

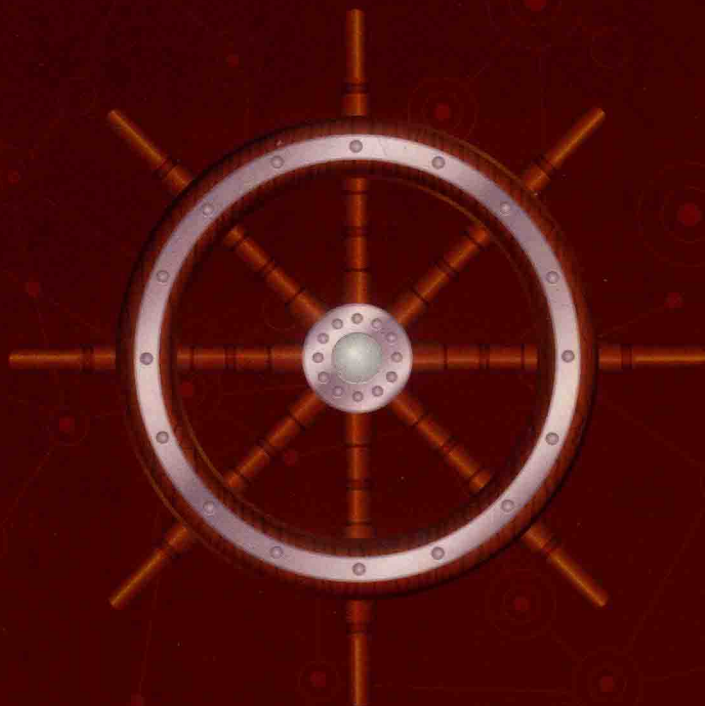
Kubernetes

权威指南



从Docker到Kubernetes
实践全接触

龚正 吴治辉
崔秀龙 闫健勇 编著



下一代分布式架构的王者

Kubernetes: The Definitive Guide



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

 博文视点云原生精品丛书

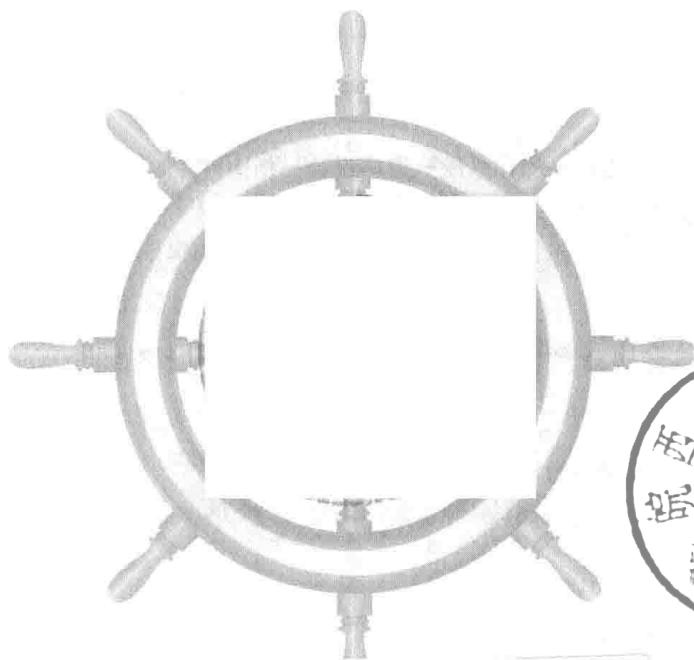
Kubernetes

权威指南



从Docker到Kubernetes
实践全接触

龚正 吴治辉 / 编著
崔秀龙 闫健勇



电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

Kubernetes 是由谷歌开源的 Docker 容器集群管理系统，为容器化的应用提供了资源调度、部署运行、服务发现、扩容及缩容等一整套功能。本书从架构师、开发人员和运维人员的角度，阐述了 Kubernetes 的基本概念、实践指南、核心原理、开发指导、运维指南、新特性演进等内容，图文并茂、内容丰富、由浅入深、讲解全面；并围绕在生产环境中可能出现的问题，给出了大量的典型案例，比如安全配置方案、网络方案、共享存储方案、高可用方案及 Trouble Shooting 技巧等，有很强的实战指导意义。本书内容随着 Kubernetes 的版本更新不断完善，目前涵盖了 Kubernetes 从 1.0 到 1.14 版本的主要特性，努力为 Kubernetes 用户提供全方位的 Kubernetes 技术指南。本书源码已上传至 GitHub 的 kubeguide/K8sDefinitiveGuide-V4-Sourcecode 目录，可自行下载本书源码进行练习。

无论是对于软件工程师、测试工程师、运维工程师、软件架构师、技术经理，还是对于资深 IT 人士，本书都极具参考价值。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

Kubernetes 权威指南：从 Docker 到 Kubernetes 实践全接触 / 龚正等编著. —4 版. —北京：电子工业出版社，2019.6

（博文视点云原生精品丛书）

ISBN 978-7-121-36235-4

I. ①K… II. ①龚… III. ①Linux 操作系统—程序设计—指南 IV. ①TP316.85-62

中国版本图书馆 CIP 数据核字（2019）第 060814 号

责任编辑：张国霞

印 刷：三河市良远印务有限公司

