

# 面向数据处理与管理的云平台系统架构设计

汪 朋,姜红玉,封 雷

(中国电子科技集团公司第十五研究所,北京 100083)

**摘 要:**随着云原生技术在云计算领域的广泛应用,应用云原生技术更好地支撑业务能力成为了开发者们关注的重点。大数据环境下,基于海量数据和信息,针对数据处理与信息管理业务,对云平台系统进行了研究工作,梳理了系统的功能组成、部署形式,从体系架构、功能结构、技术设计等方面设计了一种面向数据处理与信息管理的云平台,给出了框架、方法和技术,实现系统资源的统一管理、监控、共享和调度,实现应用的统一部署和高可用性。从整体角度分析设计了技术体系架构。通过应用容器云技术,该平台能够自动部署集群、快速扩容计算空间,用于应对大型数据处理面临的计算量巨大、优化困难等问题。该系统架构自动化程度更高、可用性更强,应用运行结果更具稳定性,能有效满足信息在存储、大规模计算、深度数据挖掘等方面的需求,从而达到通过云平台技术强力支撑数据处理与信息管理业务以及提升数据处理能力的目的。

**关键词:**云原生;云平台;容器技术;云计算;大数据

**中图分类号:**TP391

**文献标识码:**A

**文章编号:**1673-629X(2022)07-0122-06

**doi:**10.3969/j.issn.1673-629X.2022.07.021

## Design of Cloud Platform System Architecture for Data Processing and Information Management

WANG Peng, JIANG Hong-yu, FENG Lei

(The 15th Research Institute of China Electronics Group Corporation, Beijing 100083, China)

**Abstract:** With the wide application of cloud native technology in the field of cloud computing, the application of cloud native technology to better support business capabilities has become the focus of developers. In the big data environment, based on massive data and information, aiming at data processing and information management business, we study the cloud platform system, combs the functional composition and deployment form of the system, designs a cloud platform for data processing and information management from the aspects of architecture, functional structure and technical design, and gives the framework, methods and technologies, so as to realize the unified management, monitoring, sharing and scheduling of system resources, and realize the unified deployment and high availability of applications. The technical architecture is analyzed and designed from the overall perspective. Through the application of container cloud technology, the platform can automatically deploy clusters and rapidly expand computing space to deal with the huge amount of computing and optimization difficulties faced by large-scale data processing. The system architecture has higher automation, stronger availability and more stable application operation results. It can effectively meet the needs of information storage, large-scale computing and deep data mining, so as to strongly support data processing and information management business and improve data processing capacity through cloud platform technology.

**Key words:** cloud native; cloud platform; container technology; cloud computing; big data

## 0 引 言

信息技术从出现伊始到渐成主流,其发展历程经历了软件、开源、云三个阶段。从软件到开源,再到云,这也是信息技术的发展趋势。容器技术日趋成熟,在应对微服务和系统集约化方面具有巨大优势<sup>[1]</sup>。仅通过开源产品搭建并运维一个高可用、高度弹性化的平

台,进而实现近乎100%的可用性,难度可想而知。因此,在提供技术思路的同时,进一步提供整套云解决方案以保障不断扩展的非功能需求,便成了当今新一代平台的追求<sup>[2]</sup>。

随着虚拟化技术的成熟和分布式架构的普及,用来部署、管理和运行应用的云平台被越来越多地提及。

收稿日期:2021-08-01

修回日期:2021-12-02

基金项目:中国电子科技集团重点科研项目(JY201802850)

作者简介:汪 朋(1987-),男,工程师,硕士,研究方向为GIS应用。

IaaS、PaaS 和 SaaS 是云计算的三种基本服务类型,分别表示关注硬件基础设施的基础设施即服务、关注软件和中间件平台的平台即服务,以及关注业务应用的软件即服务。容器的出现,使原有的基于虚拟机的云主机应用,彻底转变为更加灵活和轻量的“容器+编排调度”的云平台应用<sup>[2]</sup>。

