用规则,对接收到的所有业务请求以统一的规则进行路由,不会因为业务请求携带的特征信息而调整路由策略。质量路由规则可以为根据运营商、机房、网络情况、机器服务能力等信息,挑选出质量较好的服务器作为目标服务器,将业务请求路由至所挑选的目标服务器,完成依据质量路由策略实现的路由,例如,质量路由策略可以设置为,优先级由高到低依次为:同机房优先,同网络优先,同运营商优先等等。

[0041] 本发明实施方式相对于现有技术而言,在接收到用户的业务请求时,获取业务请求中携带的特征信息,特征信息为某一个业务请求或某一类业务请求的标识信息,通过该特征信息即可确定业务请求的类型,从而区分不同类型的业务请求。在获取业务请求的特征信息之后,获取预先配置的该类型的业务请求的目标业务路由策略,业务路由策略由管理员根据各个类型的业务请求的实际需求进行预先配置,在业务请求的实际需求发生变化时可以对业务路由策略的配置项进行修改,从而可以灵活的实现业务请求的流量的管理。在根据预先配置的配置文件中确定目标业务路由策略之后,根据目标业务路由策略将业务请求路由至目标服务器,由于目标业务路由策略中规定了业务请求所路由的目标服务器,因此实现了对业务请求流量的定向管理,使用户可以灵活的对流量进行管理。

[0042] 本发明的第二实施例涉及一种微服务请求的流量管理方法。在本发明第二实施例中,具体说明了多个不同的流量管理策略相互配合实现流量的管理的过程,流量管理策略包括:上述提到的业务路由策略和质量路由策略,还可以包括负载均衡策略,负载均衡策略,限流策略,容错重试策略,在应用中可以将上述策略自由组合。

[0043] 具体流程如图4所示,第二实施例涉及一种微服务请求的流量管理方法,包括:

[0044] 步骤301,接收业务请求。

[0045] 步骤302,根据业务请求携带的特征信息,在多个预设业务路由策略中确定目标业务路由策略。

[0046] 步骤301、302与第一实施例中步骤101、102分别对应,在此不再赘述。

[0047] 步骤303,根据多个流量管理策略,将业务请求路由至目标服务器。具体地说,流量

管理策略可以为以下任一策略或其组合:路由策略,负载均衡策略,负载均衡策略,限流策略,容错重试策略等,通过负载均衡策略可以对处理业务请求的各个服务器的负载情况进行均衡,防止服务器过载。通过容错重试策略可以对异常路由的业务请求进行重新路由或调整路由的目标服务器,从而保证业务请求的正常路由,避免路由异常情况下无法完成业务请求的正常处理。通过限流策略可以防止在某一业务请求类型的流量过大时,导致微服务的服务器被大量的业务请求冲垮,在业务请求的并发数或访问频率超过预设阈值时满足限流策略,不再对该业务请求进行分发,保证服务器的正常运行。

[0048] 在流量管理策略为多个时,需要综合考虑多个策略决定业务请求路由的目标服务器,下面对其中一种流量管理策略的组合方式进行具体说明,实际应用中可以根据需求自由组合,在此对流量管理策略的组合方式不做限制。

[0049] 步骤3031,根据路由策略确定多个目标服务器。其中,路由策略包括业务路由策略和质量路由策略。

[0050] 具体地说,业务路由策略在上述第一实施例中进行了具体的说明,在此不再赘述,想要说明的是,在配置文件中并未获取到与业务请求携带的特征信息相匹配的匹配节点时,说明此时配置文件中并未规定该种类业务请求的业务路由规则,这时需要通过预先设置的质量路由策略确定业务请求路由的目标服务器,质量路由策略根据运营商、机房、网络情况、服务器服务能力等信息,选择质量最好的可以为用户提供服务的服务器作为目标服务器,根据质量路由策略所确定的目标服务器的数量可以为一个也可以为多个。