装 订：三河市良远印务有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本：787×980 1/16 印张：51.5 字数：1072 千字

版 次：2016 年 1 月第 1 版

2019 年 6 月第 4 版

印 次：2019 年 6 月第 1 次印刷

印 数：5000 册 定价：168.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及购电话：（010）88254888，88258888。

质量投诉请发邮件至 zltz@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819，faq@phei.com.cn。



推荐序

经过作者们多年的实践经验积累及近一年的精心准备，本书终于与我们见面了。我有幸作为首批读者，提前见证和学习了在云时代引领业界技术方向的 Kubernetes 和 Docker 的最新动态。

从内容上讲，本书从一个开发者的角度去理解、分析和解决问题：从基础入门到架构原理，从运行机制到开发源码，再从系统运维到应用实践，讲解全面。本书图文并茂，内容丰富，由浅入深，对基本原理阐述清晰，对系统架构分析透彻，对实践经验讲解深刻。

我认为本书值得推荐的原因有以下几点。

首先，作者的所有观点和经验，均是在多年建设、维护大型应用系统的过程中积累形成的。例如，读者通过学习书中的 Kubernetes 开发指南、集群管理等章节的内容，不仅可以直接提高开发技能，还可以解决在实践过程中经常遇到的各种关键问题。书中的这些内容具有很高的借鉴和推广意义。

其次，通过大量的实例操作和详尽的源码解析，本书可以帮助读者深刻理解 Kubernetes 的各种概念。例如，书中介绍了使用 Java 程序访问 Kubernetes API 的几种方法，读者参照其中的案例，只要稍做修改，再结合实际的应用需求，就可以将这些方法用于正在开发的项目中，达到事半功倍的效果，对有一定 Java 基础的专业人士快速学习 Kubernetes 的各种细节和实践操作十分有利。

再次，为了让初学者快速入门，本书配备了即时在线交流工具和专业后台技术支持团队。如果你在开发和应用的过程中遇到各类相关问题，均可直接联系该团队的开发支持专家。

最后，我们可以看到，容器化技术已经成为计算模型演化的一个开端，Kubernetes 作为谷歌开源的 Docker 容器集群管理技术，在这场新的技术革命中扮演着重要的角色。

Kubernetes 正在被众多知名公司及企业采用，例如 Google、VMware、CoreOS、腾讯、京东等，因此，Kubernetes 站在了容器新技术变革的浪潮之巅，将具有不可预估的发展前景和商业价值。