容器技术作为一种比虚拟化更轻量的资源提供技术,受到业界广泛青睐与应用,大大推动和简化了软件的开发、部署和运营。容器服务依托容器技术,通过整合软硬件资源,以容器的形式向开发者/运营者提供应用的编排、部署、管理和监控等云服务。

## 1 云原生

云一般是指一个提供资源的平台,云计算的本质是按需分配资源和弹性计算。云是指我们在哪里计算,而云原生指的是如何实现<sup>[3]</sup>。云原生应用被重点关注的地方,是让应用能够利用云平台实现资源的按需分配和弹性伸缩。云原生还关注规模,分布式系统应该具备将节点扩展到成千上万个的能力,并且这些节点应具有多租户和自愈能力。

从本质上来说,云原生是一种设计模式,它要求云原生应用具备可用性和伸缩性,以及自动化部署和管理的能力,可随处运行,并且能够通过持续集成、持续交付工具提升研发、测试与发布的效率<sup>[4-10]</sup>。

Kubernetes 是一个全新的基于容器技术的分布式架构领先方案<sup>[11]</sup>。Kubernetes 是一个容器编排系统,基于容器技术,实现了资源管理的自动化,擅长无状态

的应用部署管理,在微服务领域起到了重要作用。Kubernetes 对运行在其上的应用会进行实时状态监控,当集群判断出容器状态异常时,可自动对应用进行重启,保证业务稳定运行<sup>[12]</sup>。在云原生设计模式下,Kubernetes 作为重要基础设施,成为了云服务及应用交付的新底座。

## 2 系统概述

面向数据处理与信息管理的云平台系统是以先进的云平台技术为基础,集多源异构数据存储管理和数据处理服务、在线信息云服务等于一体的基础云平台。将云平台技术和数据处理与信息管理业务进行融合,使信息在存储、大规模计算、深度数据挖掘等方面获得云平台技术的强力支撑,提升数据处理能力。

## 3 系统设计方案

面向数据处理与信息管理的云平台系统,以容器为核心,结合业务场景与云原生技术进行整体设计,旨在建立一个能够让高度分布式的应用程序可以在不断变化的环境中运行的云平台。

### 3.1 系统功能结构设计

#### 3.1.1 系统功能组成

面向数据处理与信息管理的云平台系统由基础云平台系统、运维及安全管理系统、分布式信息存储管理系统、数据处理插件支撑系统、信息在线应用支撑系统五个分系统组成,如图1所示。

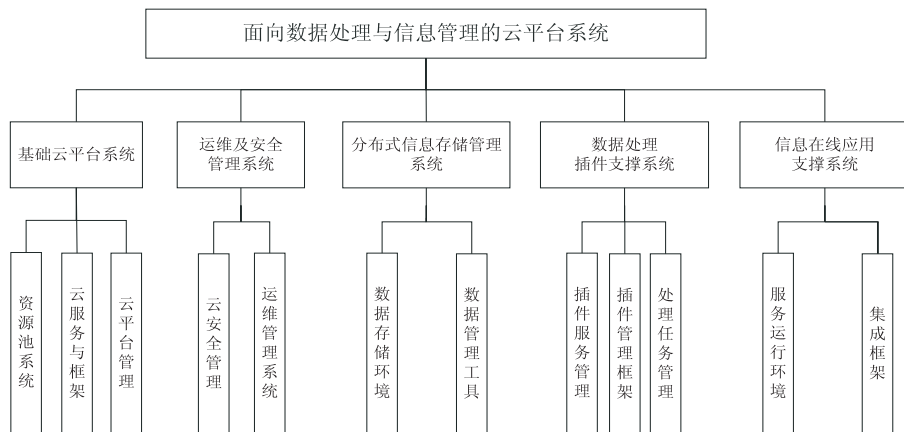


图1 系统组成

#### (1) 云平台系统。

依托云平台为上层业务系统提供资源统一管理,提供资源按需分配、系统托管和业务快速部署的能力<sup>[13-14]</sup>。云平台系统提供节点资源、存储资源的管理功能,主要包括:计算框架/数据库服务、存储服务、消息队列服务。提供应用编排、应用调度管理、应用弹性伸缩管理、应用升级管理、软件包/镜像包生命周期管

