

Native Docker Clustering with Swarm

Swarm容器编排与 Docker原生集群

用Swarm部署、配置和运行Docker容器集群

[俄] Fabrizio Soppelsa 著
[泰] Chanwit Kaewkasi
崔婧雯 钟最龙 译

Native Docker Clustering with Swarm

Swarm容器编排与 Docker原生集群

[俄] Fabrizio Soppelsa 著
[泰] Chanwit Kaewkasi
崔婧雯 钟最龙 译

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

Docker Swarm 作为 Docker 集群原生的容器编排解决方案,是 Docker 生态系统中的关键组件之一。本书涵盖了 Swarm 中的发现、调度、高可用、安全和平台伸缩性等重要主题,能帮助你了解 Swarm 如何组建包含 4700 个节点的集群,并掌握 Swarm 的使用与管理,以及如何使用实现大规模应用的伸缩。

本书适合企业架构、开发、运维等各岗位从业者阅读,同样适合广大想了解当前主流 CaaS 架构内在运行原理与真实场景实践的普通学习者。

Copyright © 2016 Packt Publishing. First published in the English language under the title 'Native Docker Clustering with Swarm'.

本书简体中文版专有出版权由 Packt Publishing 授予电子工业出版社。未经许可,不得以任何方式复制或抄袭本书的任何部分。专有出版权受法律保护。

版权贸易合同登记号 图字: 01-2017-2760

图书在版编目(CIP)数据

Swarm 容器编排与 Docker 原生集群 / (俄罗斯)法布里齐奥·索贝尔萨 (Fabrizio Soppelsa), (泰)占伟·瓦卡斯 (Chanwit Kaewkasi) 著; 崔婧雯, 钟最龙译. —北京: 电子工业出版社, 2017.7

书名原文: Native Docker Clustering with Swarm

ISBN 978-7-121-31792-7

I. ①S… II. ①法… ②占… ③崔… ④钟… III. ①Linux 操作系统—程序设计 IV. ①TP316.85

中国版本图书馆 CIP 数据核字(2017)第 129957 号

策划编辑: 张春雨

责任编辑: 徐津平

印 刷: 北京天宇星印刷厂

装 订: 北京天宇星印刷厂

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: 15 字数: 314 千字

版 次: 2017 年 7 月第 1 版

印 次: 2017 年 7 月第 1 次印刷

定 价: 69.00 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, 联系及邮购电话: (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。

作者介绍

Fabrizio Soppelsa 是一家 OpenStack 公司——Mirantis 的高级工程师。从 Docker 0.3 版本开始，他就是 Docker 的积极使用者和倡导者，他用三个国家的语言发表了 Docker 工具相关的多篇文章。他也是一些项目，特别是 Machine 项目的实际贡献者。他目前生活在俄罗斯的莫斯科，他和他的蜘蛛 Mosha 是那里 Docker 见面会的组织者。

我要感谢 ClusterHQ 的工作人员对 Flocker 的帮助，特别感谢 Ryan Wallner。也要感谢 Yandex 团队和 Denis Kutin 提供了免费的 OpenStack 实验室，让我能够很容易地使用。感谢 Mirantis 创建了（我认为是）最好的 OpenStack 发行版本。感谢 Docker 团队和 Docker 社区带给我的所有快乐。

Chanwit Kaewkasi 是泰国苏兰拉里理工大学计算机学院的助理教授。Chanwit 从 0.1 版本就开始参与 Docker Swarm 项目的贡献，他协同设计并且实现了策略过滤器、ZooKeeper 发现，以及其他特性。他目前是 Docker Swarm 的维护者以及 Docker Captain（社区给 Docker 专家的称号）。

我还要感谢我的妻子——Pitchaya，感谢她的鼓励以及对我的工作，包括这本书的大力支持。

要送给 Docker Engineering 团队特别的感谢，感谢他们开发出的伟大的软件。感谢苏兰拉里理工大学为我提供了绝佳的工作场所。感谢我的父母对我的支持。最后还要感谢 Fabrizio 邀请我合作撰写本书。

审阅者介绍

Baohua Yang 是 IBM 的资深研究员。他的兴趣包括云计算、Fintech、分布式系统和分析的核心领域。他对那些新兴技术特别感兴趣，比如，SDN/NFV、容器、大数据、区块链和认知计算。