无论你是架构师、开发者、运维人员，还是对容器技术比较好奇的读者，本书都是一本不可多得的带你从入门到进阶的 Kubernetes 精品书，值得阅读！

初瑞

中国移动业务支撑中心高级经理



自序

本书第 1 版出版于 2016 年，几年过去，Kubernetes 已从一个新生事物发展为一个影响全球 IT 技术的基础设施平台，也推动了云原生应用、微服务架构、Service Mesh 等热门技术的普及和落地。现在，Kubernetes 已经成为明星项目，其开源项目拥有超过两万名贡献者，成为开源历史上发展速度超快的项目之一。

在这几年里：

Kubernetes 背后的重要开源公司 RedHat 被 IBM 大手笔收购，使 RedHat 基于 Kubernetes 架构的先进 PaaS 平台——OpenShift 成为 IBM 在云计算基础设施中的重要筹码；

Kubernetes 的两位核心创始人 Joe Beda 和 Craig McLuckie 所创立的提供 Kubernetes 咨询和技术支持的初创公司 Heptio 也被虚拟化领域的巨头 VMware 收购；

Oracle 收购了丹麦的一家初创公司 Wercker，然后开发了 Click2Kube，这是面向 Oracle 裸机云（Oracle Bare Metal Cloud）的一键式 Kubernetes 集群安装工具；

世界 500 强中的一些大型企业也决定以 Kubernetes 为基础重构内部 IT 平台架构，大数据系统的一些用户也在努力将其生产系统从庞大的大数据专有技术栈中剥离出来靠拢 Kubernetes。

Kubernetes 是将“一切以服务（Service）为中心，一切围绕服务运转”作为指导思想创新型产品，这是它的一个亮点。它的功能和架构设计自始至终地遵循了这一指导思想，构建在 Kubernetes 上的系统不仅可以独立运行在物理机、虚拟机集群或者企业私有云上，也可以被托管在公有云上。

Kubernetes 的另一个亮点是自动化。在 Kubernetes 的解决方案中，一个服务可以自我扩展、自我诊断，并且容易升级，在收到服务扩容的请求后，Kubernetes 会触发调度流程，

最终在选定的目标节点上启动相应数量的服务实例副本，这些服务实例副本在启动成功后会自动加入负载均衡器中并生效，整个过程无须额外的人工操作。另外，Kubernetes 会定时巡查每个服务的所有实例的可用性，确保服务实例的数量始终保持为预期的数量，当它发现某个实例不可用时，会自动重启该实例或者在其他节点上重新调度、运行一个新实例，这样，一个复杂的过程无须人工干预即可全部自动完成。试想一下，如果一个包括几十个节点且运行着几万个容器的复杂系统，其负载均衡、故障检测和故障修复等都需要人工介入进行处理，其工作量将多大。

通常，我们会把 Kubernetes 看作 Docker 的上层架构，就好像 Java 与 J2EE 的关系一样：J2EE 是以 Java 为基础的企业级软件架构，Kubernetes 则以 Docker 为基础打造了一个云计算时代的全新分布式系统架构。但 Kubernetes 与 Docker 之间还存在着更为复杂的关系，从表面上看，似乎 Kubernetes 离不开 Docker，但实际上在 Kubernetes 的架构里，Docker 只是其目前支持的两种底层容器技术之一，另一种容器技术则是 Rocket，Rocket 为 CoreOS 推出的竞争产品。

Kubernetes 之所以同时支持 Docker 和 Rocket 这两种互相竞争的容器技术，是有深刻的历史原因的。快速发展的 Docker 打败了谷歌名噪一时的开源容器技术 lmcftfy，并迅速风靡世界。但是，作为一个已经对全球 IT 公司产生重要影响的技术，Docker 容器标准的制定不可能被任何一个公司主导。于是，CoreOS 推出了与 Docker 抗衡的开源容器项目 Rocket，动员一些知名 IT 公司一起主导容器技术的标准化，并与谷歌共同发起基于 CoreOS+ Rocket+Kubernetes 的新项目 Tectonic，使容器技术分裂态势加剧。最后，Linux 基金会于 2015 年 6 月宣布成立开放容器技术项目（Open Container Project），谷歌、CoreOS 及 Docker 都加入了该项目。OCP 项目成立后，Docker 公司放弃了自己的独家控制权，Docker 容器格式也被 OCP 采纳为新标准的基础，Docker 负责起草 OCP 草案规范的初稿文档，并提交自己的容器执行引擎的源码作为 OCP 项目的启动资源。

