Kubernetes on AWS

Deploy and manage production-ready Kubernetes clusters on AWS

Ed Robinson

Packt

BIRMINGHAM - MUMBAI

Kubernetes on AWS

Copyright © 2018 Packt Publishing

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the publisher, except in the case of brief quotations embedded in critical articles or reviews.

Every effort has been made in the preparation of this book to ensure the accuracy of the information presented. However, the information contained in this book is sold without warranty, either express or implied. Neither the author, nor Packt Publishing or its dealers and distributors, will be held liable for any damages caused or alleged to have been caused directly or indirectly by this book.

Packt Publishing has endeavored to provide trademark information about all of the companies and products mentioned in this book by the appropriate use of capitals. However, Packt Publishing cannot guarantee the accuracy of this information.

Commissioning Editor: Gebin George
Acquisition Editor: Rahul Nair
Content Development Editor: Abhishek Jadhav
Technical Editor: Aditya Khadye
Copy Editor: Safis Editing
Project Coordinator: Jagdish Prabhu
Proofreader: Safis Editing
Indexer: Pratik Shirodkar
Graphics: Tom Scaria
Production Coordinator: Aparna Bhagat

First published: November 2018

Production reference: 1291118

Published by Packt Publishing Ltd.
Livery Place
35 Livery Street
Birmingham
B3 2PB, UK.

ISBN 978-1-78839-007-1

www.packtpub.com

Adrian, your support made this book possible

mapt.io

Mapt is an online digital library that gives you full access to over 5,000 books and videos, as well as industry leading tools to help you plan your personal development and advance your career. For more information, please visit our website.

Why subscribe?

- Spend less time learning and more time coding with practical eBooks and Videos from over 4,000 industry professionals
- Improve your learning with Skill Plans built especially for you
- Get a free eBook or video every month
- Mapt is fully searchable
- Copy and paste, print, and bookmark content

Packt.com

Did you know that Packt offers eBook versions of every book published, with PDF and ePub files available? You can upgrade to the eBook version at www.packt.com and as a print book customer, you are entitled to a discount on the eBook copy. Get in touch with us at customercare@packtpub.com for more details.

At www.packt.com, you can also read a collection of free technical articles, sign up for a range of free newsletters, and receive exclusive discounts and offers on Packt books and eBooks.

Contributors

About the author

Ed Robinson works as a senior site reliability engineer at Cookpad's global headquarters in Bristol, UK. He has been working with Kubernetes for the last three years, deploying clusters on AWS to deliver resilient and reliable services for global audiences. He is a contributor to several open source projects and is a maintainer of Træfık, the modern HTTP reverse proxy designed for containers and microservices.

About the reviewer

Manuel Tiago Pereira is a software engineer with vast experience of automating infrastructure provisioning and configuration for systems from development environments to highly available platforms for web applications. For the last couple of years, he has been invested in providing solid platforms for application deployments using Kubernetes. He has dedicated most of his professional career to SaaS companies and he's currently working at Talkdesk in order to make life easier for call-center operators and increase their customers' happiness.

Packt is searching for authors like you

If you're interested in becoming an author for Packt, please visit `authors.packtpub.com` and apply today. We have worked with thousands of developers and tech professionals, just like you, to help them share their insight with the global tech community. You can make a general application, apply for a specific hot topic that we are recruiting an author for, or submit your own idea.

Table of Contents

Preface 1

Chapter 1: Google's Infrastructure for the Rest of Us 7
- **Why do I need a Kubernetes cluster?** 9
  - The roots of containers 10
  - Enter the container 10
    - Cgroups 11
    - Namespaces 11
    - Putting the pieces together 12
  - Here, schedule this... 12
  - The basics of Kubernetes 14
    - The pod 14
  - Labeling all the things 17
  - Replica sets 19
  - Services 20
- **Under the hood** 21
  - API server 21
  - Controller manager 22
  - Scheduler 22
  - Kubelet 23
- **Summary** 23

Chapter 2: Start Your Engines 25
- **Your own Kubernetes** 25
  - Installation 25
    - macOS 27
    - Linux 27
    - Windows 28
  - Starting Minikube 28
  - First steps with kubectl 29
  - Building Docker containers inside the cluster 31
- **Building and launching a simple application on Minikube** 32
  - What just happened? 34
    - Rolling out changes 35
    - Resilience and scaling 36
  - Using the dashboard 38
  - Configuration as code 43
  - Troubleshooting Minikube 45
- **Summary** 45

Chapter 3: Reach for the Cloud 47
- **Cluster architecture** 48

Creating an AWS account 49
Creating an IAM user 49
Getting the CLI 54
Setting up a key pair 54
Preparing the network 57
Setting up a bastion 61
sshuttle 63
Instance profiles 65
Kubernetes software 66
Docker 67
Installing Kubeadm 68
Building an AMI 69
Bootstrapping the cluster 69
What just happened? 72
Access the API from your workstation 74
Setting up pod networking 75
Launching worker nodes 76
Demo time 79
Summary 80

Chapter 4: Managing Change in Your Applications 81
Running pods directly 83
Jobs 85
CronJob 89
Cron syntax 90
Concurrency policy 92
History limits 92
Managing long running processes with deployments 92
kubectl patch 95
kubectl edit 96
kubectl apply 96
Kubernetes dashboard 97
Greater control of your deployments 98
RollingUpdate deployment 99
Recreate deployment 100
DaemonSet 101
Summary 105

Chapter 5: Managing Complex Applications with Helm 107
Installing Helm 108
macOS 108
Linux and Windows 108
Installing Tiller 109
Installing a chart 110
Configuring a chart 113

Creating your own charts 116
Chart.yaml 116
values.yaml 117
templates 117
Making it your own 118
Developing and debugging 120
Templating language 121
Functions 123
Flow control 124
Hooks 128
Packaging Helm charts 131
You can test building an index 132
Using your repository 133
Organizational patterns for Helm 134
Chart per application 134
Shared charts 135
Library charts 136
Next steps 136
Chapter 6: Planning for Production 137
The design process 138
Initial planning 138
Planning for success 140
Planning for a successful roll out 141
Discovering requirements 142
Availability 144
Capacity 148
EC2 instance types 148
EC2 instance types 149
Breadth versus depth 150
Performance 151
Disk performance 151
gp2 152
io2 153
st1 153
sc1 154
Networking 155
Security 156
Always be updating 157
In-place updates 158
Immutable images 159
Network security 160
Infra-node networking 160
Node-master networking 160
External networking 161
Kubernetes infra-pod networking 161

IAM roles 162
Validation 164
Observability 165
Logging 166
Monitoring 167
Blackbox monitoring 169
Alerting 169
Tracing 170
Summary 171

Chapter 7: A Production-Ready Cluster 173
Building a cluster 173
Getting started with Terraform 175
Variables 177
Networking 177
Plan and apply 179
Control Plane 180
Preparing node images 183
Installing Packer 183
Packer configuration 183
Node group 185
Provisioning add-ons 189
Managing change 190
Summary 192

Chapter 8: Sorry My App Ate the Cluster 193
Resource requests and limits 194
Resource units 195
How pods with resource limits are managed 196
Quality of Service (QoS) 197
Resource quotas 199
Default limits 201
Horizontal Pod Autoscaling 203
Deploying the metrics server 203
Verifying the metrics server and troubleshooting 205
Autoscaling pods based on CPU usage 206
Autoscaling pods based on other metrics 210
Autoscaling the cluster 211
Deploying the cluster autoscaler 212
Summary 218

Chapter 9: Storing State 219
Volumes 220
EBS volumes 222
Persistent volumes 225
Persistent volumes example 226