理等云应用管理功能。

#### (2) 运维及安全管控系统。

基于云平台提供了覆盖硬件、网络基础设施和应用日志的基础监控,以及基于上述数据和业务模型实现的分布式应用故障及性能管理工具,提供一个从问题发现、问题界定、问题定位、问题处置的故障全生命周期的管理平台。

提供“统一安全服务框架”,为各应用提供一站式的可视化授权管理中心和统一的鉴权服务中心。即统一用户、组织机构信息管理与查询服务;统一身份认证;统一授权管理,并可支持分级授权管理;统一鉴权服务;统一日志访问管理与安全审计。

### (3) 分布式信息存储管理系统。

分布式存储是指文件系统管理的物理存储资源不一定直接连接在本地节点上,而是通过计算机网络与存储节点相连,设计目标是在通用、廉价的硬件平台上构建大容量、高吞吐率和可伸缩性的分布式存储架构<sup>[15-16]</sup>。

基于分布式存储技术,实现分布式数据管理系统;构建多源信息的存储组织模型,实现异构数据的存储与管理;基于分布式并行事务处理框架实现海量数据高效查询访问功能。

### (4) 信息处理插件支撑系统。

基于云计算、虚拟化和并行处理技术,构建数据处理插件管理框架系统,支撑面向各业务接入集成,实现处理任务的并行化执行。

### (5) 信息在线应用支撑系统。

提供面向支撑用户应用的平台系统,在线服务运行环境提供对各类在线服务插件的支持,实现各类能力的网络服务化访问;集成框架提供自动处理和人机交互插件的集成应用软件平台,方便用户各类插件的访问;可视化框架提供面向信息显示和海量大数据可视化的平台,实现快速的数据访问和多种数据可视化视图。

#### 3.1.2 系统部署形式

系统采用集群的模式进行部署,集群包含的机器根据用处不同被分配给不同的系统来使用,系统软件的具体部署部位如图 2 所示。

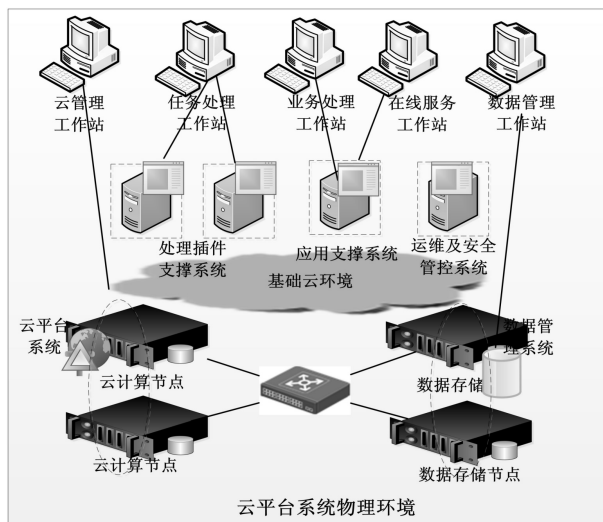


图 2 系统物理组成

将两台服务器分配为云计算节点,用于部署云平

台系统提供云平台的基础运行环境;将另外两台服务器分配为数据存储节点,用于部署数据管理系统,为数据生产、服务提供数据存储访问环境;在上述环境上构建了处理云平台部署处理插件系统,提供各类处理功能的虚拟化运行环境以及服务运行管理框架,实现处理功能的插件化、并行化运行;在其他虚拟化环境上部署应用支持系统及运维安全管控系统,实现系统必须的基础功能。

## 3.2 软件技术设计

### 3.2.1 技术体系架构

面向数据处理与信息管理的云平台系统技术结构按照云+端和平台+插件应用的架构思路,构建基于云计算与大数据的信息处理管理云平台。数据处理云平台负责提供多源、海量信息接入管理、融合处理、存储等云服务;服务应用云平台构建基于平台+插件的一体化综合应用服务框架,面向用户提供信息综合展现平台,提供数据服务、分析服务、应用服务等插件功能。系统体系架构如图 3 所示。

