



(12)发明专利申请

(10)申请公布号 CN 111367662 A

(43)申请公布日 2020.07.03

(21)申请号 202010119796.8

(22)申请日 2020.02.26

(71)申请人 普信恒业科技发展(北京)有限公司  
地址 100026 北京市朝阳区西大望路1号1  
号楼16层

(72)发明人 陈晓宇 尚伏鹏 张健伟

(74)专利代理机构 北京集佳知识产权代理有限  
公司 11227  
代理人 王小清

(51)Int.Cl.  
G06F 9/50(2006.01)

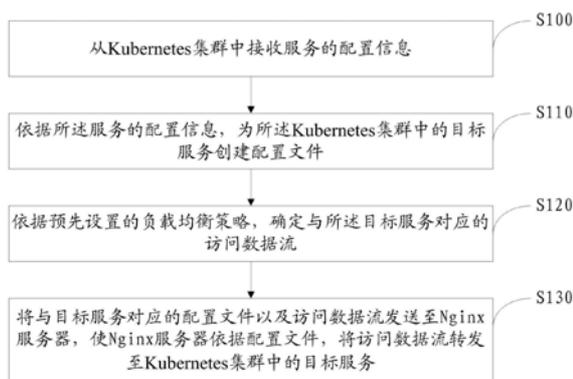
权利要求书2页 说明书9页 附图5页

(54)发明名称

一种负载均衡方法、装置及系统

(57)摘要

本发明提供了一种负载均衡方法、装置及系统,该方法包括:Kubernetes集群将服务的配置信息发送至Nginx管理器,Nginx管理器依据服务的配置信息,为Kubernetes集群中的目标服务创建配置文件;并依据预先设置的负载均衡策略,确定与目标服务对应的访问数据流;将与目标服务对应的配置文件以及访问数据流发送至Nginx服务器;Nginx服务器依据配置文件,将访问数据流转发至目标服务。Nginx管理器可以依据预先设置的负载均衡策略,确定与目标服务对应的访问数据流,实现对Kubernetes集群中目标服务对应的访问数据流的统一均衡分配,从而实现Kubernetes集群中服务的负载均衡。



1. 一种负载均衡方法,其特征在于,所述方法应用于Nginx管理器,所述方法包括:  
从Kubernetes集群中接收服务的配置信息;  
依据所述服务的配置信息,为所述Kubernetes集群中的目标服务创建配置文件;  
依据预先设置的负载均衡策略,确定与所述目标服务对应的访问数据流;  
将与所述目标服务对应的配置文件以及访问数据流发送至Nginx服务器,使所述Nginx服务器依据所述配置文件,将所述访问数据流转发至Kubernetes集群中的目标服务。
2. 根据权利要求1所述的方法,其特征在于,所述依据所述服务的配置信息,为所述Kubernetes集群中的目标服务创建配置文件包括:  
依据所述服务的配置信息,确定需要在Nginx管理器中创建配置文件的目标服务;  
依据所述目标服务的配置信息,为所述目标服务创建配置文件。
3. 根据权利要求1所述的方法,其特征在于,所述方法还包括:  
从所述Kubernetes集群中接收目标服务的更新配置信息;  
利用所述目标服务的更新配置信息,更新所述目标服务对应的配置文件。
4. 根据权利要求1所述的方法,其特征在于,所述方法还包括:  
向Nginx服务器发送流量监控指示信息,使所述Nginx服务器对发送至Kubernetes集群中的访问数据流的流量进行监控。
5. 根据权利要求1所述的方法,其特征在于,所述目标服务对应的访问数据流包括:  
Kubernetes系统中Kubernetes集群之间的访问数据流和/或所述Kubernetes系统外部的访问数据流。
6. 一种负载均衡装置,其特征在于,所述装置应用于Nginx管理器,所述装置包括:  
配置信息接收单元,用于从Kubernetes集群中接收服务的配置信息;  
配置文件创建单元,用于依据所述服务的配置信息,为所述Kubernetes集群中的目标服务创建配置文件;  
访问数据流确定单元,用于依据预先设置的负载均衡策略,确定与所述目标服务对应的访问数据流;  
发送单元,用于将与所述目标服务对应的配置文件以及访问数据流发送至Nginx服务器,使所述Nginx服务器依据所述配置文件,将所述访问数据流转发至Kubernetes集群中的目标服务。
7. 根据权利要求6所述的装置,其特征在于,所述配置文件创建单元包括:  
目标服务确定单元,用于依据所述服务的配置信息,确定需要在Nginx管理器中创建配置文件的目标服务;  
配置文件创建子单元,用于依据所述目标服务的配置信息,为所述目标服务创建配置文件。
8. 根据权利要求6所述的装置,其特征在于,所述装置还包括:  
更新配置信息接收单元,用于从所述Kubernetes集群中接收目标服务的更新配置信息;  
更新单元,用于利用所述目标服务的更新配置信息,更新所述目标服务对应的配置文件。
9. 一种负载均衡系统,其特征在于,所述系统至少包括:Nginx管理器、Nginx服务器以

及Kubernetes集群；

所述Kubernetes集群用于将服务的配置信息发送至Nginx管理器；

所述Nginx管理器用于依据所述服务的配置信息,为所述Kubernetes集群中的目标服务创建配置文件;依据预先设置的负载均衡策略,确定与所述目标服务对应的访问数据流;将与所述目标服务对应的配置文件以及访问数据流发送至Nginx服务器;

所述Nginx服务器用于依据所述配置文件,将所述访问数据流转发至Kubernetes集群中的目标服务。

10. 根据权利要求9所述的系统,其特征在于,所述系统还包括:

Kubernetes集群管理器,所述Kubernetes集群管理器用于在所述Kubernetes集群中创建服务,并配置服务的配置信息。

## 一种负载均衡方法、装置及系统

### 技术领域

[0001] 本发明涉及信息处理技术领域,更具体的说,是涉及一种负载均衡方法、装置及系统。

### 背景技术

[0002] Kubernetes是一种开源的容器编排与调度系统,Kubernetes系统是由多个Kubernetes集群构成的,每个Kubernetes集群中包含至少一个服务,每个服务都是由至少一个pod组成的,每个pod中运行着与其对应的服务的进程。由于Kubernetes系统内部的Kubernetes集群与Kubernetes集群之间可以互相进行访问,Kubernetes系统外部的主机也可以对Kubernetes系统内部的Kubernetes集群进行访问,基于此,如何实现Kubernetes集群中服务的负载均衡,成为目前需要解决的技术问题。

### 发明内容

[0003] 有鉴于此,本发明提供了一种负载均衡方法、装置及系统,以实现Kubernetes集群中服务的负载均衡。

[0004] 为实现上述目的,本发明提供如下技术方案:

[0005] 一种负载均衡方法,所述方法应用于Nginx管理器,所述方法包括:

[0006] 从Kubernetes集群中接收服务的配置信息;

[0007] 依据所述服务的配置信息,为所述Kubernetes集群中的目标服务创建配置文件;

[0008] 依据预先设置的负载均衡策略,确定与所述目标服务对应的访问数据流;

[0009] 将与所述目标服务对应的配置文件以及访问数据流发送至Nginx服务器,使所述Nginx服务器依据所述配置文件,将所述访问数据流转发至Kubernetes集群中的目标服务。

[0010] 优选的,所述依据所述服务的配置信息,为所述Kubernetes集群中的目标服务创建配置文件包括:

[0011] 依据所述服务的配置信息,确定需要在Nginx管理器中创建配置文件的目标服务;

[0012] 依据所述目标服务的配置信息,为所述目标服务创建配置文件。

[0013] 优选的,所述方法还包括:

[0014] 从所述Kubernetes集群中接收目标服务的更新配置信息;

[0015] 利用所述目标服务的更新配置信息,更新所述目标服务对应的配置文件。

[0016] 优选的,所述方法还包括:

[0017] 向Nginx服务器发送流量监控指示信息,使所述Nginx服务器对发送至Kubernetes集群中的访问数据流的流量进行监控。

[0018] 优选的,所述目标服务对应的访问数据流包括:

[0019] Kubernetes系统中Kubernetes集群之间的访问数据流和/或所述Kubernetes系统外部的访问数据流。

[0020] 一种负载均衡装置,所述装置应用于Nginx管理器,所述装置包括:

- [0021] 配置信息接收单元,用于从Kubernetes集群中接收服务的配置信息;
- [0022] 配置文件创建单元,用于依据所述服务的配置信息,为所述Kubernetes集群中的目标服务创建配置文件;
- [0023] 访问数据流确定单元,用于依据预先设置的负载均衡策略,确定与所述目标服务对应的访问数据流;
- [0024] 发送单元,用于将与所述目标服务对应的配置文件以及访问数据流发送至Nginx服务器,使所述Nginx服务器依据所述配置文件,将所述访问数据流转发至Kubernetes集群中的目标服务。
- [0025] 优选的,所述配置文件创建单元包括:
- [0026] 目标服务确定单元,用于依据所述服务的配置信息,确定需要在Nginx管理器中创建配置文件的目標服务;
- [0027] 配置文件创建子单元,用于依据所述目标服务的配置信息,为所述目标服务创建配置文件。
- [0028] 优选的,所述装置还包括:
- [0029] 更新配置信息接收单元,用于从所述Kubernetes集群中接收目标服务的更新配置信息;
- [0030] 更新单元,用于利用所述目标服务的更新配置信息,更新所述目标服务对应的配置文件。
- [0031] 一种负载均衡系统,所述系统至少包括:Nginx管理器、Nginx服务器以及Kubernetes集群;
- [0032] 所述Kubernetes集群用于将服务的配置信息发送至Nginx管理器;
- [0033] 所述Nginx管理器用于依据所述服务的配置信息,为所述Kubernetes集群中的目标服务创建配置文件;依据预先设置的负载均衡策略,确定与所述目标服务对应的访问数据流;将与所述目标服务对应的配置文件以及访问数据流发送至Nginx服务器;
- [0034] 所述Nginx服务器用于依据所述配置文件,将所述访问数据流转发至Kubernetes集群中的目标服务。
- [0035] 优选的,所述系统还包括:
- [0036] Kubernetes集群管理器,所述Kubernetes集群管理器用于在所述Kubernetes集群中创建服务,并配置服务的配置信息。
- [0037] 经由上述技术方案可知,与现有技术相比,本发明提供了一种负载均衡方法、装置及系统,该方法包括:Kubernetes集群将服务的配置信息发送至Nginx管理器,Nginx管理器依据所述服务的配置信息,为所述Kubernetes集群中的目标服务创建配置文件;并依据预先设置的负载均衡策略,确定与所述目标服务对应的访问数据流;将与所述目标服务对应的配置文件以及访问数据流发送至Nginx服务器;Nginx服务器依据所述配置文件,将所述访问数据流转发至目标服务。Nginx管理器可以依据预先设置的负载均衡策略,确定与目标服务对应的访问数据流,实现对Kubernetes集群中目标服务对应的访问数据流的统一均衡分配,从而实现Kubernetes集群中服务的负载均衡。

## 附图说明

[0038] 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,下面描述中的附图仅仅是本发明的实施例,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,还可以根据提供的附图获得其他的附图。

