

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2023 年 10 月 19 日 (19.10.2023)



(10) 国际公布号
WO 2023/197874 A1

(51) 国际专利分类号:
H04L 47/10 (2022.01) **G06F 11/20** (2006.01)

(21) 国际申请号: PCT/CN2023/084815

(22) 国际申请日: 2023 年 3 月 29 日 (29.03.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202210371384.2 2022年4月11日 (11.04.2022) CN

(71) 申请人: 苏州浪潮智能科技有限公司
(**INSUR SUZHOU INTELLIGENT TECHNOLOGY CO., LTD.**) [CN/CN]; 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。

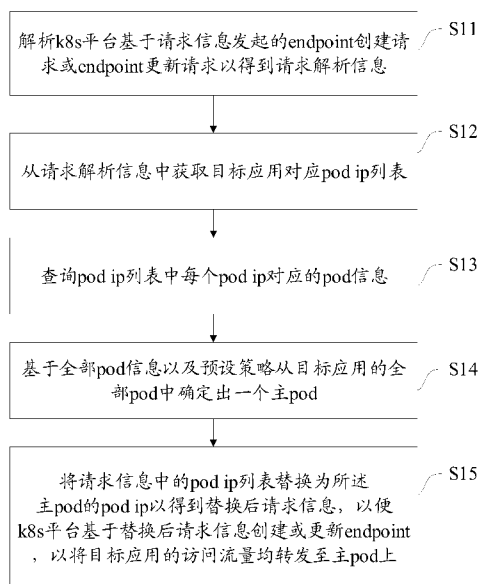
(72) 发明人: 黄吉旺 (**HUANG, Jiwang**); 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。 吴栋 (**WU, Dong**); 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。 王玉东 (**WANG, Yudong**); 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。

(74) 代理人: 北京润泽恒知识产权代理有限公司 (**BEIJING RUN ZEHENG INTELLECTUAL PROPERTY LAW FIRM**); 中国北京市海淀区中关村南大街甲18号北京国际C座6层606, Beijing 100081 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG,

(54) **Title:** APPLICATION ACCESS METHOD AND APPARATUS, AND DEVICE AND MEDIUM

(54) 发明名称: 一种应用访问方法、装置、设备及介质



S11 Parse an endpoint creation request or an endpoint update request, which is initiated by a k8s platform on the basis of request information, so as to obtain request parsing information

S12 Acquire, from the request parsing information, a pod ip list corresponding to a target application

S13 Query pod information corresponding to each pod ip in the pod ip list

S14 On the basis of all pieces of pod information and a preset policy, determine a master pod from all pods of the target application

S15 Replace the pod ip list in the request information with a pod ip of the master pod, so as to obtain request information which has been subjected to replacement, such that the k8s platform creates or updates an endpoint on the basis of the request information which has been subjected to replacement, so as to forward all access traffic of the target application to the master pod

图 1

(57) **Abstract:** Disclosed in the present application are an application access method and apparatus, and a device and a medium, which are applied to the technical field of application access. The method comprises: parsing an endpoint creation request or an endpoint update request, which is initiated by a k8s platform on the basis of request information, so as to obtain request parsing information; acquiring, from the request parsing information, a pod ip list corresponding to a target application; querying pod information corresponding to each pod ip in the pod ip list; on the basis of all pieces of pod information and a preset policy, determining a master pod from all

BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

Pods of the target application; and replacing the pod ip list in the request information with a pod ip of the master pod, so as to obtain request information which has been subjected to replacement, such that the k8s platform creates or updates an endpoint on the basis of the request information which has been subjected to replacement, so as to forward all access traffic of the target application to the master pod. Therefore, the complexity of application development can be reduced, and the response performance of application access can also be improved.

(57) 摘要: 本申请公开了一种应用访问方法、装置、设备及介质, 应用于应用访问技术领域, 包括: 解析k8s平台基于请求信息发起的endpoint创建请求或endpoint更新请求以得到请求解析信息; 从请求解析信息中获取目标应用对应pod ip列表; 查询pod ip列表中每个pod ip对应的pod信息; 基于全部pod信息以及预设策略从目标应用的全部pod中确定出一个主pod; 将请求信息中的pod ip列表替换为主pod的pod ip以得到替换后请求信息, 以便k8s平台基于替换后请求信息创建或更新endpoint, 以将目标应用的访问流量均转发至主pod上。能够降低应用开发的复杂度, 并且提升应用访问的响应性能。

一种应用访问方法、装置、设备及介质

相关申请的交叉引用

本申请要求于 2022 年 04 月 11 日提交中国专利局，申请号为 202210371384.2，申请名称为“一种应用访问方法、装置、设备及介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及应用访问技术领域，特别涉及一种应用访问方法、装置、设备及介质。

背景技术

在生产环境中，应用的高可用通常通过多副本来实现，在容器平台上，表现为多个 pod，并有多主、主从等不同的请求处理机制。其中，主从的机制表现为其中一个副本作为主副本角色响应对该应用的请求，其余副本作为从副本角色运行，在主副本异常时由其中一个从副本变为主副本来接管流量。这种机制能够保证原子性，避免多个副本同时响应请求可能带来的数据处理不一致问题。主从的技术实现方案一般为在源码中开发或集成一套选举机制，通过持续运行该选举进程不停的维持某副本的主副本角色，由该主副本响应请求，当主副本失联时再从备份副本中重新选举出新的主副本，当流量到达从副本时，从副本需要把流量转发给主副本，避免从副本操作数据。这种方案不但技术原理复杂，不易维护，并且存在从副本要再次转发流量给主副本带来的响应性能问题。