作为首席架构师，他领导企业产品的架构设计和系统实现，并且帮助解决了行业解决方案的关键技术难题。

作为开源社区的贡献者，他向数个项目提交代码、方案和演讲，包括 OpenStack、Hyperledger、OpenvSwitch、Docker、OpenDaylight 和 Kubernetes，并且领导了一些项目，包括 easyOVS、Hyperledger Fabric-SDK-py 和 Cello。他现在是中国的 Hyperledger 技术工作组的主席。

他在顶尖的互联网会议和期刊（包括 IEEE INFOCOM, IEEE Trans on Computers）上发表了十多篇文章，并且参与撰写了一些技术书籍和专利。他现在是数个学术会议和期刊的 TPC 成员。

他的主页：<https://yeasy.github.com>。

前言

欢迎来到《Swarm 容器编排与 Docker 原生集群》一书！这是一本关于容器和分布式系统的书。本书将介绍如何使用原生的 Docker 工具建模微服务、生成任务、扩大应用程序的规模，以及将容器推送到 Docker 集群里！一句话来说，本书将讨论 Docker 的编排。

随着最近 Swarm Mode 的崛起，以及 Docker Engine 启用了 Swarm 功能，编排 Docker 的最佳方式其实还是 Docker！

听上去不错，但是“编排 Docker”是什么意思呢？什么是编排？更确切的说法是，什么是管弦乐队？



管弦乐队指的是音乐家的全体，它由指挥家指挥，指挥家负责控制节奏、旋律，塑造出音乐的整体。弦乐队、管乐队、打击乐队、键盘乐队以及其他乐队都会遵循指挥家的指挥，共同演奏出惊人的交响乐曲，比如贝多芬的《第九交响乐》。

类似地，在容器编排系统里，音乐家是任务，指挥家则是领导者服务(Swarm primitives)。任务并不演奏旋律，或者并不仅仅做这些：更为抽象地说，它们执行一些计算型工作，比如，运行 Web 服务器。而指挥家——Swarm，则负责它们的预配，它们的可用性，它们的

链接，它们的扩展。这也就是大家所说的“Docker 编排”。

本书讲述如何预配这样的 Docker “管弦乐队”，如何保证服务的可用性，如何连接任务，以及如何扩展平台，从而演奏出属于应用程序的动人交响乐。

本书范围

第 1 章“欢迎来到 Docker Swarm”会介绍 Swarm，并且解释用户为什么需要集群解决方案来管理容器。这一章介绍 Swarm 的特性，介绍其架构的高层级描述。这一章还设计了一些示例，讲述 Swarm 和 Fleet、Kubernetes、Mesos 的不同之处。之后也会介绍 Docker 工具的安装以及两种 Swarm 的预配方式：本地的 Swarm Standalone 和远程在 DigitalOcean 上的 Swarm Mode 集群。

第 2 章“探索发现服务”是描述性语言最多、最抽象的一章。这一章介绍发现机制和共识算法是什么，以及它们为什么对于分布式系统来说至关重要。本章会详细介绍 Swarm Mode 包含的共识机制 Raft 及其实现 Etcd。还会介绍第 1 章“欢迎来到 Docker Swarm”里所使用的发现机制的局限性，并且使用 Consul 扩展上一章的本地示例，之后重新将其部署。

第 3 章“遇见 Docker Swarm Mode”介绍全新的 Docker kit，它能够帮助用户创建任何规模的任务集群。本章会介绍 Docker Swarm Mode 的基础——SwarmKit，介绍它在 Docker 1.12+ 版本里是如何工作的，讨论其架构、理念，它和“旧”Swarm 的不同之处在哪里，以及它是如何通过抽象出服务和任务来组织工作负载的。

第 4 章“创建生产级别 Swarm”介绍并且讨论了社区驱动的项目——Swarm2k 和 Swarm3k，我们实验了 2300 和 4800 个节点的 Swarm 集群，可以运行成千上万个容器。最后总结了可用计划，预配多大规模的集群，以及实验中的经验教训。

第 5 章“管理 Swarm 集群”主要探讨基础架构。这一章展示如何增加或者降低 Swarm 的规模，如何 promote 以及 demote 节点，以及如何更新集群和节点的属性。这一章还会介绍 Shipyard 和 Portainer.io，其可以作为 Swarm 的图形 UI。

第 6 章“Swarm 上真实应用的部署”介绍了将真实应用程序放到 Swarm 上，并且讨论了 Compose、Docker Stacks 和 Docker Application Bundles。这一章展示了典型的部署 workflow 如何在集群里过滤并且调度容器，将其作为服务启动，将容器作为任务处理。这一章从