[0039] 图1为本发明实施例提供的负载均衡系统的一种结构示意图;

[0040] 图2为本发明实施例提供的负载均衡方法的信令流程图;

[0041] 图3为本发明实施例提供的负载均衡系统的另一种结构示意图;

[0042] 图4为本发明实施例提供的负载均衡方法的一种流程图;

[0043] 图5为本发明实施例提供的负载均衡方法的另一种流程图;

[0044] 图6为本发明实施例提供的负载均衡装置的结构框图。

## 具体实施方式

[0045] 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、完整地描述,显然,所描述的实施例仅仅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。

[0046] Kubernetes是一种开源的容器编排与调度系统,Kubernetes系统是由多个Kubernetes集群构成的,每个Kubernetes集群中包含至少一个服务,每个服务都是由至少一个pod组成的,每个pod中运行着与其对应的服务的进程。由于Kubernetes系统内部的Kubernetes集群与Kubernetes集群之间可以互相进行访问,Kubernetes系统外部的主机也可以对Kubernetes系统内部的Kubernetes集群进行访问,基于此,如何实现Kubernetes集群内部的服务负载均衡,以及Kubernetes集群与Kubernetes集群之间的服务负载均衡成为目前需要解决的技术问题。

[0047] 本发明提供了一种负载均衡方法、装置及系统,该方法包括:Kubernetes集群将服务的配置信息发送至Nginx管理器,Nginx管理器依据所述服务的配置信息,为所述Kubernetes集群中的目标服务创建配置文件;并依据预先设置的负载均衡策略,确定与所述目标服务对应的访问数据流;将与所述目标服务对应的配置文件以及访问数据流发送至Nginx服务器;Nginx服务器依据所述配置文件,将所述访问数据流转发至Kubernetes集群中的目标服务。Nginx管理器可以依据预先设置的负载均衡策略,确定与目标服务对应的访问数据流,实现对Kubernetes集群中目标服务对应的访问数据流的统一均衡分配,从而实现Kubernetes集群中服务的负载均衡。

[0048] 图1为本发明实施例提供的负载均衡系统的结构示意图,本发明实施例提供的负载均衡方法可由该负载均衡系统实施;参照图1,该负载均衡系统可以包括:Nginx管理器10、Nginx服务器11以及Kubernetes集群12。

[0049] 基于图1所示的负载均衡系统,图2示出了本发明实施例提供的负载均衡方法的信令流程图,参照图2,该流程可以包括:

[0050] 步骤S10、Kubernetes集群将服务的配置信息发送至Nginx管理器;

[0051] 步骤S20、Nginx管理器依据所述服务的配置信息,为所述Kubernetes集群中的目

标服务创建配置文件；

[0052] 步骤S30、Nginx管理器依据预先设置的负载均衡策略，确定与所述目标服务对应的访问数据流；

[0053] 步骤S40、Nginx管理器将与所述目标服务对应的配置文件以及访问数据流发送至Nginx服务器；

[0054] 步骤S50、Nginx服务器依据所述配置文件，将所述访问数据流转发至Kubernetes集群中的目标服务。

[0055] 需要说明的是，本发明实施例中的一个Nginx管理器下对应控制多个Nginx服务器，Kubernetes集群也存在多个，因此，一个Nginx服务器可以与一个Kubernetes集群对应，一个Nginx服务器也可以与多个Kubernetes集群对应，本发明实施例不做具体限定。

[0056] 图3为本发明实施例提供的负载均衡系统的另一结构示意图，参照图3，该负载均衡系统可以包括：Nginx管理器20、Nginx服务器21、Kubernetes集群22以及Kubernetes集群管理器23。

[0057] 所述Kubernetes集群管理器用于在所述Kubernetes集群中创建服务，并配置服务的配置信息。

[0058] 下面以Nginx管理器的角度，对本发明实施例提供的负载均衡方法进行介绍，下文描述的负载均衡方法可与上文描述的信令流程内容相互对应参照。

[0059] 图4为本发明实施例提供的负载均衡方法的流程图，该方法可应用于所述Nginx管理器，参照图4，所述方法可以包括：

[0060] 步骤S100、从Kubernetes集群中接收服务的配置信息；

[0061] Nginx管理器会与多个kubernetes集群的API (Application Programming Interface, 应用程序接口) 相连接，kubernetes集群的API监听到在Kubernetes集群中存在新服务的创建后，则获取新创建的服务的配置信息，然后将获取到的新创建的服务的配置信息发送至Nginx管理器。

[0062] 另一方面，本发明实施例中的Nginx管理器也会通过watch机制获取Kubernetes集群中新服务的创建。

[0063] 同样的，在kubernetes集群的API监听到在Kubernetes集群中有服务的配置信息发生更新后，则会将更新的服务的配置信息发送给Nginx管理器。

[0064] 一种服务对应一种配置信息，不同服务的配置信息的内容不同，服务的配置信息至少包括：容器地址、服务域名、服务端口、以及服务的各种注解信息等，本发明实施例不做具体限定。

[0065] 服务域名用于唯一标识一服务，可选的，服务域名的形式可以为：“服务名称-命名空间-统一后缀”，命名空间中主要承载的是服务所在kubernetes集群中的位置信息，对于统一后缀的具体内容本发明实施例不做具体限定。本发明实施例中的服务的域名主要通过DNS (Domain Name System, 域名系统 (服务) 协议) 解析到Nginx管理器上。

[0066] 一种服务对应唯一的一个服务端口，同一服务在不同的kubernetes集群中具有同一个服务端口。

[0067] 步骤S110、依据所述服务的配置信息，为所述Kubernetes集群中的目标服务创建配置文件；

[0068] Nginx管理器在接收到所有服务的配置信息之后,会依据所述服务的配置信息,确定需要在Nginx管理器中创建配置文件的目标服务;依据所述目标服务的配置信息,为所述目标服务创建配置文件。