发明内容

有鉴于此，本申请的目的在于提供一种应用访问方法、装置、设备及介质，能够降低应用开发的复杂度，并且提升应用访问的响应性能。其具体方案如下：

本申请的一些实施例公开了一种应用访问方法，应用于预设准入控制器，包括：

解析 k8s 平台基于请求信息发起的 endpoint 创建请求或 endpoint 更新请求以得到请求解析信息；

从请求解析信息中获取目标应用对应 pod ip 列表；

查询 pod ip 列表中每个 pod ip 对应的 pod 信息；

基于全部 pod 信息以及预设策略从目标应用的全部 pod 中确定出一个主 pod；

将请求信息中的 pod ip 列表替换为主 pod 的 pod ip 以得到替换后请求信息，以便 k8s

平台基于替换后请求信息创建或更新 endpoint，以将目标应用的访问流量均转发至主 pod 上。

在本申请的一些实施例中，还包括：

从请求解析信息中获取 endpoint 名称；

相应的，在查询 pod ip 列表中每个 pod ip 对应的 pod 信息之前，还包括：

基于 endpoint 名称查询 service 信息；其中，service 信息为 endpoint 对应的 service 的信息；

判断 service 信息中是否存在预设主从标签；

若存在预设主从标签且预设主从标签的值为第一预设值，则触发查询 pod ip 列表中每个 pod ip 对应的 pod 信息的步骤。

在本申请的一些实施例中，还包括：

若不存在预设主从标签，或存在预设主从标签且预设主从标签的值为第二预设值，则不进行任何处理。

在本申请的一些实施例中，基于全部 pod 信息以及预设策略从目标应用的全部 pod 中确定出一个主 pod，包括：

基于全部 pod 信息以及预设策略从目标应用的全部 pod 中确定出创建时间最早的一个 pod，作为主 pod。

在本申请的一些实施例中，基于全部 pod 信息以及预设策略从目标应用的全部 pod 中确定出一个主 pod，包括：

基于全部 pod 信息以及预设策略从目标应用的全部 pod 中确定出创建时间最晚的一个 pod，作为主 pod。

在本申请的一些实施例中，解析 k8s 平台基于请求信息发起的 endpoint 创建请求或 endpoint 更新请求之前，还包括：

基于 admission webhook 机制获取 k8s 平台发起的 endpoint 创建请求或 endpoint 更新请求。

本申请的一些实施例还公开了一种应用访问装置，应用于预设准入控制器，包括：

请求解析模块，用于解析 k8s 平台基于请求信息发起的 endpoint 创建请求或 endpoint 更新请求以得到请求解析信息；

列表获取模块，用于从请求解析信息中获取目标应用对应 pod ip 列表；

信息查询模块，用于查询 pod ip 列表中每个 pod ip 对应的 pod 信息；

主 pod 确定模块，用于基于全部 pod 信息以及预设策略从目标应用的全部 pod 中确定出一个主 pod；

信息更新模块，用于将请求信息中的 pod ip 列表替换为主 pod 的 pod ip 以得到替换后请求信息，以便 k8s 平台基于替换后请求信息创建或更新 endpoint，以将目标应用的访问流量均转发至主 pod 上。

在本申请的一些实施例中，还包括：

endpoint 名称获取模块，用于从请求解析信息中获取 endpoint 名称；

service 信息查询模块，用于基于 endpoint 名称查询 service 信息；其中，service 信息为 endpoint 对应的 service 的信息；

主从标签判断模块，用于判断 service 信息中是否存在预设主从标签；若存在预设主从标签且预设主从标签的值为第一预设值，则触发信息查询模块执行查询 pod ip 列表中每个 pod ip 对应的 pod 信息的步骤。

本申请的一些实施例还公开了一种电子设备，包括处理器和存储器；其中，

存储器，用于保存计算机程序；

处理器，用于执行计算机程序以实现前述的应用访问方法。

本申请的一些实施例还公开了一种计算机非易失性可读存储介质，用于保存计算机程序，其中，计算机程序被处理器执行时实现前述的应用访问方法。

可见，本申请先解析 k8s 平台基于请求信息发起的 endpoint 创建请求或 endpoint 更新请求以得到请求解析信息，之后从请求解析信息中获取目标应用对应 pod ip 列表，并查询 pod ip 列表中每个 pod ip 对应的 pod 信息，然后基于全部 pod 信息以及预设策略从目标应用的全部 pod 中确定出一个主 pod，最后将请求信息中的 pod ip 列表替换为主 pod 的 pod ip 以得到替换后请求信息，以便 k8s 平台基于替换后请求信息创建或更新 endpoint，以将目标应用的访问流量均转发至主 pod 上。也即，本申请通过预设准入控制器对 k8s 平台基于请求信息发起的 endpoint 创建请求或 endpoint 更新请求进行解析，得到请求解析信息，并基于请求解析信息和预设策略确定出一个主 pod，利用该主 pod 的 pod ip 替换掉请求信息中的 pod ip 列表，k8s 平台基于该请求信息创建或更新的 endpoint，能够将目标应用的访问流量均转发至主 pod 上，避免了复杂且不易维护的选举机制，以及从副本的转发流量带来的性能问题，能够降低应用开发的复杂度，并且提升应用访问的响应性能。