2015 年 7 月，谷歌正式宣布加入 OpenStack 阵营，其目标是确保 Linux 容器及其关联的容器管理技术 Kubernetes 能够被 OpenStack 生态圈所接纳，这也意味着对数据中心控制平面的争夺已经结束，以容器为代表的形态与以虚拟化为代表的系统形态将会完美融合于 OpenStack 之上，并与软件定义网络和软件定义存储一起主导下一代数据中心。

谷歌凭借着几十年大规模容器使用的丰富经验，步步为营，先是祭出 Kubernetes 这个神器，然后掌控了容器技术的制定标准，最后入驻 OpenStack 阵营全力支持 Kubernetes 的发展。可以预测，Kubernetes 的影响力可能超过十年，所以，我们每个 IT 人都有理由重视这门新技术。

谁能比别人领先一步掌握新技术，谁就能在竞争中赢得先机。慧与中国通信和媒体解决方案领域的资深专家团一起分工协作、并行研究，并废寝忘食地合力撰写，才促成了这部巨著的出版。经过这些年的高速发展，Kubernetes 先后发布了十几个大版本，每个版本都带来了大量的新特性，能够处理的应用场景也越来越丰富。本书遵循从入门到精通的学习路线，涵盖了入门、安装指南、实践指南、核心原理、开发指南、运维指南、新特性演进等内容，内容翔实、图文并茂，几乎囊括了 Kubernetes 当前主流版本的方方面面，无论是对于软件工程师、测试工程师、运维工程师、软件架构师、技术经理，还是对于资深 IT 人士，本书都极具参考价值。

吴治辉

HPE 资深架构师

读者服务

轻松注册成为博文视点社区用户 (www.broadview.com.cn)，扫码直达本书页面。

- **提交勘误**：您对书中内容的修改意见可在 [提交勘误](#) 处提交，若被采纳，将获赠博文视点社区积分（在您购买电子书时，积分可用来抵扣相应金额）。
- **交流互动**：在页面下方 [读者评论](#) 处留下您的疑问或观点，与我们和其他读者一同学习交流。

页面入口：<http://www.broadview.com.cn/36235>





目 录

第 1 章	Kubernetes 入门	1
1.1	Kubernetes 是什么	2
1.2	为什么要用 Kubernetes	5
1.3	从一个简单的例子开始	6
1.3.1	环境准备	7
1.3.2	启动 MySQL 服务	7
1.3.3	启动 Tomcat 应用	10
1.3.4	通过浏览器访问网页	12
1.4	Kubernetes 的基本概念和术语	13
1.4.1	Master	16
1.4.2	Node	16
1.4.3	Pod	19
1.4.4	Label	24
1.4.5	Replication Controller	28
1.4.6	Deployment	31
1.4.7	Horizontal Pod Autoscaler	34
1.4.8	StatefulSet	36
1.4.9	Service	37
1.4.10	Job	45
1.4.11	Volume	45
1.4.12	Persistent Volume	49
1.4.13	Namespace	51
1.4.14	Annotation	52