[0051] 步骤3032,根据负载均衡策略,在多个目标服务器中选择一个目标服务器。

[0052] 具体地说,在根据路由策略确定的目标服务器的数量为多个时,根据负载均衡策略可以将业务请求的流量分发至多个服务器均衡处理,从而优化响应时长,并且防止服务器过载。负载均衡策略至少包括以下三种策略:简单策略,带权重循环策略和连接数策略。简单策略不对服务器的路由顺序进行排序,根据自定义的路由服务器直接路由至目标服务器。带权重循环策略是根据设置的权重大小确定各个服务器的调用次数。连接数策略是根据服务器连接数量进行选择,优先使用与客户端连接数最少的服务器。以上三种负载均衡策略可以由用户自行配置,根据用户的需求选择合适的负载均衡策略。

[0053] 步骤3033,若根据路由策略对业务请求进行路由的过程中出现异常,则根据容错重试策略确定目标服务器。

[0054] 具体地说,在由路由策略,或负载均衡策略确定的目标服务器出现异常时,比如目标服务器负载过大或出现故障无法被调用时,采用容错重试策略对这种异常情况进行处理,容错重试策略可以包括:只允许组内重试,分组外重试,优先重试本机房等,只允许组内重试中的分组为针对这一业务请求的服务器分组,组内包含多个服务器,各个服务器均可为用户提供服务,分组外重试中的分组也同样为针对这一业务请求的服务器分组。上述容错重试策略也可以由用户进行配置,而管理员可以设置是否允许重试,若管理员设置不允许重试,则用户无法配置容错重试策略。

[0055] 另外,还包括其他流量管理策略,例如,限流策略等,限流策略可以防止业务请求流量过大导致服务器被冲垮的问题,保证服务器的正常运行。用户可以通过配置业务请求的并发数和访问频率等实现限流策略的设置,当服务器检测到某一服务的业务请求的并发数或访问频率达到预先设置的阈值时,不再接收该服务的业务请求,实现限流管理。

[0056] 通过以上各个流量管理策略的组合,用户可以轻松实现灰度升级,测试引流,资源隔离等功能,下面针对灰度升级功能的实现方式进行说明。

[0057] 假设某一服务部署在A、B两台服务器上,现需要对这两台机器进行灰度升级,将A服务器升级到新版本,B服务器暂时不升级。规定将该服务的业务请求中的10%的流量转发至A服务器做灰度测试,剩余90%的流量转发至B服务器。该灰度测试功能对应配置配置文件的配置过程如下:首先在页面服务的标签定义中定义tagA、tagB两个标签,用于指示A、B两台服务器。其次,在路由规则定义中定义本次灰度测试所需的路由规则,将10%的流量转发至A服务器,将90%的流量转发至B服务器,由于灰度测试功能无需定义匹配规则,所以不需要定义匹配节点,只需配置route节点即可。在路由方案定义中新增路由方案,路由方案选择上述路由规则。在服务路由管理中将该灰度升级功能的服务配置上述新增的路由方案。在服务标签管理中将A服务器的标签配置为tagA,B服务器的标签配置为tagB,完成配置文件的配置。通过上述配置可以使A、B两台服务器收到的业务请求流量的比例趋近1:9,即可实现将少部分的流量引到A服务器进行灰度测试。在灰度测试之后可以通过调整路由至各服务器的权重,最终将所有的业务请求引入上级后的A服务器。

[0058] 另外,测试引流和资源隔离等功能同样可以通过不同的配置实现,测试引流是将业务请求引至特定的服务器实现测试功能,资源隔离则是将业务请求隔离,不允许业务请求流至特定的服务器,修改route节点的配置即可实现,在此就不一一说明了。

[0059] 本发明的第三实施例涉及一种微服务请求的流量管理方法,应用于注册中心,方法包括:读取页面服务设备的数据库中的流量管理数据;其中,流量管理数据中至少包括:多个与业务请求携带的特征信息相匹配的预设业务路由策略;根据读取的流量管理数据生成配置文件,以供微服务架构根据配置文件对接收的业务请求进行路由。下面对本实施例的微服务请求的流量管理方法的实现细节进行具体的说明。

[0060] 具体流程如图6所示,第三实施例涉及一种微服务请求的流量管理方法,包括:

[0061] 步骤601,注册中心读取页面服务设备的数据库中的流量管理数据。具体地说,在管理员通过页面服务设备对配置文件进行配置的过程中,页面服务设备将所配置的流量管理数据存储至数据库中,在配置完成之后或需要生成配置文件时,注册中心从数据库中读取流量管理数据。

[0062] 步骤602,根据读取的流量管理数据生成配置文件。具体地说,在获取到流量管理数据之后,根据流量管理数据生成配置文件,生成后的配置文件即可根据管理员设置的规则对业务请求的流量进行定向管理。

[0063] 在实际应用中,可以通过设定定时任务,定时从数据库中读取流量管理数据并生成配置文件,从而可以及时的对管理员设置的流量管理策略进行更新,更好的对业务请求进行管理。

[0064] 本发明实施方式相对于现有技术而言,注册中心读取页面服务设备的数据库中的流量管理数据;并根据读取的流量管理数据生成配置文件,生成的配置文件中包括了多个与业务请求携带的特征信息相匹配的预设业务路由策略,从而规定了不同类型的业务请求所路由的目标服务器,使微服务架构可以根据该生成的配置文件对接收的业务请求进行定向的路由,实现了对业务请求流量的定向管理。

[0065] 本发明的第四实施例涉及一种配置文件的生成方法,应用于页面服务设备,方法

包括:设置多个与业务请求携带的特征信息相匹配的预设业务路由策略;将多个预设业务路由策略作为流量管理数据存储于数据库,以供注册中心根据流量管理数据生成配置文件。下面对本实施例的微服务请求的流量管理方法的实现细节进行具体的说明。

[0066] 具体流程如图7所示,第四实施例涉及一种微服务请求的流量管理方法,包括:

[0067] 步骤701,页面服务设备设置多个与业务请求携带的特征信息相匹配的预设业务路由策略。页面服务设备显示标签定义,路由规则定义,路由方案定义,服务路由管理,服务标签管理等模块,管理员通过上述模块进行匹配节点即预设业务路由策略等配置文件的配置,配置过程在第一实施方式中有具体的说明,在此不再赘述。

[0068] 步骤702,将多个预设业务路由策略作为流量管理数据存储于数据库,以供注册中心根据流量管理数据生成配置文件。

[0069] 本发明实施方式相对于现有技术而言,页面服务设备设置多个与业务请求携带的特征信息相匹配的预设业务路由策略;将多个预设业务路由策略作为流量管理数据存储于数据库,以供注册中心根据流量管理数据生成配置文件,微服务架构则可以通过生成的配置文件对业务请求进行定向路由,从而实现业务请求的定向管理。

[0070] 上面各种方法的步骤划分,只是为了描述清楚,实现时可以合并为一个步骤或者对某些步骤进行拆分,分解为多个步骤,只要包括相同的逻辑关系,都在本专利的保护范围内;对算法中或者流程中添加无关紧要的修改或者引入无关紧要的设计,但不改变其算法和流程的核心设计都在该专利的保护范围内。

[0071] 本发明第三实施例涉及一种微服务请求的流量管理装置,如图6所示,包括:接收模块61,确定模块62,路由模块63;接收模块61用于接收业务请求;确定模块62用于根据业务请求携带的特征信息,在多个预设业务路由策略中确定目标业务路由策略;其中,预设业务路由策略在配置文件中设置;路由模块63用于将业务请求按照目标业务路由策略路由至目标服务器。

[0072] 不难发现,本实施例为与第一实施例相对应的装置实施例,本实施例可与第一实施例互相配合实施。第一实施例中提到的相关技术细节在本实施例中依然有效,为了减少重复,这里不再赘述。相应地,本实施例中提到的相关技术细节也可应用在第一实施例中。

[0073] 另外,确定模块62用于获取各个预设业务路由策略对应的匹配节点定义的信息;若预设业务路由策略的匹配节点定义的信息与特征信息相同,或预设业务路由策略的匹配节点未定义信息,则将预设业务路由策略作为目标业务路由策略。