附图说明

为了更清楚地说明本申请一些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申请的实施例，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。

图 1 为本申请一些实施例公开的一种应用访问方法流程图；

图 2 为本申请一些实施例提供的一种具体的应用访问方案示意图；

图 3 为本申请一些实施例公开的一种应用访问装置结构示意图；

图 4 为本申请一些实施例公开的一种电子设备结构图；

图 5 为本申请一些实施例所提供的一种计算机非易失性可读存储介质的结构示意图。

具体实施方式

下面将结合本申请一些实施例中的附图，对本申请一些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申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，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，都属于本申请保护的范围。

首先，对本申请方案涉及到的关键术语进行解释：kubernetes：容器云平台通用的开源容器编排架构，能够实现容器的自动调度、扩展、故障恢复等，简称k8s。kube-apiserver：k8s的对外服务API（Application Programming Interface, 应用程序编程接口），k8s中所有资源的相关操作需要调用kube-apiserver相关接口来实现。工作负载：k8s上应用组件的部署形式，包括deployment（无状态负载）/statefulset（有状态负载）/job（任务）/cronjob（定时任务）/daemonset（守护进程）等多种形式。每个工作负载可以管理1个或多个pod。pod：k8s的最小管理单元，由1个或多个容器组成，这些容器共享进程、网络等命名空间。service：k8s为满足对工作负载的请求而设计的一个对象资源，通过labelselector匹配工作负载下的pod，并提供负载均衡能力将流量转发到不同pod。endpoint：一组pod ip和端口的集合，当service创建后，根据service的labelselector选择对应工作负载的pod。对service的访问将通过负载均衡后负载到endpoint上的某一个pod ip和端口上。

目前，主从的技术实现方案一般为在源码中开发或集成一套选举机制，通过持续运行该选举进程不停的维持某副本的主副本角色，由该主副本响应请求，当主副本失联时再从备份副本中重新选举出新的主副本，当流量到达从副本时，从副本需要把流量转发给主副本，避

免从副本操作数据。这种方案不但技术原理复杂，不易维护，并且存在从副本要再次转发流量给主副本带来的响应性能问题。为此，本申请公开了一种应用访问方案，能够降低应用开发的复杂度，并且提升应用访问的响应性能。

参见图1所示，本申请一些实施例公开了一种应用访问方法，应用于预设准入控制器，包括：

步骤S11：解析k8s平台基于请求信息发起的endpoint创建请求或endpoint更新请求以得到请求解析信息。

其中，请求信息包括k8s平台查找到的目标应用的所有pod的pod ip，以及待创建或待更新endpoint的endpoint名称。在具体的实施方式中，本申请一些实施例基于admission webhook机制（准入控制插件的一种，可以用于拦截所有向APISERVER发送的请求，并且可以修改请求或拒绝请求）获取k8s平台发起的endpoint创建请求或endpoint更新请求。

需要指出的是，当前大部分容器云平台均基于k8s框架，面向用户提供以应用管理为核心的功能体系。容器化应用以工作负载的形式部署在k8s集群上，每个工作负载可以同时运行1个或多个实例（pod）。K8s提供service机制实现对工作负载的访问，每个service有对应的集群外或内部的访问地址，并且提供负载均衡功能来实现对多个pod的负载访问。每创建一个service后，k8s会实时创建一个同名的endpoint资源，该资源包含service要负载到的所有pod ip和端口，对service的访问将通过负载均衡后负载到endpoint上的某一个pod ip和端口上。进一步的，k8s提供admission webhook机制，允许在创建资源的过程中，调用自定义的admission controller，进行校验、修改动作，基于此，本申请一些实施例基于k8s的service、endpoint资源管理并结合webhook机制，开发主从高可用的准入控制器，得到预设准入控制器，并且，预先将预设准入控制器部署在容器平台上，并注册到k8s平台中。k8s在创建endpoint时，调用预设准入控制器，将endpoint创建请求传入预设准入控制器。

步骤S12：从请求解析信息中获取目标应用对应pod ip列表。

进一步的，本申请一些实施例还可以从请求解析信息中获取endpoint名称；基于endpoint名称查询service信息；其中，service信息为endpoint对应的service的信息；判断service信息中是否存在预设主从标签；若存在预设主从标签且预设主从标签的值为第一预设值，则触发查询pod ip列表中每个pod ip对应的pod信息的步骤。若不存在预设主从标签，或存在预设主从标签且预设主从标签的值为第二预设值，则不进行任何处理。

需要指出的是，endpoint名称即为service名称，因此，可以基于endpoint名称查询

service信息。并且，应用开发人员在为应用创建service时，可以根据是否有主从高可用的需求，为service添加一个标签即预设主从标签。在预设准入控制器中可以判断service是否有此标签，并针对有此标签的service，再去执行后续的主pod选择过程，并且，可以根据需求，通过对此标签配置不同的值，开启或关闭后续主pod选择过程，达到配置简单、不影响其他应用的效果。比如，第一预设值为true，第二预设值为false。