1.4.15	ConfigMap.....	53
1.4.16	小结.....	54
第 2 章	Kubernetes 安装配置指南.....	55
2.1	系统要求.....	56
2.2	使用 kubeadm 工具快速安装 Kubernetes 集群.....	57
2.2.1	安装 kubeadm 和相关工具.....	57
2.2.2	kubeadm config.....	58
2.2.3	下载 Kubernetes 的相关镜像.....	59
2.2.4	运行 kubeadm init 命令安装 Master.....	59
2.2.5	安装 Node, 加入集群.....	61
2.2.6	安装网络插件.....	62
2.2.7	验证 Kubernetes 集群是否安装完成.....	63
2.3	以二进制文件方式安装 Kubernetes 集群.....	64
2.3.1	Master 上的 etcd、kube-apiserver、kube-controller-manager、 kube-scheduler 服务.....	66
2.3.2	Node 上的 kubelet、kube-proxy 服务.....	71
2.4	Kubernetes 集群的安全设置.....	73
2.4.1	基于 CA 签名的双向数字证书认证方式.....	73
2.4.2	基于 HTTP Base 或 Token 的简单认证方式.....	78
2.5	Kubernetes 集群的网络配置.....	80
2.6	内网中的 Kubernetes 相关配置.....	80
2.6.1	Docker Private Registry (私有 Docker 镜像库).....	80
2.6.2	kubelet 配置.....	81
2.7	Kubernetes 的版本升级.....	81
2.7.1	二进制升级.....	81
2.7.2	使用 kubeadm 进行集群升级.....	82
2.8	Kubernetes 核心服务配置详解.....	84
2.8.1	公共配置参数.....	84
2.8.2	kube-apiserver 启动参数.....	85
2.8.3	kube-controller-manager 启动参数.....	97
2.8.4	kube-scheduler 启动参数.....	107

2.8.5	kubelet 启动参数.....	113
2.8.6	kube-proxy 启动参数.....	128
2.9	CRI（容器运行时接口）详解.....	132
2.9.1	CRI 概述.....	132
2.9.2	CRI 的主要组件.....	133
2.9.3	Pod 和容器的生命周期管理.....	133
2.9.4	面向容器级别的设计思路.....	135
2.9.5	尝试使用新的 Docker-CRI 来创建容器.....	136
2.9.6	CRI 的进展.....	137
2.10	kubectl 命令行工具用法详解.....	137
2.10.1	kubectl 用法概述.....	137
2.10.2	kubectl 子命令详解.....	139
2.10.3	kubectl 参数列表.....	142
2.10.4	kubectl 输出格式.....	143
2.10.5	kubectl 操作示例.....	145
第 3 章	深入掌握 Pod	149
3.1	Pod 定义详解.....	150
3.2	Pod 的基本用法.....	156
3.3	静态 Pod.....	161
3.4	Pod 容器共享 Volume.....	162
3.5	Pod 的配置管理.....	165
3.5.1	ConfigMap 概述.....	165
3.5.2	创建 ConfigMap 资源对象.....	165
3.5.3	在 Pod 中使用 ConfigMap.....	173
3.5.4	使用 ConfigMap 的限制条件.....	179
3.6	在容器内获取 Pod 信息（Downward API）.....	180
3.6.1	环境变量方式：将 Pod 信息注入为环境变量.....	180
3.6.2	环境变量方式：将容器资源信息注入为环境变量.....	182
3.6.3	Volume 挂载方式.....	184
3.7	Pod 生命周期和重启策略.....	186
3.8	Pod 健康检查和服务可用性检查.....	187

3.9	玩转 Pod 调度	190
3.9.1	Deployment 或 RC: 全自动调度	193
3.9.2	NodeSelector: 定向调度	194
3.9.3	NodeAffinity: Node 亲和性调度	197
3.9.4	PodAffinity: Pod 亲和与互斥调度策略	198
3.9.5	Taints 和 Tolerations (污点和容忍)	202
3.9.6	Pod Priority Preemption: Pod 优先级调度	206
3.9.7	DaemonSet: 在每个 Node 上都调度一个 Pod	209
3.9.8	Job: 批处理调度	211
3.9.9	Cronjob: 定时任务	215
3.9.10	自定义调度器	219
3.10	Init Container (初始化容器)	220
3.11	Pod 的升级和回滚	224
3.11.1	Deployment 的升级	225
3.11.2	Deployment 的回滚	231
3.11.3	暂停和恢复 Deployment 的部署操作, 以完成复杂的修改	234
3.11.4	使用 kubectl rolling-update 命令完成 RC 的滚动升级	236
3.11.5	其他管理对象的更新策略	239
3.12	Pod 的扩缩容	240
3.12.1	手动扩缩容机制	240
3.12.2	自动扩缩容机制	241
3.13	使用 StatefulSet 搭建 MongoDB 集群	264
3.13.1	前提条件	264
3.13.2	创建 StatefulSet	265
3.13.3	查看 MongoDB 集群的状态	269
3.13.4	StatefulSet 的常见应用场景	271
第 4 章	深入掌握 Service	276
4.1	Service 定义详解	277
4.2	Service 的基本用法	279
4.2.1	多端口 Service	282
4.2.2	外部服务 Service	283