[0074] 另外,确定模块62用于在目标业务路由策略的数量大于一时,将目标业务路由策略按预设优先级顺序进行排序;路由模块63用于按预设优先级顺序的目标业务路由策略确定对应的目标服务器,直至所确定的目标服务器满足路由条件;将业务请求路由至满足路由条件的目标服务器。

[0075] 另外,为服务请求的流量管理模块还包括:下载模块,下载模块用于通过注册中心的网络接口下载配置文件;解析下载的配置文件。

[0076] 另外,路由模块63还用于根据目标业务路由策略中包括的权重,对业务请求路由至目标服务器的流量进行管理。

[0077] 另外,路由模块63还用于按照路由策略,并结合负载均衡策略,限流策略或容错策略确定目标服务器;将业务请求路由至目标服务器。

[0078] 值得一提的是,本实施例中所涉及到的各模块均为逻辑模块,在实际应用中,一个逻辑单元可以是一个物理单元,也可以是一个物理单元的一部分,还可以以多个物理单元的组合实现。此外,为了突出本发明的创新部分,本实施方式中并没有将与解决本发明所提出的技术问题关系不太密切的单元引入,但这并不表明本实施例中不存在其它的单元。

[0079] 本发明第四实施方式涉及一种微服务请求的流量管理系统,如图7所示,包括:页面服务设备701,注册中心702和代理服务器703;页面服务设备701用于设置多个与业务请求携带的特征信息相匹配的预设业务路由策略,并将多个预设业务路由策略作为流量管理数据存储至数据库中;注册中心702用于获取流量管理数据,并根据流量管理数据生成配置文件;代理服务器703用于在接收到业务请求之后,根据业务请求携带的特征信息,在注册中心下载的配置文件中的多个预设业务路由策略中确定目标业务路由策略;并将业务请求按照目标业务路由策略路由至目标服务器。

[0080] 由于本发明实施例中的代理服务器可以用于执行上述微服务请求的流量管理方法流程,因此,其所能获得的技术效果也可参考上述方法实施例,本发明实施例在此不再赘述。

[0081] 本发明第五实施例涉及一种服务器,如图8所示,包括:至少一个处理器801;以及,与至少一个处理器801通信连接的存储器802;其中,存储器802存储有可被至少一个处理器801执行的指令,指令被至少一个处理器801执行,以使至少一个处理器801能够执行上述微服务请求的流量管理方法。

[0082] 其中,存储器802和处理器801采用总线方式连接,总线可以包括任意数量的互联的总线和桥,总线将一个或多个处理器801和存储器802的各种电路连接在一起。总线还可以将诸如外围设备、稳压器和功率管理电路等之类的各种其他电路连接在一起,这些都是本领域所公知的,因此,本文不再对其进行进一步描述。总线接口在总线和收发机之间提供接口。收发机可以是一个元件,也可以是多个元件,比如多个接收器和发送器,提供用于在传输介质上与各种其他装置通信的单元。经处理器处理的数据通过天线在无线介质上进行传输,进一步,天线还接收数据并将数据传送给处理器801。

[0083] 处理器801负责管理总线和通常的处理,还可以提供各种功能,包括定时,外围接口,电压调节、电源管理以及其他控制功能。而存储器802可以被用于存储处理器801在执行操作时所使用的数据。

[0084] 本发明第六实施方式涉及一种计算机可读存储介质,存储有计算机程序。计算机程序被处理器执行时实现上述方法实施例。

[0085] 即,本领域技术人员可以理解,实现上述实施例方法中的全部或部分步骤是可以通程序来指令相关的硬件来完成,该程序存储在一个存储介质中,包括若干指令用以使得一个设备(可以是单片机,芯片等)或处理器(processor)执行本申请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。

[0086] 本领域的普通技术人员可以理解,上述各实施方式是实现本发明的具体实施例,而在实际应用中,可以在形式上和细节上对其作各种改变,而不偏离本发明的精神和范围。