步骤S13：查询pod ip列表中每个pod ip对应的pod信息。

其中，pod信息可以包括创建时间、名称等。

步骤S14：基于全部pod信息以及预设策略从目标应用的全部pod中确定出一个主pod。

在一种实施方式中，可以基于全部pod信息以及预设策略从目标应用的全部pod中确定出创建时间最早的一个pod，作为主pod。可以理解的是，最早创建的pod比较稳定。

在另一种实施方式中，可以基于全部pod信息以及预设策略从目标应用的全部pod中确定出创建时间最晚的一个pod，作为主pod。可以理解的是，最晚创建的pod为基于最新应用程序创建的pod，便于提供更新的应用服务。

在具体的实施方式中，可以根据创建时间对全部pod进行排序，根据排序结果确定主pod。

当然，在一些实施例中，还可以通过其他策略确定主pod。

步骤S15：将请求信息中的pod ip列表替换为主pod的pod ip以得到替换后请求信息，以便k8s平台基于替换后请求信息创建或更新endpoint，以将目标应用的访问流量均转发至主pod上。

需要指出的是，本申请中，请求信息可以为全局变量，这样，预设准入控制器可以直接对请求信息进行修改。

其中，当k8s平台监测到目标应用的pod发生变化时，则发起endpoint更新请求；当目标应用中任一pod出现异常、或被删除、或新增了pod，则判定目标应用的pod发生变化。也即，本申请一些实施例在为service创建endpoint的过程中，调用预设准入控制器，根据当前endpoint创建请求中的所有pod ip，查询关联的pod信息，并根据预设策略从所有pod中筛选一个主pod，使用该pod的pod ip替换请求信息中的pod ip列表，从而实现通过service的请求流量只转发到该pod，该pod将作为主副本响应请求。当该应用对应的pod发生变化时，k8s会实时自动触发endpoint更新请求，同样会通过admission webhook机制调用预设准入控制器，基于创建endpoint时的同样逻辑再次选择新的pod作为主pod。这样，基于k8s的service/endpoint资源管理及admission webhook机制，通过开发webhook程序，在为

service创建或更新endpoint的过程中，根据预置的策略选择某一个pod ip作为endpoint中的pod ip，也即选择了一个pod作为service负载的后端，从而实现流量直接转发到该pod。这样，实现了多副本应用的主从高可用访问机制，无需应用自主开发主从选举机制，应用可以专注于应用本身业务逻辑，避免流量到从副本转发带来的性能问题，并且一次开发部署，可以给集群所有有需求的应用使用。整体方案简洁灵活，易于扩展，有新的pod选择策略直接修改控制器处理逻辑即可，具有良好的复用性和可扩展性。

例如，参见图2所示，图2为本申请一些实施例公开的一种具体的应用访问方案示意图。客户端发起创建service请求，并为service添加主从处理标签，以及将service的labelselector设置为应用的label。k8s接收到该请求，创建service，并根据labelselector查询该应用对应的所有pod，得到pod ip列表，进而确定请求信息，基于请求信息发起endpoint创建请求，并调用已注册好的admission controller（预设准入控制器），在检测到有pod异常、删除、或新增，基于当前的请求信息自动发起endpoint更新请求，预设准入控制器的业务处理逻辑为：解析endpoint创建或更新请求，得到请求解析信息；从请求解析信息中获取endpoint名称、pod ip列表等；根据endpoint名称查询service信息；判断service是否存在主从标签，如果存在该标签并且值为true，说明该服务需要主从访问，则进行后续的处理逻辑，如果不存在该标签，说明该服务不需要主从访问，则不再继续执行，直接完成。后续的处理逻辑包括：根据pod ip列表查询所有pod信息，按创建时间由早到晚排序，获取第一个pod ip；用筛选出的pod ip替换原请求信息中的ip列表，以便k8基于更新的请求信息创建或者更新endpoint。

可见，本申请一些实施例先解析k8s平台基于请求信息发起的endpoint创建请求或endpoint更新请求以得到请求解析信息，之后从请求解析信息中获取目标应用对应pod ip列表，并查询pod ip列表中每个pod ip对应的pod信息，然后基于全部pod信息以及预设策略从目标应用的全部pod中确定出一个主pod，最后将请求信息中的pod ip列表替换为主pod的pod ip以得到替换后请求信息，以便k8s平台基于替换后请求信息创建或更新endpoint，以将目标应用的访问流量均转发至主pod上。也即，本申请一些实施例通过预设准入控制器对k8s平台基于请求信息发起的endpoint创建请求或endpoint更新请求进行解析，得到请求解析信息，并基于请求解析信息和预设策略确定出一个主pod，利用该主pod的pod ip替换掉请求信息中的pod ip列表，k8s平台基于该请求信息创建或更新的endpoint，能够将目标应用的访问流量均转发至主pod上，避免了复杂且不易维护的选举机制，以及从副本的转发流量带来的性能问题，能够降低应用开发的复杂度，并且提升应用访问的响应性能。

参见图3所示，本申请一些实施例公开了一种应用访问装置，应用于预设准入控制器，包括：

请求解析模块11，用于解析k8s平台基于请求信息发起的endpoint创建请求或endpoint更新请求以得到请求解析信息；

列表获取模块12，用于从请求解析信息中获取目标应用对应pod ip列表；

信息查询模块13，用于查询pod ip列表中每个pod ip对应的pod信息；

主pod确定模块14，用于基于全部pod信息以及预设策略从目标应用的全部pod中确定出一个主pod；

信息更新模块15，用于将请求信息中的pod ip列表替换为主pod的pod ip以得到替换后请求信息，以便k8s平台基于替换后请求信息创建endpoint或更新，以将目标应用的访问流量均转发至主pod上。