[0069] 具体的,Nginx管理器会判断任一服务的配置信息中是否包含预设标识,如果包含,则为包含预设标识的目标服务创建配置文件;对于不包含预设标识的目标服务则不为其创建配置文件。具体的,会判断任一服务的配置信息中的注解信息中是否包含domain这个注解,如果包含,则为包含domain这个注解的目标服务创建配置文件。

[0070] 为所述Kubernetes集群中的目标服务创建配置文件的过程为:将该目标服务对应的相关配置信息整合成配置文件的形式,具体的,一个服务对应唯一一个配置文件,配置文件中至少包含如下信息:容器地址、服务的域名、服务的类型(例如是否为直接注册容器的服务等)、服务的端口Nodeport等。

[0071] 具体的,本发明实施例中的Nginx管理器还会从所述Kubernetes集群中接收目标服务的更新配置信息;利用所述目标服务的更新配置信息,更新所述目标服务对应的配置文件。

[0072] 在Kubernetes集群中无论目标服务对应的容器的配置信息(例如Pod地址)发生了更新还是目标服务的配置信息发生了更新,都会将更新的信息发送至Nginx管理器,Nginx管理器利用更新的信息更新所述目标服务对应的配置文件。

[0073] 更新的过程具体可以为将新的配置信息添加到配置文件中,旧的配置信息可以删除、也可以保留。

[0074] 当Kubernetes集群中某一服务被删除时,相应的Nginx管理器中存储的该服务对应的配置文件也会被删除。

[0075] 可选的,本发明实施例中主要是通过Kubernetes集群管理器(例如ipaas)在Kubernetes集群中进行服务创建的,Kubernetes集群管理器用于在所述Kubernetes集群中创建服务,并配置服务的配置信息,在Kubernetes集群管理器进行服务创建时,Kubernetes集群管理器会调用多个kubernetes集群的接口,分别在多个集群创建服务以及服务对应的属性信息Deployment以及服务的配置文件Configmap,Deployment中至少包括:服务的域名、服务的镜像版本、服务的副本个数、服务的环境变量等信息,本发明实施例不做具体限定。

[0076] 步骤S120、依据预先设置的负载均衡策略,确定与所述目标服务对应的访问数据流;

[0077] 负载均衡策略例如为:负载平均策略、权重策略、最少连接数策略、轮询策略、IPhash策略等。

[0078] 负载平均策略指的是实现Kubernetes集群中目标服务上的负载平均;权重策略主要是依据目标服务的响应时间计算权重,根据得出的权重来选择目标服务,相应时间越短的目标服务被选择的概率越大;最少连接数策略是依据根据目标服务当前的连接情况进行负载均衡的,当访问数据流到来时,会选取当前连接数最少的目标服务来处理访问数据流;轮询策略是将访问数据流轮流分配给Kubernetes集群中目标服务;IPhash策略是将源IP地址相同的访问数据流定向到同一目标服务中。

[0079] 可选的,本发明实施例中的目标服务对应的访问数据流包括:

[0080] Kubernetes系统中Kubernetes集群之间的访问数据流和/或所述Kubernetes系统外部的访问数据流。

[0081] 也就是说, Nginx管理器针对的待分配的数据流可以为Kubernetes集群之间的访问数据流,还可以为Kubernetes系统外部的访问数据流。由此可知,本发明实施例不仅可以对Kubernetes集群与Kubernetes集群之间的访问数据流进行均衡分配,还可以对来自Kubernetes系统外部的访问数据流进行均衡分配,本发明实施例并不做具体限定。

[0082] 步骤S130、将与所述目标服务对应的配置文件以及访问数据流发送至Nginx服务器,使所述Nginx服务器依据所述配置文件,将所述访问数据流转发至Kubernetes集群中的目标服务。

[0083] Nginx服务器可以从配置文件中获取目标服务的相关配置信息,例如服务的端口、或容器的IP地址等定位到Kubernetes集群中的目标服务的存储位置,进而将访问数据流转发至Kubernetes集群中的目标服务。

[0084] Nginx管理器在将目标服务对应的配置文件以及访问数据流发送至Nginx服务器之后,还会向Nginx服务器发送“nginx-sreload”加载指示信息,使所述目标Nginx服务器依据所述“nginx-sreload”,加载所述配置信息。

[0085] nginx-sreload可以指示Nginx服务器在不中断自身运行的服务的情况下,加载所述配置文件以及访问数据流,避免了对Nginx服务器运转状态的影响。

[0086] 需要说明的是,目标服务具体可以指的是位于同一个Kubernetes集群中的服务,还可以是位于不同Kubernetes集群中的服务,也就是说,本发明实施例可以对同一个Kubernetes集群中的服务进行负载均衡调节,还可以对位于不同Kubernetes集群中的服务进行负载均衡调节。

[0087] 本发明实施例中的上述方法,包括:Kubernetes集群将服务的配置信息发送至Nginx管理器,Nginx管理器依据所述服务的配置信息,为所述Kubernetes集群中的目标服务创建配置文件;并依据预先设置的负载均衡策略,确定与所述目标服务对应的访问数据流;将与所述目标服务对应的配置文件以及访问数据流发送至Nginx服务器;Nginx服务器依据所述配置文件,将所述访问数据流转发至Kubernetes集群中的目标服务。Nginx管理器可以依据预先设置的负载均衡策略,确定与目标服务对应的访问数据流,实现对Kubernetes集群中目标服务对应的访问数据流的统一均衡分配,从而实现Kubernetes集群中服务的负载均衡,进而实现了对Kubernetes集群中服务的统一管理,同时也实现了访问数据流的统一分配。