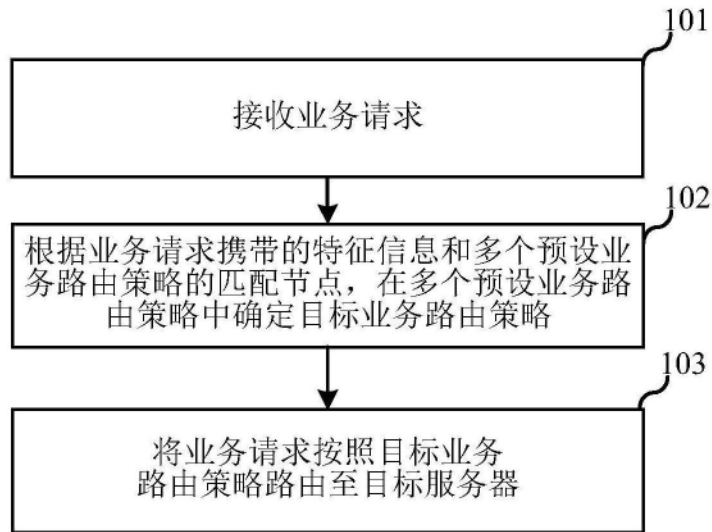


图1

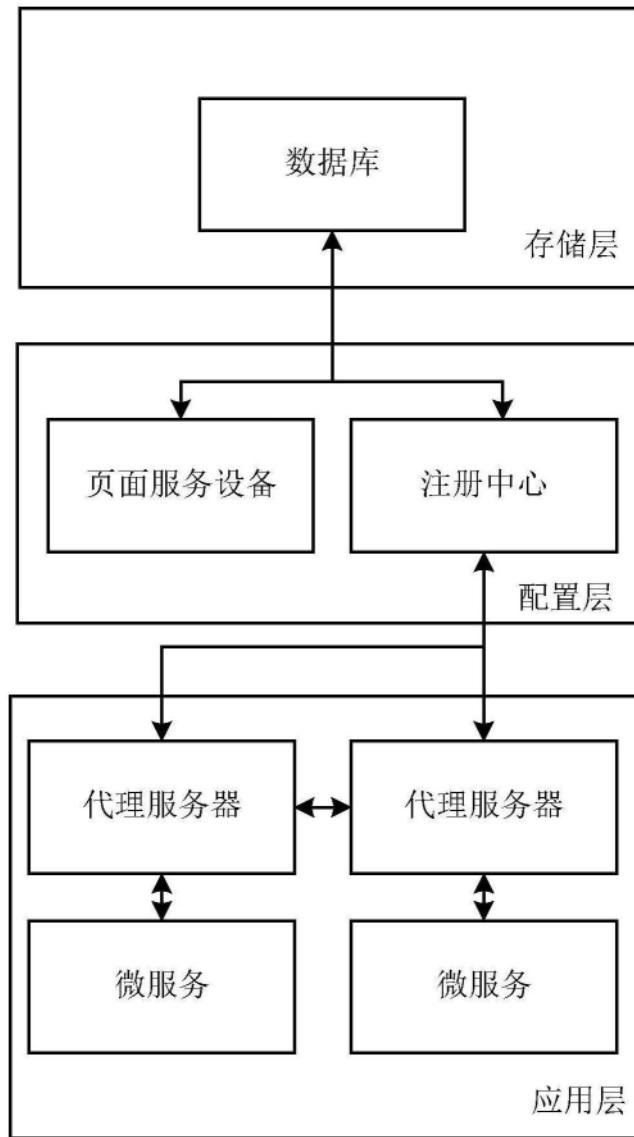


图2

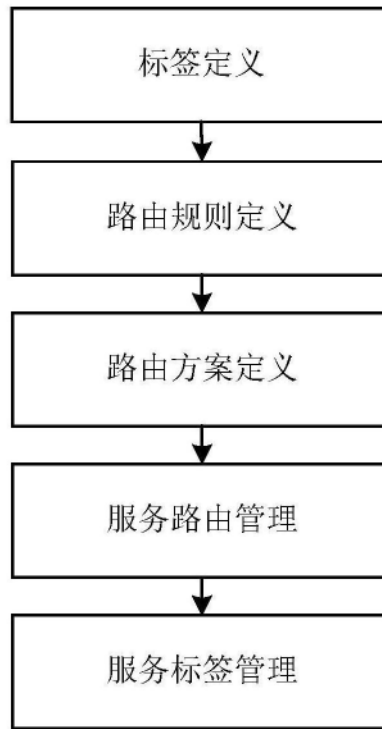


图3