可见，本申请一些实施例先解析k8s平台基于请求信息发起的endpoint创建请求或endpoint更新请求以得到请求解析信息，之后从请求解析信息中获取目标应用对应pod ip列表，并查询pod ip列表中每个pod ip对应的pod信息，然后基于全部pod信息以及预设策略从目标应用的全部pod中确定出一个主pod，最后将请求信息中的pod ip列表替换为主pod的pod ip以得到替换后请求信息，以便k8s平台基于替换后请求信息创建或更新endpoint，以将目标应用的访问流量均转发至主pod上。也即，本申请一些实施例通过预设准入控制器对k8s平台基于请求信息发起的endpoint创建请求或endpoint更新请求进行解析，得到请求解析信息，并基于请求解析信息和预设策略确定出一个主pod，利用该主pod的pod ip替换掉请求信息中的pod ip列表，k8s平台基于该请求信息创建或更新的endpoint，能够将目标应用的访问流量均转发至主pod上，避免了复杂且不易维护的选举机制，以及从副本的转发流量带来的性能问题，能够降低应用开发的复杂度，并且提升应用访问的响应性能。

在本申请一些实施例中，应用访问装置还包括：endpoint名称获取模块，用于从请求解析信息中获取endpoint名称。

相应的，应用访问装置还包括：

service信息查询模块，用于基于endpoint名称查询service信息；其中，service信息为endpoint对应的service的信息；

主从标签判断模块，用于判断service信息中是否存在预设主从标签；若存在预设主从标签且预设主从标签的值为第一预设值，则触发信息查询模块13执行查询pod ip列表中每个

pod ip对应的pod信息的步骤。若不存在预设主从标签，或存在预设主从标签且预设主从标签的值为第二预设值，则不进行任何处理。

在一种实施方式中，主pod确定模块14，具体用于基于全部pod信息以及预设策略从目标应用的全部pod中确定出创建时间最早的一个pod，作为主pod。

在另一种实施方式中，主pod确定模块14，具体用于基于全部pod信息以及预设策略从目标应用的全部pod中确定出创建时间最晚的一个pod，作为主pod。

在本申请一些实施例中，应用访问装置还包括请求获取模块，用于基于admission webhook机制获取k8s平台发起的endpoint创建请求或endpoint更新请求。

参见图4所示，本申请一些实施例公开了一种电子设备20，包括处理器21和存储器22；其中，存储器22，用于保存计算机程序；处理器21，用于执行计算机程序，前述实施例公开的应用访问方法。

关于上述应用访问方法的具体过程可以参考前述实施例中公开的相应内容，在此不再进行赘述。

并且，存储器22作为资源存储的载体，可以是只读存储器、随机存储器、磁盘或者光盘等，存储方式可以是短暂存储或者永久存储。

另外，电子设备20还包括电源23、通信接口24、输入输出接口25和通信总线26；其中，电源23用于为电子设备20上的各硬件设备提供工作电压；通信接口24能够为电子设备20创建与外界设备之间的数据传输通道，其所遵循的通信协议是能够适用于本申请技术方案的任意通信协议，在此不对其进行具体限定；输入输出接口25，用于获取外界输入数据或向外界输出数据，其具体的接口类型可以根据具体应用需要进行选取，在此不进行具体限定。

进一步的，本申请一些实施例还公开了一种计算机非易失性可读存储介质，用于保存计算机程序，如图5所示，计算机非易失性可读存储介质5上存储有计算机程序51，其中，计算机程序51被处理器执行时实现前述实施例公开的应用访问方法。

关于上述应用访问方法的具体过程可以参考前述实施例中公开的相应内容，在此不再进行赘述。

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它实施例的不同之处，各个实施例之间相同或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。

结合本文中所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，Random Access Memory）、内存、只读存储器（ROM，Read-Only Memory）、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。

以上对本申请所提供的一种应用访问方法、装置、设备及介质进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，综上，本说明书内容不应理解为对本申请的限制。

权 利 要 求 书

1、一种应用访问方法，其特征在于，应用于预设准入控制器，包括：

解析 k8s 平台基于请求信息发起的 endpoint 创建请求或 endpoint 更新请求以得到请求解析信息；

从所述请求解析信息中获取目标应用对应 pod ip 列表；

查询所述 pod ip 列表中每个 pod ip 对应的 pod 信息；

基于全部所述 pod 信息以及预设策略从所述目标应用的全部 pod 中确定出一个主 pod；

将所述请求信息中的所述 pod ip 列表替换为所述主 pod 的 pod ip 以得到替换后请求信息，以便所述 k8s 平台基于所述替换后请求信息创建或更新 endpoint，以将所述目标应用的访问流量均转发至所述主 pod 上。

2、根据权利要求 1 所述的应用访问方法，其特征在于，所述请求信息包括 k8s 平台查找到的目标应用的所有 pod 的 pod ip，以及待创建或待更新 endpoint 的 endpoint 名称。