4.3	Headless Service	284
4.3.1	自定义 SeedProvider	285
4.3.2	通过 Service 动态查找 Pod	286
4.3.3	Cassandra 集群中新节点的自动添加	289
4.4	从集群外部访问 Pod 或 Service	291
4.4.1	将容器应用的端口号映射到物理机	291
4.4.2	将 Service 的端口号映射到物理机	292
4.5	DNS 服务搭建和配置指南	294
4.5.1	在创建 DNS 服务之前修改每个 Node 上 kubelet 的启动参数	296
4.5.2	创建 CoreDNS 应用	297
4.5.3	服务名的 DNS 解析	301
4.5.4	CoreDNS 的配置说明	302
4.5.5	Pod 级别的 DNS 配置说明	304
4.6	Ingress: HTTP 7 层路由机制	306
4.6.1	创建 Ingress Controller 和默认的 backend 服务	307
4.6.2	定义 Ingress 策略	311
4.6.3	客户端访问 http://mywebsite.com/demo	313
4.6.4	Ingress 的策略配置技巧	316
4.6.5	Ingress 的 TLS 安全设置	320
第 5 章	核心组件运行机制	326
5.1	Kubernetes API Server 原理解析	327
5.1.1	Kubernetes API Server 概述	327
5.1.2	API Server 架构解析	330
5.1.3	独特的 Kubernetes Proxy API 接口	334
5.1.4	集群功能模块之间的通信	336
5.2	Controller Manager 原理解析	337
5.2.1	Replication Controller	338
5.2.2	Node Controller	339
5.2.3	ResourceQuota Controller	341
5.2.4	Namespace Controller	343
5.2.5	Service Controller 与 Endpoints Controller	343

5.3	Scheduler 原理解析	344
5.4	kubelet 运行机制解析	348
5.4.1	节点管理	349
5.4.2	Pod 管理	349
5.4.3	容器健康检查	351
5.4.4	cAdvisor 资源监控	352
5.5	kube-proxy 运行机制解析	354
第 6 章	深入分析集群安全机制	358
6.1	API Server 认证管理	359
6.2	API Server 授权管理	361
6.2.1	ABAC 授权模式详解	362
6.2.2	Webhook 授权模式详解	365
6.2.3	RBAC 授权模式详解	368
6.3	Admission Control	384
6.4	Service Account	388
6.5	Secret 私密凭据	393
6.6	Pod 的安全策略配置	396
6.6.1	PodSecurityPolicy 的工作机制	397
6.6.2	PodSecurityPolicy 配置详解	399
6.6.3	Pod 的安全设置详解	406
第 7 章	网络原理	410
7.1	Kubernetes 网络模型	411
7.2	Docker 网络基础	413
7.2.1	网络命名空间	413
7.2.2	Veth 设备对	416
7.2.3	网桥	419
7.2.4	iptables 和 Netfilter	421
7.2.5	路由	424
7.3	Docker 的网络实现	426
7.4	Kubernetes 的网络实现	435
7.4.1	容器到容器的通信	435