[0088] 本发明实施例通过将Kubernetes集群中的服务统一注册到Nginx管理器中,无需借助第三方的注册中心,即可完成服务的注册。

[0089] 通过上述方法,本发明实施例可以实现多个Kubernetes集群中服务的负载均衡,可以将访问数据流分配到多个集群中,在某个集群出现故障之后,还可以往其他未出现故障的集群中分配访问数据流,保障Kubernetes集群内部服务的高可用性。

[0090] 可选的,本发明实施例还提供了连接Nginx管理器的终端可视化界面,用户可以通过终端可视化界面查看Kubernetes集群中各个服务的运行状况以及访问数据流的分配情况,并可以通过该可视化界面,直观的呈现Kubernetes集群中各个服务配置和Nginx配置之间存在的差异,提高了用户体验水平。

[0091] 本发明实施例中的Nginx管理器还可以监控Nginx的外部请求量,包括正常的访问请求以及HTTP返回码为4xx、5xx的异常请求流量。

[0092] 为了进一步提高负载均衡效果,本发明实施例提出如下另外一种负载均衡方法,图5为本发明实施例提供的负载均衡方法的流程图,该方法可应用于所述Nginx管理器,参照图5,所述方法可以包括:

[0093] 步骤S200、从Kubernetes集群中接收服务的配置信息;

[0094] 步骤S210、依据所述服务的配置信息,为所述Kubernetes集群中的目标服务创建配置文件;

[0095] 步骤S220、依据预先设置的负载均衡策略,确定与所述目标服务对应的访问数据流;

[0096] 步骤S230、将与所述目标服务对应的配置文件以及访问数据流发送至Nginx服务器,使所述Nginx服务器依据所述配置文件,将所述访问数据流转发至Kubernetes集群中的目标服务;

[0097] 本发明实施例中的步骤S200-步骤S230与上述步骤中的步骤S100-步骤S130相同,具体内容可以相互对应参照,本发明实施例在此不做详细说明。

[0098] 为了实现对Nginx服务器的监测,本发明实施例中的Nginx管理器还会向Nginx服务器发送心跳检查指令,并依据所述Nginx服务器的响应信息,监测所述Nginx服务器的运行状况,在监测到Nginx服务器运行出现故障的情况下,及时发出异常告警提示,确保Nginx集群的正常运行。

[0099] 同时,Nginx服务器如果检测到发送给Kubernetes集群中的访问数据流发送失败,也会及时向Nginx管理器发送告警信息,提示访问数据流下发失败,请求进行重新发送或进行其他处理。

[0100] 步骤S240、向Nginx服务器发送流量监控指示信息,使所述Nginx服务器对发送至Kubernetes集群中的访问数据流的流量进行监控。

[0101] 通过向Nginx服务器发送流量监控指示信息的方式,所述Nginx服务器对发送至Kubernetes集群中的访问数据流的流量进行监控,具体的,所要监控的流量既包括了Kubernetes集群与Kubernetes集群之间的访问流量,又包括了Kubernetes系统外部的访问流量。在Nginx服务器监控到某一Kubernetes集群对应的访问数据流的流量达到该Kubernetes集群对应的访问流量阈值的情况下,则向所述Nginx管理器发送指示信息,指示所述Nginx管理器调整针对该Kubernetes集群的负载均衡策略,例如Nginx管理器可以依据所述指示信息,限时部分访问数据流,例如限制低优先级业务访问数据流的下发,仅允许高优先级业务访问数据流的下发。

[0102] 本发明实施例中,Nginx服务器对发送至Kubernetes集群中的访问数据流的流量进行监控,并将监控结果向所述Nginx管理器发送,从而使得Nginx管理器及时对Kubernetes集群的负载均衡策略做出调整,即通过反馈调节的方式,提高了负载均衡策略调节的准确性。

[0103] 下面对本发明实施例提供的负载均衡装置进行介绍,下文描述的负载均衡装置可与上文描述的负载均衡方法相互对应参照。

[0104] 图6为本发明实施例提供的负载均衡装置的结构框图,该负载均衡装置具体可以

为Nginx管理器,参照图6,该负载均衡装置可以包括:

[0105] 配置信息接收单元300,用于从Kubernetes集群中接收服务的配置信息;

[0106] 配置文件创建单元310,用于依据所述服务的配置信息,为所述Kubernetes集群中的目标服务创建配置文件;

[0107] 访问数据流确定单元320,用于依据预先设置的负载均衡策略,确定与所述目标服务对应的访问数据流;

[0108] 发送单元330,用于将与所述目标服务对应的配置文件以及访问数据流发送至Nginx服务器,使所述Nginx服务器依据所述配置文件,将所述访问数据流转发至Kubernetes集群中的目标服务。

[0109] 所述配置文件创建单元310包括:

[0110] 目标服务确定单元,用于依据所述服务的配置信息,确定需要在Nginx管理器中创建配置文件的目标服务;

[0111] 配置文件创建子单元,用于依据所述目标服务的配置信息,为所述目标服务创建配置文件。

[0112] 所述负载均衡装置还包括:

[0113] 更新配置信息接收单元,用于从所述Kubernetes集群中接收目标服务的更新配置信息;

[0114] 更新单元,用于利用所述目标服务的更新配置信息,更新所述目标服务对应的配置文件。

[0115] 所述负载均衡装置还包括:

[0116] 监控指示信息发送单元,用于向Nginx服务器发送流量监控指示信息,使所述Nginx服务器对发送至Kubernetes集群中的访问数据流的流量进行监控。