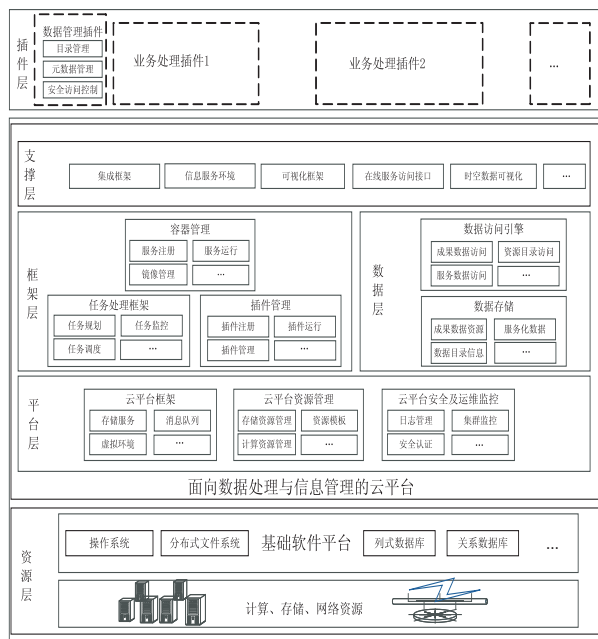


图 3 系统体系架构

### 3.2.2 功能体系

采用“共用平台+数据+应用插件”的集成框架,整合各种业务数据等信息资源,在统一的共用平台上组织管理和共享使用,实现资源统一管理、信息统一服务。在统一平台下将数据服务对外发布,构建网络服务云,实现数据栅格化信息服务能力,按需随遇接入的各种应用终端,支持四类信息服务,包括共用支撑服务、软件服务、业务功能服务和高级信息产品服务,形成“网络服务云”+“应用终端”的服务体系。面向数据处理与信息管理的云平台服务视图如图 4 所示。

共用支撑服务主要包括系统管理、资源管理、状态

监控、租户管理、人员权限与角色管理等用户使用的基 础性服务。

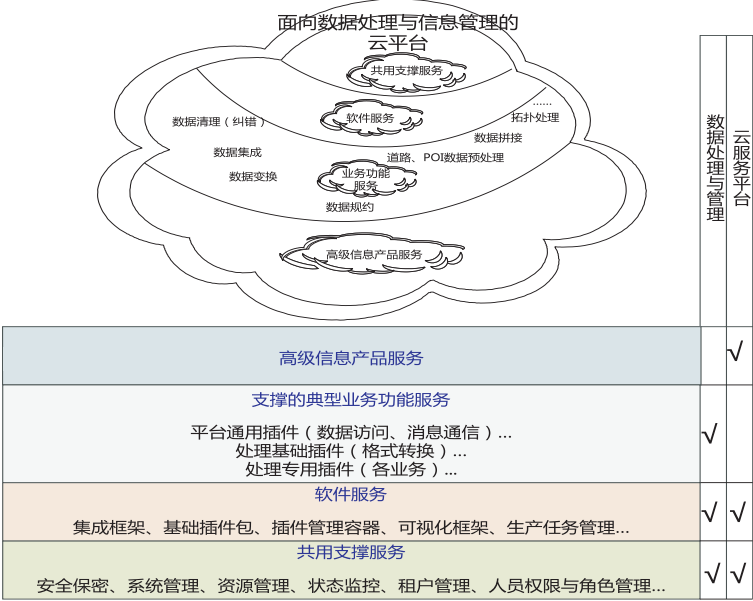


图 4 系统服务视图

软件服务主要包括集成框架、通用插件包、插件管理容器、云端应用平台、集成规范等支撑应用系统构建和运行的服务。

业务功能服务主要由各类处理插件构成,分为平台通用、处理基础和处理专用插件三类。

高级信息产品服务包括各种业务数据等面向用户的多样化处理功能服务。

4 系统关键技术

4.1 在线服务智能管理调度技术

在线服务智能化管理调度技术是解决分布式网络环境下多服务之间协同的问题,为建立高并发支撑能力、高性能、高可靠性在线信息服务系统奠定技术基

础,主要包括两方面:

(1)在线服务智能化调度。

主要解决如何将分布式上的多个在线服务虚拟成一个逻辑上统一的“超级”云服务中心的问题,实现统一的在线信息服务管理平台,实现对各个注册的所有信息服务的虚拟化管理,以便支持服务的动态绑定、发现和替换,以及为用户即时提供更面向用户业务领域的、种类更丰富的服务,达到信息服务的快速访问与全局负载均衡效果。

在线服务智能化调度技术主要包括服务目录、状态监控、访问控制、动态路由、服务发现及智能缓存等技术(见图 5),服务智能化调度的过程是:当客户端发送访问请求时,系统根据服务目录搜索能够满足需要

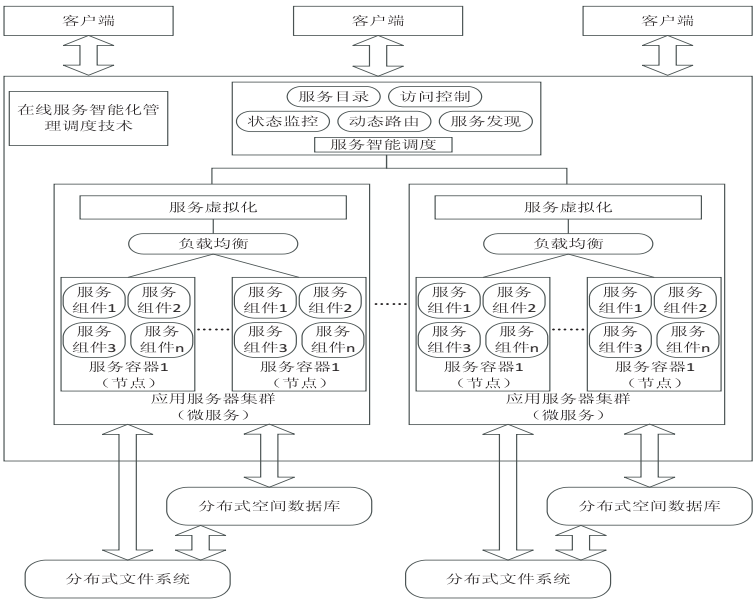


图 5 在线服务智能管理调度技术原理图

的所有在线服务,并监听服务之间的多应用服务器集群的状态,评估集群服务负载后,通过访问控制来指派最优的微服务来执行<sup>[17]</sup>,因服务扩充或服务节点停止工作等各类原因导致服务节点发生变化时,网络能进行发现服务、动态路由,为减少服务与应用终端的网络负载,在应用终端采用智能缓存的方法实现本地化快速服务和离线条件下的应用保障。

#### (2) 信息服务虚拟化。

虚拟化管理是指将多应用服务器集群机制下的各类服务进行聚类、组合、封装、转换,生成新的适合用户要求的虚拟服务,实现在线服务的聚合。

服务虚拟化指在不改变原有服务功能的前提下,对服务进行加工、抽象、转换,生成新的适合用户要求的虚拟服务。根据虚拟化层次的不同,服务虚拟化技术手段包括服务聚类技术、服务组合技术等。其中,服务内聚是将功能相同或者相似的服务抽象成具有虚拟特性的聚类服务;服务组合是将多个细粒度的服务组合成具有一定业务逻辑的粗粒度服务。

### 4.2 多级并行计算架构设计

目前,部分业务信息处理周期长、服务能力单一,并且信息数据种类繁多,采集复杂,生产力量少而使用

者众多,基础数据量庞大而更新频率低、并发访问数据量大、服务计算密度大的特点,基于云平台提供的多级并行计算框架能力,使信息在存储、大规模计算、深度数据挖掘等方面获得 IT 技术的强力支撑。