3、根据权利要求 1 所述的应用访问方法，其特征在于，还包括：

从所述请求解析信息中获取 endpoint 名称；

相应的，在所述查询所述 pod ip 列表中每个 pod ip 对应的 pod 信息之前，还包括：

基于所述 endpoint 名称查询 service 信息；其中，所述 service 信息为所述 endpoint 对应的 service 的信息；

判断所述 service 信息中是否存在预设主从标签；

若存在所述预设主从标签且所述预设主从标签的值为第一预设值，则触发所述查询所述 pod ip 列表中每个 pod ip 对应的 pod 信息的步骤。

4、根据权利要求 3 所述的应用访问方法，其特征在于，还包括：

若不存在所述预设主从标签，或存在所述预设主从标签且所述预设主从标签的值为第二预设值，则不进行任何处理。

5、根据权利要求 3 所述的应用访问方法，其特征在于，所述 endpoint 名称为 service 名称。

6、根据权利要求 3 所述的应用访问方法，其特征在于，所述预设主从标签经由客户端在向 k8s 平台发起创建 service 请求时为 service 进行添加。

7、根据权利要求 1 所述的应用访问方法，其特征在于，所述 pod 信息包括创建时

间、名称。

8、根据权利要求 7 所述的应用访问方法，其特征在于，所述基于全部所述 pod 信息以及预设策略从所述目标应用的全部 pod 中确定出一个主 pod，包括：

基于全部所述 pod 信息以及预设策略从所述目标应用的全部 pod 中确定出创建时间最早的一个 pod，作为主 pod。

9、根据权利要求 7 所述的应用访问方法，其特征在于，所述基于全部所述 pod 信息以及预设策略从所述目标应用的全部 pod 中确定出一个主 pod，包括：

基于全部所述 pod 信息以及预设策略从所述目标应用的全部 pod 中确定出创建时间最晚的一个 pod，作为主 pod。

10、根据权利要求 7 所述的应用访问方法，其特征在于，所述基于全部所述 pod 信息以及预设策略从所述目标应用的全部 pod 中确定出一个主 pod，包括：

根据所述 pod 信息中的创建时间对全部 pod 进行排序，根据排序结果确定主 pod。

11、根据权利要求 1 至 9 任一项所述的应用访问方法，其特征在于，所述解析 k8s 平台基于请求信息发起的 endpoint 创建请求或 endpoint 更新请求之前，还包括：

基于 admission webhook 机制获取 k8s 平台发起的所述 endpoint 创建请求或 endpoint 更新请求。

12、根据权利要求 1 至 9 任一项所述的应用访问方法，其特征在于，所述 endpoint 更新请求基于 k8s 平台监测到目标应用的 pod 发生变化时发起；所述目标应用的 pod 发生变化基于目标应用中任一 pod 出现异常或被删除、或新增 pod 的情况判定。

13、根据权利要求 1 至 9 任一项所述的应用访问方法，其特征在于，所述预设准入控制器部署在容器平台上，所述预设准入控制器基于 k8s 平台的 service、endpoint 资源管理并结合 webhook 机制开发得到。

14、根据权利要求 13 所述的应用访问方法，其特征在于，还包括：

在 k8s 平台创建 endpoint 时，调用预设准入控制器，并将 endpoint 创建请求传入预设准入控制器。

15、根据权利要求 13 所述的应用访问方法，其特征在于，还包括：

在为 service 创建 endpoint 的过程中，调用预设准入控制器，根据当前 endpoint 创建请求中的所有 pod ip，查询关联的 pod 信息；

根据预设策略从所有 pod 中筛选一个主 pod，使用该 pod 的 pod ip 替换请求信息中的 pod ip 列表，以实现通过 service 的请求流量只转发到该 pod，该 pod 将作为主副本响应请求。

16、根据权利要求 14 或 15 所述的应用访问方法，其特征在于，所创建的 endpoint 为所述 k8s 平台在每创建一个 service 后，实时创建同名的 endpoint 资源，所述 endpoint 资源包含 service 所要负载的所有 pod ip 和端口。

17、一种应用访问装置，其特征在于，应用于预设准入控制器，包括：

请求解析模块，用于解析 k8s 平台基于请求信息发起的 endpoint 创建请求或 endpoint 更新请求以得到请求解析信息；

列表获取模块，用于从所述请求解析信息中获取目标应用对应 pod ip 列表；

信息查询模块，用于查询所述 pod ip 列表中每个 pod ip 对应的 pod 信息；

主 pod 确定模块，用于基于全部所述 pod 信息以及预设策略从所述目标应用的全部 pod 中确定出一个主 pod；

信息更新模块，用于将所述请求信息中的所述 pod ip 列表替换为所述主 pod 的 pod ip 以得到替换后请求信息，以便所述 k8s 平台基于所述替换后请求信息创建或更新 endpoint，以将所述目标应用的访问流量均转发至所述主 pod 上。

18、根据权利要求 17 所述的应用访问装置，其特征在于，还包括：

endpoint 名称获取模块，用于从所述请求解析信息中获取 endpoint 名称；

service 信息查询模块，用于基于所述 endpoint 名称查询 service 信息；其中，所述 service 信息为所述 endpoint 对应的 service 的信息；

主从标签判断模块，用于判断所述 service 信息中是否存在预设主从标签；若存在所述预设主从标签且所述预设主从标签的值为第一预设值，则触发所述信息查询模块执行所述查询所述 pod ip 列表中每个 pod ip 对应的 pod 信息的步骤。