[0117] 所述目标服务对应的访问数据流包括:

[0118] Kubernetes系统中Kubernetes集群之间的访问数据流和/或所述Kubernetes系统外部的访问数据流。

[0119] 综上,本发明提供了一种负载均衡方法、装置及系统,该方法包括:Kubernetes集群将服务的配置信息发送至Nginx管理器,Nginx管理器依据所述服务的配置信息,为所述Kubernetes集群中的目标服务创建配置文件;并依据预先设置的负载均衡策略,确定与所述目标服务对应的访问数据流;将与所述目标服务对应的配置文件以及访问数据流发送至Nginx服务器;Nginx服务器依据所述配置文件,将所述访问数据流转发至目标服务。Nginx管理器可以依据预先设置的负载均衡策略,确定与目标服务对应的访问数据流,实现对Kubernetes集群中目标服务对应的访问数据流的统一均衡分配,从而实现Kubernetes集群中服务的负载均衡。

[0120] 本说明书中各个实施例采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似部分互相参见即可。对于实施例公开的装置而言,由于其与实施例公开的方法相对应,所以描述的比较简单,相关之处参见方法部分说明即可。

[0121] 专业人员还可以进一步意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、计算机软件或者二者的结合来实现,为了清楚地说明硬件和

软件的可互换性,在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本发明的范围。

[0122] 结合本文中公开的实施例描述的方法或算法的步骤可以直接用硬件、处理器执行的软件模块,或者二者的结合来实施。软件模块可以置于随机存储器(RAM)、内存、只读存储器(ROM)、电可编程ROM、电可擦除可编程ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。

[0123] 对所公开的实施例的上述说明,使本领域专业技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下,在其它实施例中实现。因此,本发明将不会被限制于本文所示的这些实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

