Table of Contents

Preface xiii

1. Python Essentials for DevOps 1
Installing and Running Python 2
The Python Shell 2
Jupyter Notebooks 3
Procedural Programming 3
Variables 4
Basic Math 4
Comments 5
Built-in Functions 5
Print 5
Range 6
Execution Control 6
if/elif/else 7
for Loops 8
while Loops 9
Handling Exceptions 10
Built-in Objects 10
What Is an Object? 11
Object Methods and Attributes 11
Sequences 12
Functions 23
Anatomy of a Function 24
Functions as Objects 25

Anonymous Functions 26
Using Regular Expressions 26
Searching 27
Character Sets 27
Character Classes 28
Groups 29
Named Groups 29
Find All 29
Find Iterator 30
Substitution 30
Compiling 30
Lazy Evaluation 31
Generators 31
Generator Comprehensions 32
More IPython Features 33
Using IPython to Run Unix Shell Commands 33
Exercises 34

2. Automating Files and the Filesystem. 35
Reading and Writing Files 35
Using Regular Expressions to Search Text 44
Dealing with Large Files 46
Encrypting Text 47
Hashing with Hashlib 47
Encryption with Cryptography 47
The os Module 49
Managing Files and Directories Using os.path 51
Walking Directory Trees Using os.walk 54
Paths as Objects with Pathlib 55

3. Working with the Command Line. 57
Working with the Shell 57
Talking to the Interpreter with the sys Module 57
Dealing with the Operating System Using the os Module 58
Spawn Processes with the subprocess Module 59
Creating Command-Line Tools 61
Using sys.argv 62
Using argparse 65
Using click 69

fire 73
Implementing Plug-ins 78
Case Study: Turbocharging Python with Command-Line Tools 79
Using the Numba Just-in-Time (JIT) Compiler 80
Using the GPU with CUDA Python 82
Running True Multicore Multithreaded Python Using Numba 83
KMeans Clustering 84
Exercises 85

4. Useful Linux Utilities. 87
Disk Utilities 88
Measuring Performance 88
Partitions 90
Retrieving Specific Device Information 91
Network Utilities 93
SSH Tunneling 93
Benchmarking HTTP with Apache Benchmark (ab) 93
Load Testing with molotov 95
CPU Utilities 97
Viewing Processes with htop 98
Working with Bash and ZSH 99
Customizing the Python Shell 100
Recursive Globbing 101
Searching and Replacing with Confirmation Prompts 101
Removing Temporary Python Files 103
Listing and Filtering Processes 103
Unix Timestamp 104
Mixing Python with Bash and ZSH 104
Random Password Generator 104
Does My Module Exist? 105
Changing Directories to a Module's Path 106
Converting a CSV File to JSON 107
Python One-Liners 107
Debuggers 108
How Fast Is this Snippet? 108
strace 109
Exercises 112
Case Study Question 112

5. Package Management. 113
Why Is Packaging Important? 114
When Packaging Might Not Be Needed 114
Packaging Guidelines 114
Descriptive Versioning 115
The changelog 116
Choosing a Strategy 117
Packaging Solutions 118
Native Python Packaging 118
Debian Packaging 124
RPM Packaging 131
Management with systemd 137
Long-Running Processes 138
Setting It Up 138
The systemd Unit File 140
Installing the Unit 141
Log Handling 143
Exercises 144
Case Study Question 144

6. Continuous Integration and Continuous Deployment. 145
Real-World Case Study: Converting a Poorly Maintained WordPress Site to Hugo 145
Setting Up Hugo 147
Converting WordPress to Hugo Posts 148
Creating an Algolia Index and Updating It 150
Orchestrating with a Makefile 151
Deploying with AWS CodePipeline 152
Real-World Case Study: Deploying a Python App Engine Application with Google Cloud Build 153
Real-World Case Study: NFSOPS 160

7. Monitoring and Logging. 163
Key Concepts in Building Reliable Systems 163
Immutable DevOps Principles 164
Centralized Logging 164
Case Study: Production Database Kills Hard Drives 165
Did You Build It or Buy It? 166
Fault Tolerance 166

Monitoring 168
Graphite 168
StatsD 169
Prometheus 169
Instrumentation 173
Naming Conventions 176
Logging 177
Why Is It Hard? 177
The basicconfig 178
Deeper Configuration 179
Common Patterns 183
The ELK Stack 184
Logstash 185
Elasticsearch and Kibana 187
Exercises 190
Case Study Question 191

8. Pytest for DevOps. 193
Testing Superpowers with pytest 193
Getting Started with pytest 194
Testing with pytest 194
Differences with unittest 196
pytest Features 198
conftest.py 198
The Amazing assert 199
Parametrization 200
Fixtures 202
Getting Started 202
Built-in Fixtures 204
Infrastructure Testing 206
What Is System Validation? 207
Introduction to Testinfra 208
Connecting to Remote Nodes 209
Features and Special Fixtures 212
Examples 213
Testing Jupyter Notebooks with pytest 216
Exercises 216
Case Study Question 217

9. Cloud Computing. 219
Cloud Computing Foundations 220
Types of Cloud Computing 222
Types of Cloud Services 223
Infrastructure as a Service 223
Metal as a Service 227
Platform as a Service 227
Serverless Computing 228
Software as a Service 231
Infrastructure as Code 232
Continuous Delivery 232
Virtualization and Containers 232
Hardware Virtualization 232
Software Defined Networks 233
Software Defined Storage 233
Containers 234
Challenges and Opportunities in Distributed Computing 235
Python Concurrency, Performance, and Process Management in the Cloud Era 237
Process Management 237
Manage Processes with Subprocess 237
Using Multiprocessing to Solve Problems 240
Forking Processes with Pool() 241
Function as a Service and Serverless 243
High Performance Python with Numba 243
Using Numba Just in Time Compiler 243
Using High-Performance Servers 244
Conclusion 245
Exercises 245
Case Study Questions 246