19、一种电子设备，其特征在于，包括处理器和存储器；其中，

所述存储器，用于保存计算机程序；

所述处理器，用于执行所述计算机程序以实现如权利要求 1 至 16 任一项所述的应用访问方法。

20、一种计算机非易失性可读存储介质，其特征在于，用于保存计算机程序，其中，所述计算机程序被处理器执行时实现如权利要求 1 至 16 任一项所述的应用访问方法。



图 1

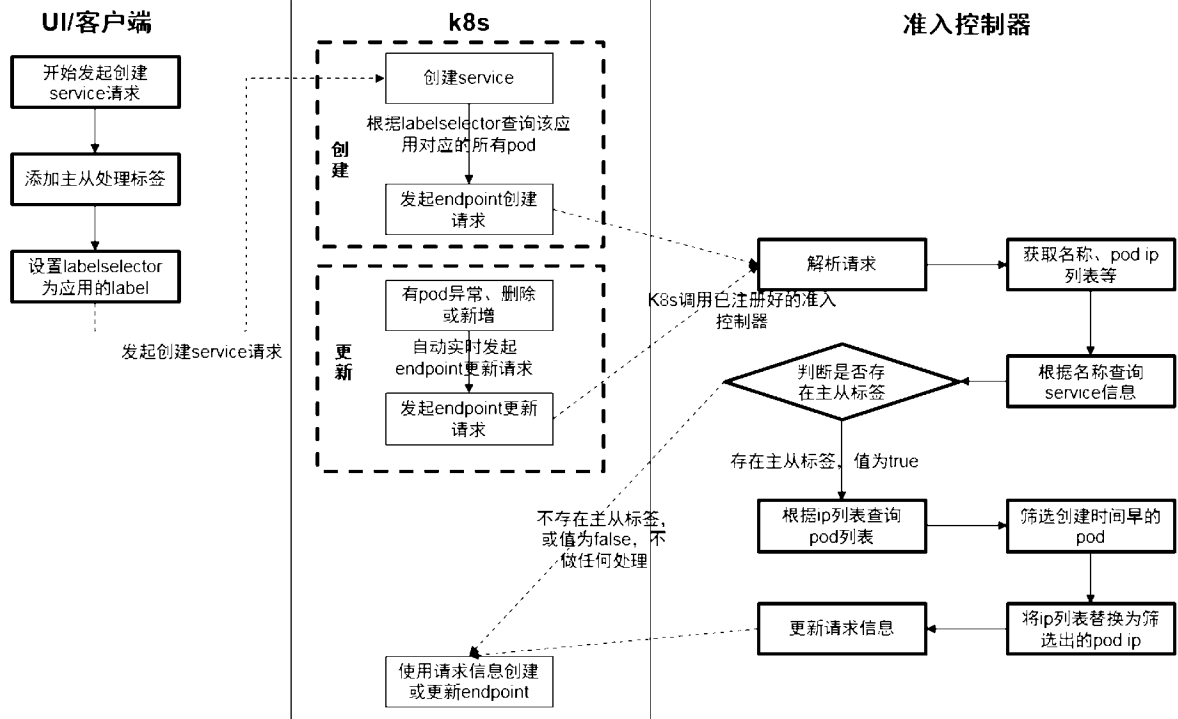


图 2

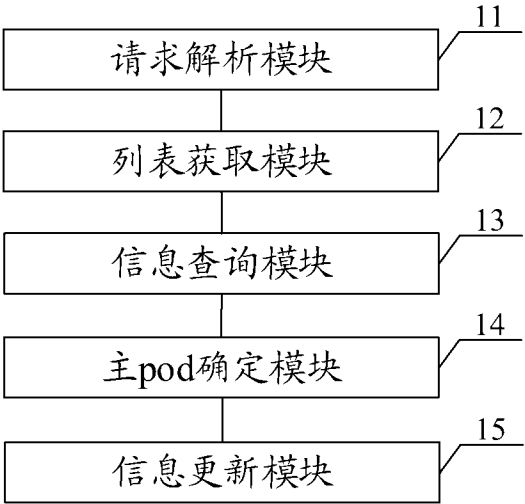


图 3

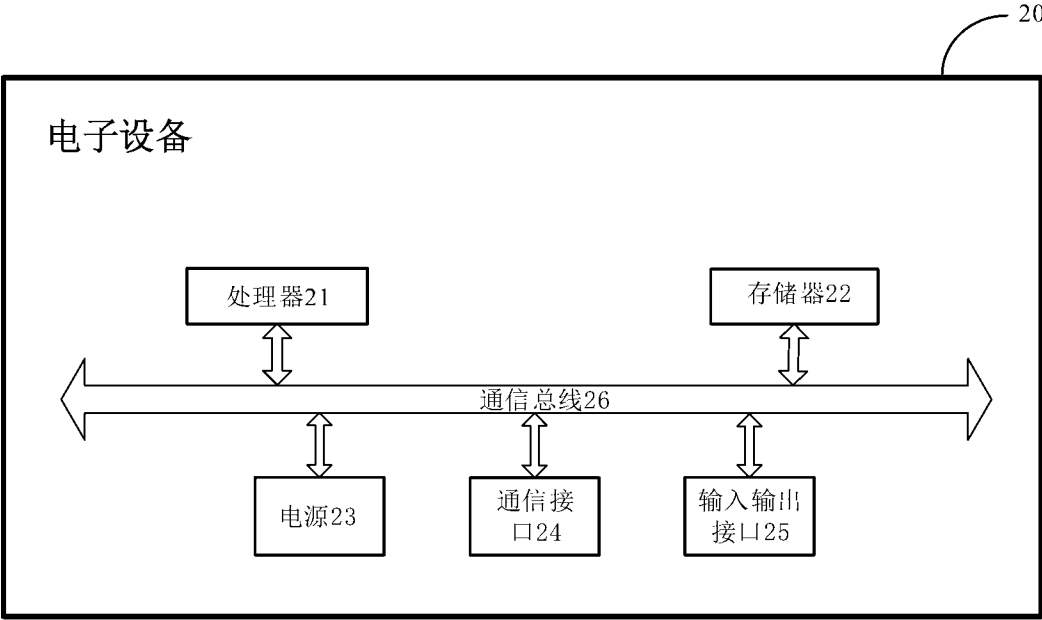


图 4

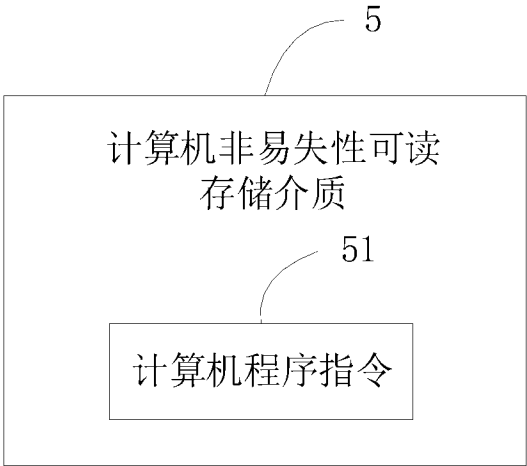


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/084815

A. CLASSIFICATION OF SUBJECT MATTER

H04L 47/10(2022.01)i; G06F 11/20(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: H04L G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS; CNTXT; CNKI; VEN; WOTXT; USTXT; EPTXT; 3GPP; IEEE: 应用, 访问, 容器, 编排, 主备, 主从, 容灾, 选举, 创建, 更新, 请求, k8s, Kubernetes, endpoint, pod, IP, application, APP, service, container, standby, create, update, request

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 114448895 A (SUZHOU INSPUR INTELLIGENT TECHNOLOGY CO., LTD.) 06 May 2022 (2022-05-06) description, paragraphs [0017]-[0052]	1-20
X	CN 112153143 A (SINA TECHNOLOGY (CHINA) CO., LTD.) 29 December 2020 (2020-12-29) description, paragraphs [0067]-[0112]	1-20
A	CN 114124948 A (JINAN INSPUR DATA TECHNOLOGY CO., LTD.) 01 March 2022 (2022-03-01) entire document	1-20
A	US 2021328858 A1 (RIBBON COMMUNICATIONS OPERATING CO., INC.) 21 October 2021 (2021-10-21) entire document	1-20
A	WO 2021063339 A1 (TRANSWARP TECHNOLOGY (SHANGHAI) CO., LTD.) 08 April 2021 (2021-04-08) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

31 May 2023

Date of mailing of the international search report

07 July 2023