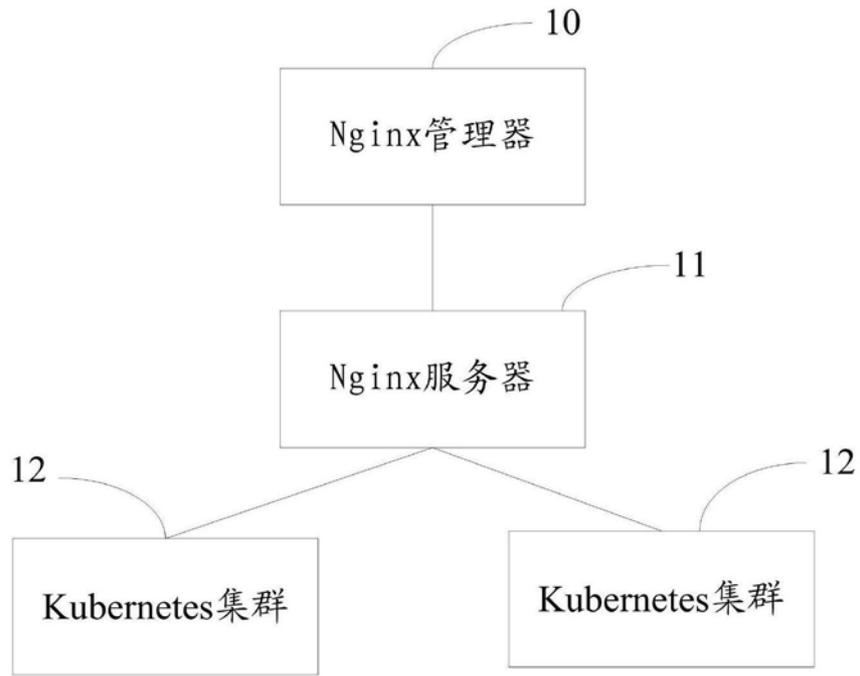


图1

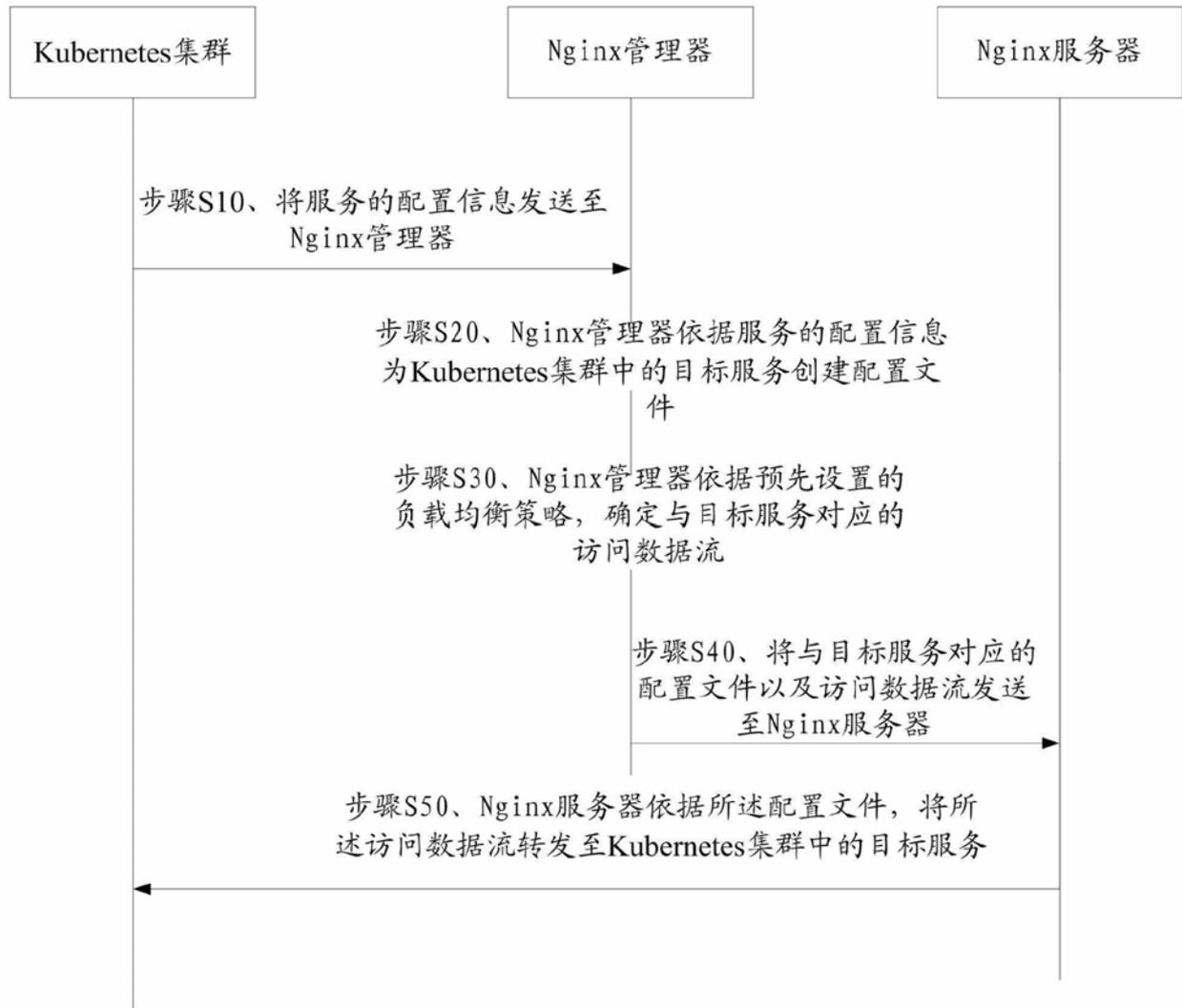


图2

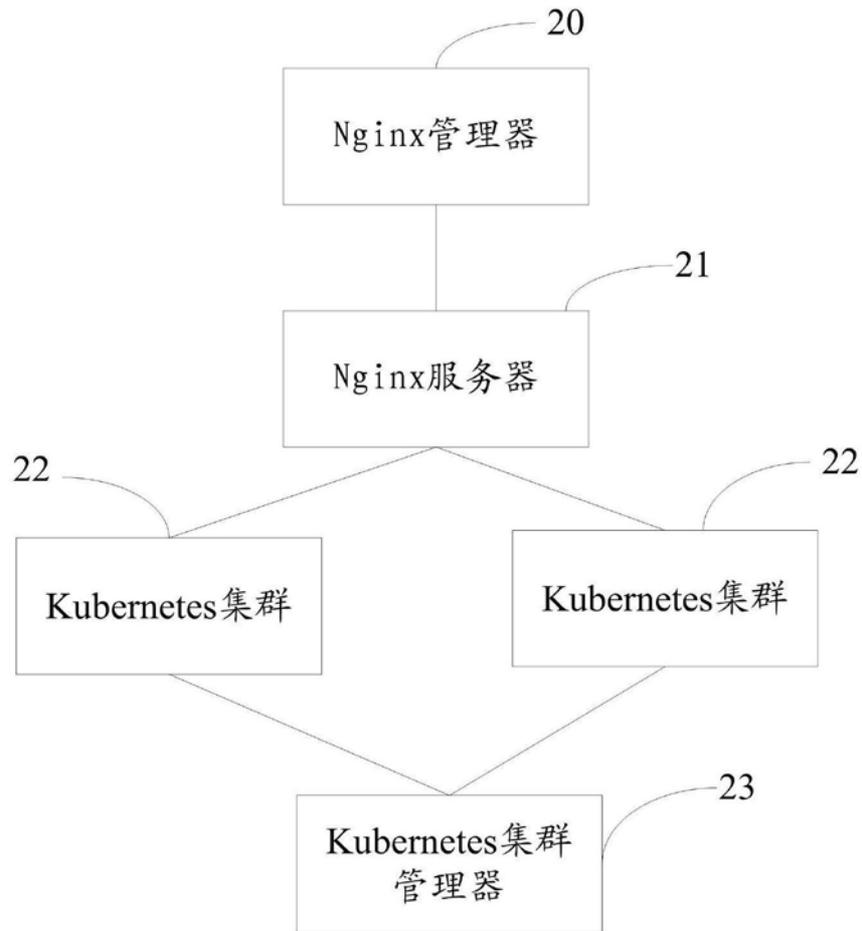