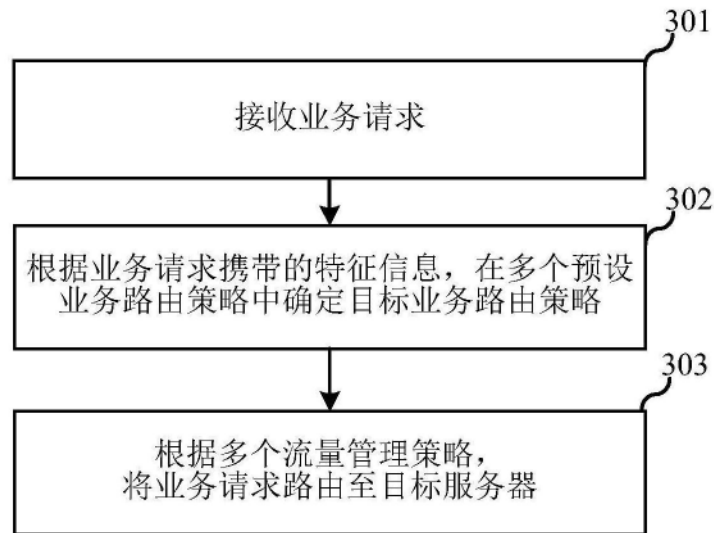


图4

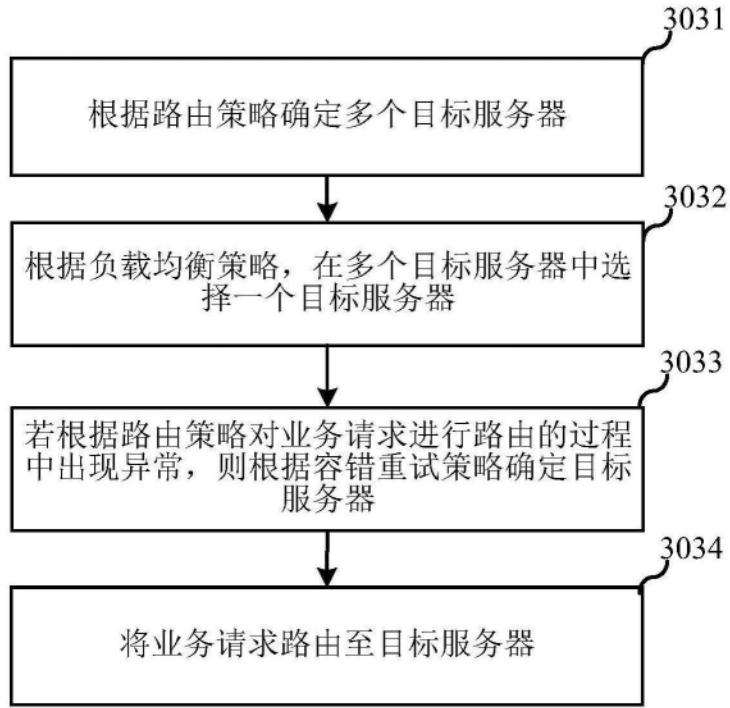


图5



图6



图7

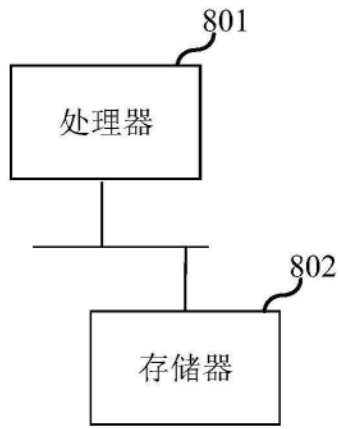


图8