云计算的分布式特性天然具备了并行计算的基础条件,在设计数据处理功能过程中,需要充分利用云计算提供的能力实现业务处理效能的提升。首先在物理层面,云环境会将处理功能发布到多个实体物理机上运行,保证处理的并发和可靠性;在虚拟化层面,云平台会根据任务配置情况开设合适数量的虚拟环境给每个处理任务,一个处理任务可以对应多个相同的处理插件服务同时进行处理工作,实现处理任务的多进程执行;在处理层面每个处理插件服务内部可以采用多线程模式进行数据内容单元的获取,实现数据的并行访问,而每个数据内容单元存储在分布式环境中,可以保证数据获取的多 I/O 访问。海量数据分布式存储融合接入各种海量业务数据类型,支持结构化和非结构化大数据的统一存储管理,支持弹性扩容、大容量、高并发和高速一体化的数据访问<sup>[18]</sup>。原理图如图 6 所示。

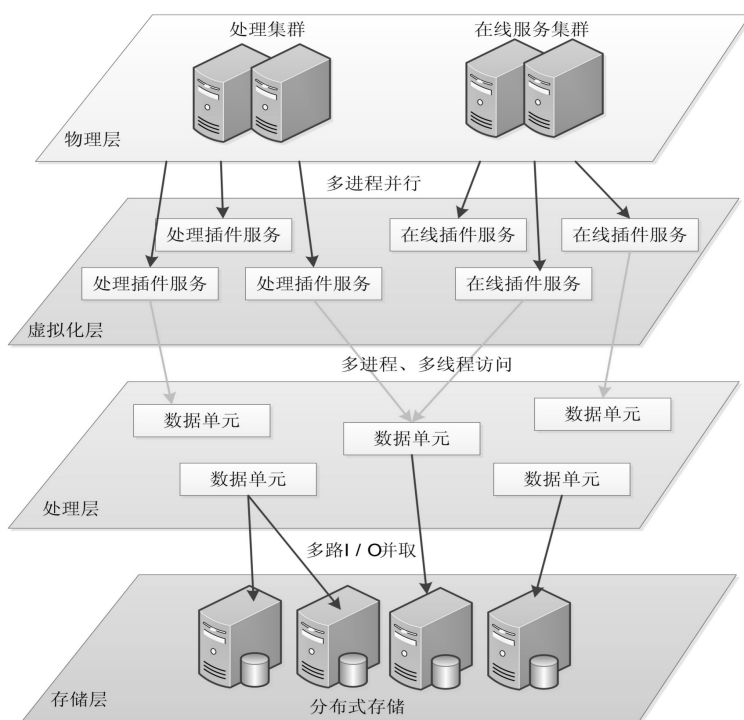


图6 多级并行计算技术原理图

## 5 实用效果分析

该文结合某云平台项目及测评报告,从基础云平台、集群计算能力验证、并发任务调度、存储资源调度、租户生命周期管理、租户资源隔离、数据加密传输、数据安全、数据存储管理能力、数据吞吐效率、数据检索

性能测试等 18 项指标对该云平台系统进行了分析总结,主要结果表明:

- (1) 单机群计算节点部署能力节点数量  $\geq 1\,000$  个;
- (2) 自动化处理插件的并发调度处理流程个数  $\geq 100$  个;

(3)并发任务调度测试响应时间 $\leq 1$ 秒;

(4)针对真实数据,万兆网络环境下,单节点不低于200 MB/s处理吞吐能力,简单业务处理延时毫秒级;

(5)通过性测试项测试均为通过;

(6)有些数据处理插件运算耗时较多,需要进一步优化处理工具算法效率。

## 6 结束语

大多数开发者都是基于多个现有解决方案进行软件开发的。即使是在开发一个新的软件,可能也需要与那些现有的系统进行对接,而且已经运行的软件几乎都不是云原生的。云原生的绝妙之处在于它最终是由许多不同的组件组成的,即使其中一些组件的模式不是最新的,云原生的组件仍然可以与它们进行交互。即使你的软件使用了旧的设计模式,应用云原生模式依然可以带来立竿见影的价值。