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/084815

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	114448895	A	06 May 2022	None	
CN	112153143	A	29 December 2020	None	
CN	114124948	A	01 March 2022	None	
US	2021328858	A1	21 October 2021	None	
WO	2021063339	A1	08 April 2021	None	

A. 主题的分类

H04L 47/10(2022.01)i; G06F 11/20(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC: H04L G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS;CNTXT;CNKI;VEN;WOTXT;USTXT;EPTXT;3GPP;IEEE: 应用, 访问, 容器, 编排, 主备, 主从, 容灾, 选举, 创建, 更新, 请求, k8s, Kubernetes, endpoint, pod, IP, application, APP, service, container, standby, create, update, request

C. 相关文件

类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 114448895 A (苏州浪潮智能科技有限公司) 2022年5月6日 (2022 - 05 - 06) 说明书第[0017]-[0052]段	1-20
X	CN 112153143 A (新浪网技术(中国)有限公司) 2020年12月29日 (2020 - 12 - 29) 说明书第[0067]-[0112]段	1-20
A	CN 114124948 A (济南浪潮数据技术有限公司) 2022年3月1日 (2022 - 03 - 01) 全文	1-20
A	US 2021328858 A1 (RIBBON COMM OPERATING CO INC) 2021年10月21日 (2021 - 10 - 21) 全文	1-20
A	WO 2021063339 A1 (TRANSWARP TECH SHANGHAI CO LTD) 2021年4月8日 (2021 - 04 - 08) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2023年5月31日

国际检索报告邮寄日期

2023年7月7日

ISA/CN的名称和邮寄地址

中国国家知识产权局

中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

燕璐

电话号码 (+86) 0512-88996236

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2023/084815

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	114448895	A	2022年5月6日	无	
CN	112153143	A	2020年12月29日	无	
CN	114124948	A	2022年3月1日	无	
US	2021328858	A1	2021年10月21日	无	
WO	2021063339	A1	2021年4月8日	无	