10. Infrastructure as Code. 247
A Classification of Infrastructure Automation Tools 249
Manual Provisioning 250
Automated Infrastructure Provisioning with Terraform 251
Provisioning an S3 Bucket 252
Provisioning an SSL Certificate with AWS ACM 255
Provisioning an Amazon CloudFront Distribution 256
Provisioning a Route 53 DNS Record 258

Copying Static Files to S3 259
Deleting All AWS Resources Provisioned with Terraform 260
Automated Infrastructure Provisioning with Pulumi 260
Creating a New Pulumi Python Project for AWS 261
Creating Configuration Values for the Staging Stack 265
Provisioning an ACM SSL Certificate 266
Provisioning a Route 53 Zone and DNS Records 267
Provisioning a CloudFront Distribution 269
Provisioning a Route 53 DNS Record for the Site URL 270
Creating and Deploying a New Stack 271
Exercises 273

11. Container Technologies: Docker and Docker Compose. 275
What Is a Docker Container? 276
Creating, Building, Running, and Removing Docker Images and Containers 276
Publishing Docker Images to a Docker Registry 280
Running a Docker Container with the Same Image on a Different Host 281
Running Multiple Docker Containers with Docker Compose 283
Porting the docker-compose Services to a New Host and Operating System 295
Exercises 298

12. Container Orchestration: Kubernetes. 299
Short Overview of Kubernetes Concepts 300
Using Kompose to Create Kubernetes Manifests from docker-compose.yaml 301
Deploying Kubernetes Manifests to a Local Kubernetes Cluster Based on minikube 302
Launching a GKE Kubernetes Cluster in GCP with Pulumi 316
Deploying the Flask Example Application to GKE 319
Installing Prometheus and Grafana Helm Charts 325
Destroying the GKE Cluster 330
Exercises 331

13. Serverless Technologies. 333
Deploying the Same Python Function to the "Big Three" Cloud Providers 336
Installing Serverless Framework 336
Deploying Python Function to AWS Lambda 336
Deploying Python Function to Google Cloud Functions 339
Deploying Python Function to Azure 344
Deploying a Python Function to Self-Hosted FaaS Platforms 348

Deploying Python Function to OpenFaaS 349
Provisioning DynamoDB Table, Lambda Functions, and API Gateway Methods Using the AWS CDK 356
Exercises 375

14. MLOps and Machine learning Engineering. 377
What Is Machine Learning? 377
Supervised Machine Learning 377
Modeling 380
Python Machine learning Ecosystem 382
Deep Learning with PyTorch 383
Cloud Machine learning Platforms 387
Machine learning Maturity Model 388
Machine Learning Key Terminology 388
Level 1: Framing, Scope Identification, and Problem Definition 389
Level 2: Continuous Delivery of Data 390
Level 3: Continuous Delivery of Clean Data 391
Level 4: Continuous Delivery of Exploratory Data Analysis 393
Level 5: Continuous Delivery of Traditional ML and AutoML 393
Level 6: ML Operational Feedback Loop 394
Sklearn Flask with Kubernetes and Docker 395
Sklearn Flask with Kubernetes and Docker 398
EDA 399
Modeling 400
Tune Scaled GBM 401
Fit Model 402
Evaluate 402
adhoc_predict 403
JSON Workflow 404
Scale Input 404
adhoc_predict from Pickle 405
Scale Input 406
Exercises 406
Case Study Question 406
Learning Assessments 407

15. Data Engineering. 409
Small Data 410
Dealing with Small Data Files 410

Write a File 411
Read a File 411
Generator Pipeline to Read and Process Lines 411
Using YAML 412
Big Data 413
Big Data Tools, Components, and Platforms 415
Data Sources 415
Filesystems 416
Data Storage 417
Real-Time Streaming Ingestion 418
Case Study: Building a Homegrown Data Pipeline 419
Serverless Data Engineering 420
Using AWS Lambda with CloudWatch Events 421
Using Amazon CloudWatch Logging with AWS Lambda 421
Using AWS Lambda to Populate Amazon Simple Queue Service 422
Wiring Up CloudWatch Event Trigger 427
Creating Event-Driven Lambdas 428
Reading Amazon SQS Events from AWS Lambda 428
Conclusion 432
Exercises 433
Case Study Question 433

16. DevOps War Stories and Interviews. 435
Film Studio Can't Make Film 436
Game Studio Can't Ship Game 438
Python Scripts Take 60 Seconds to Launch 440
Putting Out a Fire with a Cache and Intelligent Instrumentation 441
You'll Automate Yourself Out of a Job! 442
DevOps Antipatterns 443
No Automated Build Server Antipattern 443
Flying Blind 444
Difficulties in Coordination as an Ongoing Accomplishment 444
No Teamwork 445
Interviews 451
Glenn Solomon 451
Andrew Nguyen 451
Gabriella Roman 453
Rigoberto Roche 454
Jonathan LaCour 456

Ville Tuulos 458
Joseph Reis 460
Teijo Holzer 461
Matt Harrison 463
Michael Foord 464
Recommendations 467
Exercises 468
Challenges 468
Capstone Project 469

Index **471**