7.4.2	Pod 之间的通信	436
7.5	Pod 和 Service 网络实战	439
7.6	CNI 网络模型	454
7.6.1	CNM 模型	454
7.6.2	CNI 模型	455
7.6.3	在 Kubernetes 中使用网络插件	467
7.7	Kubernetes 网络策略	467
7.7.1	网络策略配置说明	468
7.7.2	在 Namespace 级别设置默认的网络策略	470
7.7.3	NetworkPolicy 的发展	472
7.8	开源的网络组件	472
7.8.1	Flannel	472
7.8.2	Open vSwitch	477
7.8.3	直接路由	483
7.8.4	Calico 容器网络和网络策略实战	486
第 8 章	共享存储原理	508
8.1	共享存储机制概述	509
8.2	PV 详解	510
8.2.1	PV 的关键配置参数	511
8.2.2	PV 生命周期的各个阶段	515
8.3	PVC 详解	516
8.4	PV 和 PVC 的生命周期	518
8.4.1	资源供应	518
8.4.2	资源绑定	519
8.4.3	资源使用	519
8.4.4	资源释放	519
8.4.5	资源回收	519
8.5	StorageClass 详解	521
8.5.1	StorageClass 的关键配置参数	521
8.5.2	设置默认的 StorageClass	524

8.6	动态存储管理实战：GlusterFS	524
8.6.1	准备工作	525
8.6.2	创建 GlusterFS 管理服务容器集群	525
8.6.3	创建 Heketi 服务	528
8.6.4	为 Heketi 设置 GlusterFS 集群	530
8.6.5	定义 StorageClass	533
8.6.6	定义 PVC	534
8.6.7	Pod 使用 PVC 的存储资源	536
8.7	CSI 存储机制详解	537
8.7.1	CSI 的设计背景	538
8.7.2	CSI 存储插件的关键组件和部署架构	539
8.7.3	CSI 存储插件的使用示例	540
8.7.4	CSI 的发展	556
第 9 章	Kubernetes 开发指南	560
9.1	REST 简述	561
9.2	Kubernetes API 详解	563
9.2.1	Kubernetes API 概述	563
9.2.2	Kubernetes API 版本的演进策略	570
9.2.3	API Groups (API 组)	571
9.2.4	API REST 的方法说明	573
9.2.5	API Server 响应说明	575
9.3	使用 Java 程序访问 Kubernetes API	577
9.3.1	Jersey	577
9.3.2	Fabric8	590
9.3.3	使用说明	591
9.3.4	其他客户端库	615
9.4	Kubernetes API 的扩展	616
9.4.1	使用 CRD 扩展 API 资源	617
9.4.2	使用 API 聚合机制扩展 API 资源	626

第 10 章	Kubernetes 集群管理	635
10.1	Node 的管理	636
10.1.1	Node 的隔离与恢复	636
10.1.2	Node 的扩容	637
10.2	更新资源对象的 Label	638
10.3	Namespace: 集群环境共享与隔离	639
10.3.1	创建 Namespace	639
10.3.2	定义 Context（运行环境）	640
10.3.3	设置工作组在特定 Context 环境下工作	641
10.4	Kubernetes 资源管理	643
10.4.1	计算资源管理	645
10.4.2	资源配置范围管理（LimitRange）	655
10.4.3	资源服务质量管理（Resource QoS）	662
10.4.4	资源配额管理（Resource Quotas）	670
10.4.5	ResourceQuota 和 LimitRange 实践	676
10.4.6	资源管理总结	685
10.5	资源紧缺时的 Pod 驱逐机制	686
10.5.1	驱逐策略	686
10.5.2	驱逐信号	686
10.5.3	驱逐阈值	688
10.5.4	驱逐监控频率	689
10.5.5	节点的状况	689
10.5.6	节点状况的抖动	690
10.5.7	回收 Node 级别的资源	690
10.5.8	驱逐用户的 Pod	691
10.5.9	资源最少回收量	692
10.5.10	节点资源紧缺情况下的系统行为	692
10.5.11	可调度的资源和驱逐策略实践	694
10.5.12	现阶段的问题	694
10.6	Pod Disruption Budget（主动驱逐保护）	695
10.7	Kubernetes 集群的高可用部署方案	697
10.7.1	手工方式的高可用部署方案	698