该文针对数据处理与信息管理云化改造需求,从体系架构、关键技术等方面设计了一种面向数据处理与信息管理的云平台,给出了框架、方法和技术。通过应用容器云技术,该平台能够自动部署集群、快速扩容计算空间,用于应对大型数据处理与信息项目面临的计算量巨大、优化困难等问题。通过容器技术和立体化监控,可全方位监控各集群的资源整理利用率,方便了集群的管理和维护。并通过多集群间动态部署容器数量,实现灵活的水平扩展能力。研究证明,容器云技术有力保证了数据处理与信息管理平台的高自动化与高可用性,具有良好的应用前景,为云平台系统的架构设计改造给出了参考。但考虑容器技术作为一个较新的技术,在推动容器技术的标准化和规范化、实现更多应用的重构及统管、构建统一化的服务组件等方面,仍需要持续关注和研究。

## 参考文献:

- [1] 关兴民,张兆伟,王祝先,等. 基于 Kubernetes 的气象私有云自动化部署与优化[J]. 信息技术,2019(5):76-80.
- [2] 张 亮,吴 晟,敖小剑,等. 未来架构从服务化到云原生[M]. 北京:电子工业出版社,2019:1-40.
- [3] 张若飞,宋净超. 云原生模式设计拥抱变化的软件[M]. 北京:电子工业出版社,2020:20-100.
- [4] MARKO L. Kubernetes in action[M]. Beijing:Publishing House of Electronics Industry,2018.
- [5] GUO Yanghui. Research and implementation of docker container scheduling strategy in micro service[D]. Beijing:Beijing University of Posts and Telecommunications,2018.
- [6] 曾德泽,陈律昊,顾 琳,等. 云原生边缘计算:探索与展望[J]. 物联网学报,2021,5(2):7-17.
- [7] LI W B,LEMIEUX Y,GAO J,et al. Service mesh:challenges,state of the art,and future research opportunities[C]//2019 IEEE international conference on service-oriented system engineering (SOSE). San Francisco East Bay:IEEE,2019:122-127.
- [8] BURNS B,GRANT B,OPPENHEIMER D,et al. Borg,omega and kubernetes[J]. Queue,2016,59(5):50-57.
- [9] HIGHTOWER K,BURNS B,BEDA J. Kubernetes:up and running dive into the future of infrastructure[M]. Massachusetts:O'Reilly Media,Inc,2017.
- [10] 单朋荣,杨美红,赵志刚,等. 基于 Kubernetes 云平台的弹性伸缩方案设计与实现[J]. 计算机工程,2021,47(1):312-320.
- [11] 龚 正,吴治辉,崔秀龙,等. Kubernetes 权威指南第4版[M]. 北京:电子工业出版社,2019:1-30.
- [12] 夏 畅. 基于 Kubernetes 的企业级容器云平台建设[J]. 电信快报,2021(1):12-16.
- [13] 何锡点,马桂勤. 基于云平台的数据中心改造架构设计及关键技术[J]. 网络安全技术与应用,2018(12):73-75.
- [14] 齐 磊,张海峰,张天晓,等. 基于容器技术的 PaaS 云平台方案[J]. 电信科学,2017,33(4):177-182.
- [15] 许春聪,黄小猛,吴 诺,等. 分布式文件系统存储介质评测与分析[J]. 计算机学报,2010,33(10):1873-1880.
- [16] 刘福鑫,李劲巍,王熠弘,等. 基于 Kubernetes 的云原生海量数据存储系统设计与实现[J]. 计算机应用,2020,40(2):547-552.
- [17] 张若飞,宋净超. 云原生 Java Spring Boot、Spring Cloud 与 Cloud Foundry 弹性系统设计[M]. 北京:电子工业出版社,2018:1-30.
- [18] 马晓亭,陈 臣. 基于虚拟化技术的数字图书馆分布式云计算数据中心研究[J]. 情报科学,2012,30(7):1025-1029.