定义一个使用 Nginx 的 Web 服务开始，然后部署一个使用 MySQL 的 WordPress，最终介绍一个更为实际的应用：Apache Spark。

第 7 章“平台的向上伸缩”将在前几章的基础上开始新的话题讨论。这一章将介绍 Flocker，给 Swarm 上运行的 Spark 增加存储能力，并且会展示如何安装，以及如何和 Swarm 一起大规模、自动地使用它。这一章将完善之前的 Spark 示例，运行一些真实的大数据 job，并且为该基础架构搭建基础的监控系统。

第 8 章“Swarm 附加特性的探索”讨论了一些对于 Swarm 来说很重要的高级话题，包括 Libnetwork 和 libkv。

第 9 章“Swarm 集群和 Docker 软件供应链的安全加固”关注 Swarm 集群的安全方面，会介绍平台的参数、证书、防火墙等概念，并且会介绍 Notary。

第 10 章“Swarm 和云”介绍了在云供应商那里运行 Swarm 的最为流行的几种方案。将在 AWS 和 Azure 上安装 Swarm，然后介绍 Docker Datacenter，最后会转向 OpenStack，介绍在 Magnum 上如何安装以及管理 Swarm，Magnum 是 OpenStack 提供的容器即服务方案。

第 11 章“Swarm 的未来展望”展望了 Docker 编排的趋势，比如软件定义的基础架构、InfraKit、unikernel 以及 Caas。伟大的征途尚未结束！

阅读本书的要求

本书假定读者有在命令行里使用 Docker 的经验：本书通篇会持续地拉取镜像、运行容器、定义服务、暴露端口以及创建网络。

另外，读者最好对网络协议有一些基本了解，并且熟悉公有云和私有云的概念，比如虚拟机和 tenant 网络。

要实践本书的示例，读者需要 Docker 及其工具。第 1 章“欢迎来到 Docker Swarm”介绍了它们的安装方式。

另外，要想完全学习到示例里的知识，读者还需要能够访问一种公有云（比如 AWS，Azure 或者 DigitalOcean）或者私有云（比如 OpenStack）来初始化出虚拟机。

目标读者

本书写给 Docker 的用户——开发人员和系统管理员，那些想要利用现有的 Swarm 和 Swarmkit 的功能，借助容器大幅扩展应用程序的人们。

约定的格式

本书使用了一系列文本格式来区分不同类型的信息。这里列出了这些格式的示例，并且解释了其含义。

文本里的代码、数据库表名、文件夹名、文件名、文件扩展、路径名、URL、用户输入，以及 Twitter 处理器格式如下：“当执行 `docker swarm init` 时，只需复制和粘贴所输出的行”。

代码如下所示：

```
digitalocean:
  image: "docker-1.12-rc4"
  region: nyc3
  ssh_key_fingerprint: "your SSH ID"
  ssh_user: root
```

命令行输入或输入如下所示：

```
# Set $GOPATH here
go get https://github.com/chanwit/belt
```

新的术语以及重要单词会加重表示。屏幕上可以看到的文本，比如菜单或者对话框，显示为这样的样式：“UI 有预期选项，包括启动容器的一系列模板，比如 **MySQL** 或者私有 **Registry**，但是撰写本书时还不支持 Swarm 服务”。



警告或者重要提示使用这个图标。



建议或技巧使用这个图标。

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/31792>



目录

第 1 章 欢迎来到 Docker Swarm	1
集群工具和容器管理器	3
Swarm 的目标	3
为什么使用 Swarm	4
真实的示例	5
宠物模型 vs 牛群模型	5
Swarm 特性	6
类似项目	7
Kubernetes	7
CoreOS Fleet	8
Apache Mesos	9
Kubernetes vs Fleet vs Mesos	10
Swarm vs 所有	10
Swarm v1 架构	10
术语	12
开始使用 Swarm	13
Mac 系统上的 Docker	14
Windows 系统上的 Docker	16
使用 Linux	18
检查 Docker Machine 是否可用——所有系统	19
以前的 Swarm	19
Boot2Docker	21
使用 Docker Machine 创建 4 个集群节点	21
配置 Docker 主机	24
启动 Docker Swarm	25

启动 Docker Swarm	25
测试 Swarm 集群	29
如今的 Swarm	31
本章小结	35
第 2 章 探索发现服务	36
发现服务	37
Token	38
使用 token 重新架构第 1 章示例	38
Token 的限制	43
Raft	43
Raft 理论	43
实际的 Raft	45
EtcD	47
使用 EtcD 重新架构第 1 章示例	47
ZooKeeper	50
Consul	50
使用 Consul 重新架构第 1 章示例	50
实现去中心化的发现服务	52
本章小结	52
第 3 章 遇见 Docker Swarm Mode	53
SwarmKit	53
版本和支持	54
SwarmKit 架构	54
SwarmKit 的核心: swarmd	56
SwarmKit 的控制器: swarmctl	57
使用 Ansible 预配 SwarmKit 集群	58
在 SwarmKit 上创建服务	62
Swarm Mode	63
Swarm v1 vs Swarm Mode vs SwarmKit	64
深入了解 Swarm Mode 部署	65