图3

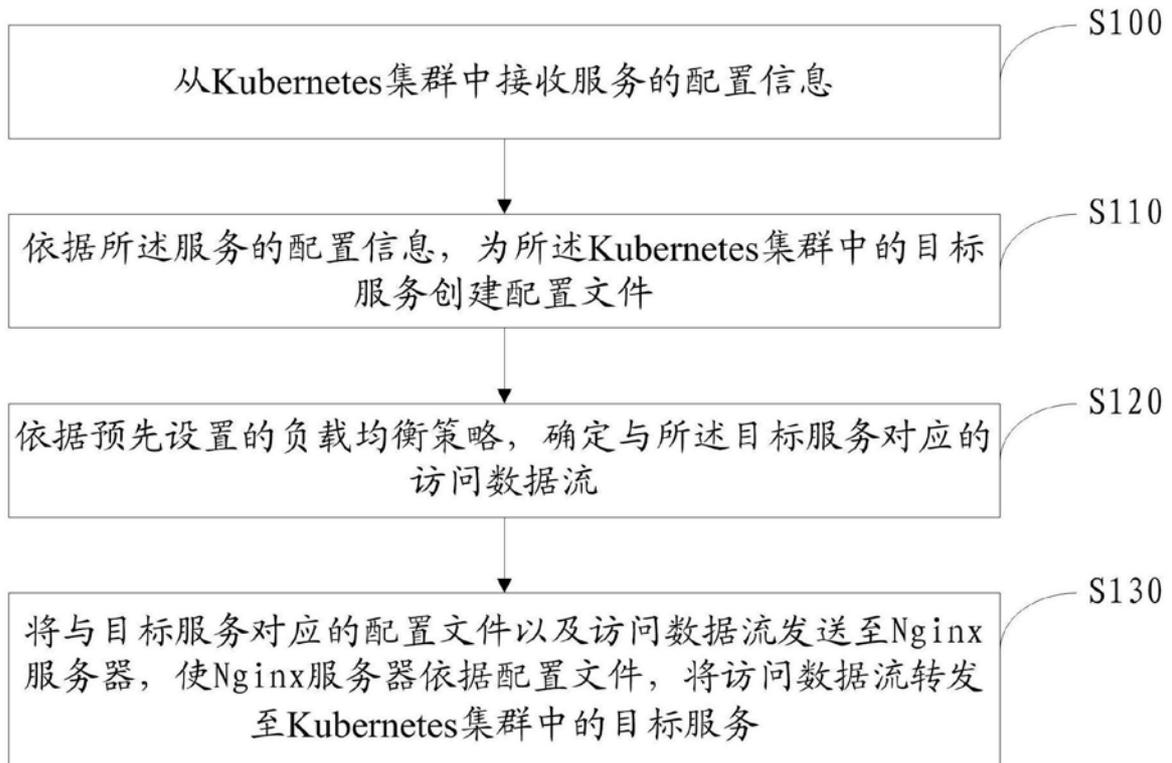


图4

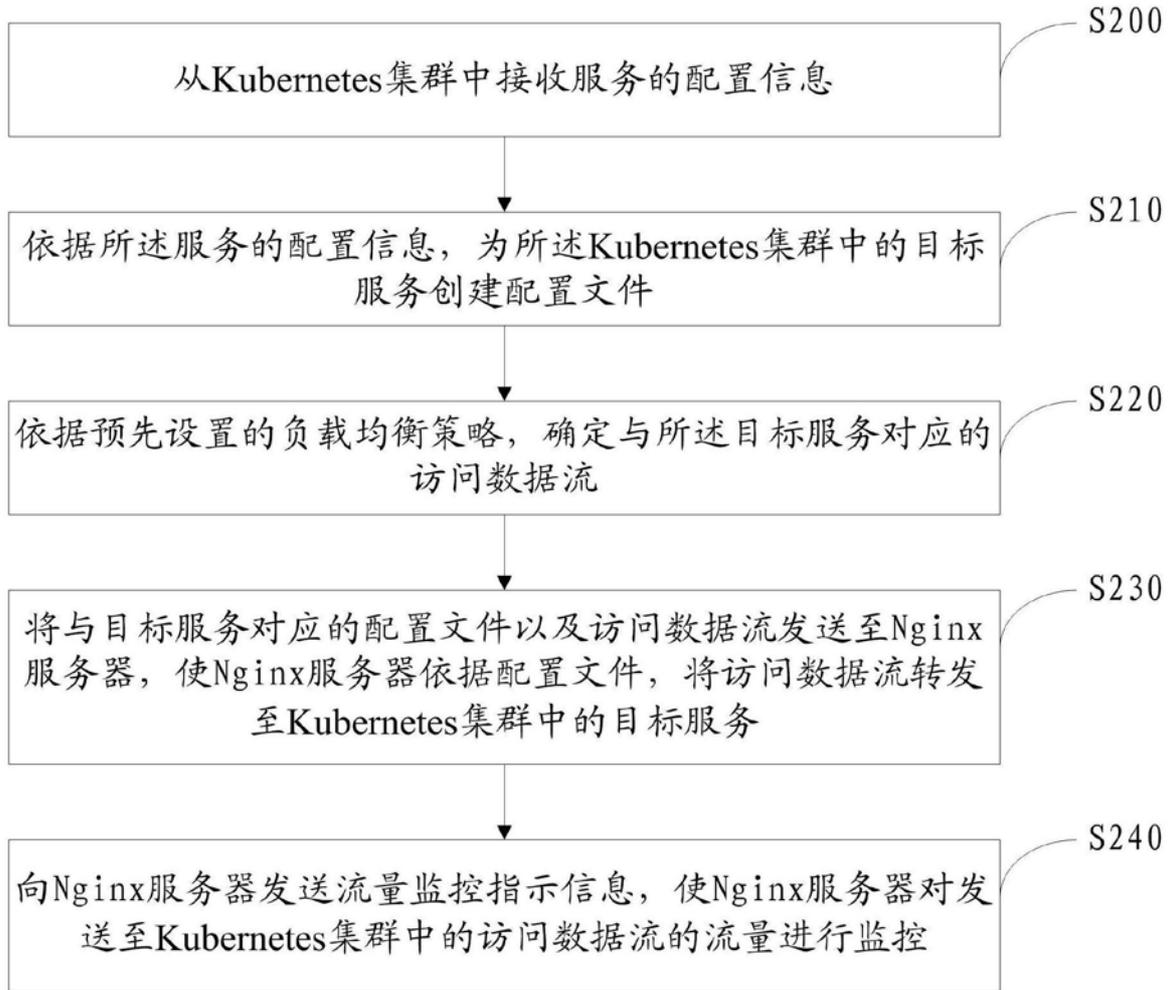


图5



图6