本章小结	72
第 4 章 创建生产级别 Swarm	73
工具	73
Swarm2k 的 HA 拓扑	74
管理器配置	75
Raft 恢复场景	75
Raft 文件	76
运行任务	76
管理器拓扑	76
使用 belt 预配基础架构	79
使用 Docker Machine 保护管理器安全	81
理解 Swarm 内部机制	83
加入 worker	84
升级管理器	85
监控 Swarm2k	87
InfluxDB 时间序列数据库	87
Swarm3k	90
Swarm3k 的搭建和工作负载	90
大规模 Swarm 的性能	92
总结 Swarm2k 和 Swarm3k 的经验教训	95
本章小结	96
第 5 章 管理 Swarm 集群	97
Docker Swarm standalone	97
Docker Swarm Mode	98
手动添加节点	99
管理器	99
Worker 数量	100
添加脚本化节点	100
belt	102
使用 Ansible	103

集群管理	105
操作节点	106
降级和升级	107
标记节点	108
移除节点	110
Swarm 健康	111
备份集群配置	111
灾难恢复	112
Swarm 的图形化界面	112
Shipyards	112
Portainer	114
本章小结	115
第 6 章 Swarm 上真实应用的部署	116
微服务	116
部署一个复制的 Nginx	117
一个极简的 Swarm	118
Docker Service	120
overlay 网络	124
集成的负载均衡	124
服务的连接：用 WordPress 例子展示	124
Swarm 的调度策略	127
现在，WordPress	127
Docker Compose 和 Swarm Mode	130
Docker stacks 介绍	130
分布式应用包	131
Docker deploy	132
另外一个应用：Apache Spark	133
为什么要在 Docker 上运行 Spark	134
没有 Swarm 的 Spark 单机	134
在 Swarm 上的 Spark 单机	137

在 Swarm 上启动 Spark.....	138
本章小结	140
第 7 章 平台的向上伸缩.....	141
再次登场的 Spark 例子.....	142
Docker 插件	142
实验室环境	143
一个独一无二的秘钥	143
Docker Machine.....	144
安全组	145
网络配置	146
存储配置和架构	146
安装 Flocker	148
生成 Flocker 证书	149
安装软件	151
安装控制节点	153
安装集群节点	153
测试一切是否正常	154
安装并配置 Swarm.....	156
为 Spark 添加一个卷.....	157
再次部署 Spark.....	157
测试 Spark	159
使用 Flocker 存储.....	161
伸缩 Spark.....	164
监控 Swarm 托管的应用.....	165
Prometheus	165
安装一个监控系统	166
在 Grafana 中导入 Prometheus.....	167
本章小结	169
第 8 章 Swarm 附加特性的探索.....	171
Libnetwork.....	171

Networking 插件	172
容器网络模型	173
加密和路由矩阵	174
MacVLAN	174
overlay 网络	175
网络控制面板	177
Libkv	179
如何使用 libkv	180
本章小结	181
第 9 章 Swarm 集群和 Docker 软件供应链的安全加固	182
软件供应链	182
Swarm 集群的安全加固	183
安全加固 Swarm: 最佳实践	184
证书颁发机构	185
证书和相互 TLS	185
集群加入令牌	185
在 Docker Machine 中添加 TLS	186
Docker Notary	187
Docker Secret 介绍	190
本章小结	192
第 10 章 Swarm 和云	193
Docker for AWS 和 Docker for Azure	193
Docker for AWS	194
Docker for Azure	198
Docker Datacenter	201
OpenStack 上的 Swarm	202
OpenStack Nova	204
当下现实: OpenStack 友好的方式	205
OpenStack Heat	205
OpenStack Magnum	206

本章小结	215
第 11 章 Swarm 的未来展望	216
Provisioning 的挑战	216
软件定义基础设施	216
InfraKit	217
TUF—— The Update Framework	219
Docker Stacks 和 Compose	220
Caas —— 容器即服务	220
Unikernel	220
为 Docker 做贡献	222
Github	222
提交 issue	222
代码	223
belt 和其他项目	223
本章小结	223