

CONTENTS IN DETAIL

ACKNOWLEDGMENTS	xix
------------------------	------------

INTRODUCTION	xxi
---------------------	------------

Who This Book Is For	xxii
What's in This Book	xxii
Prerequisites	xxiv
Online Resources	xxv

PART I: FOUNDATIONS OF SECURE DATA ENGINEERING **1**

1	
DATA ENGINEERING BASICS	3

Common Data Engineering Tasks	3
Getting Data to the SIEM	3
Managing Throughput and Latency	4
Enriching and Standardizing Events	5
Example Architectures	6
A Basic Data Pipeline	6
Proactive Data Retrieval	7
Temporary Data Centralization	7
Event Caching	8
Data Serialization Formats	8
JSON	9
YAML	9
Elastic Common Schema	10
Virtual Machine Setup	10
Windows File Transfer and Shell Access Tools	11
Other Useful Tools	11
Summary	11

2	
NETWORK ENCRYPTION	13

TLS Components	14
Interactions	14
Mutual TLS	15
Generating a Certificate Authority	15
Root Configuration	16
Root Generation	19
Intermediate Configuration	20
Intermediate Signing	21
Chain Files	22

Setting Up TLS for Logstash	23
Flex Configuration	23
Server Keys and Certificates	24
Directory Preparation	25
Logstash Configuration	26
Firewall Ports	27
SSH	27
Enabling SSH	28
Creating SSH Keys	28
Starting the Agent	29
Creating Configuration Files	29
Distributing Public Keys	30
Verifying Connectivity	30
Summary	31

3

SOURCE AND CONFIGURATION MANAGEMENT 33

The Git Version Control Workflow	34
Installation	34
Setting Up GitHub or GitLab	35
Creating New Directories	36
Initializing Locally	36
Cloning the Remote Repository	37
Updating README.md	37
Creating a New File	39
Ignoring Files	40
Creating a Development Branch	41
Adding the Branch	41
Creating a New File	42
Setting the Branch's Remote Origin	43
Merging the Branch into main	44
Performing Optional Cleanup	44
Rebasing and Resetting Your Code	45
Creating a New Test Branch	45
Modifying the Remote main Branch	45
Temporarily Stashing Changes	48
Cleanup	49
Creating a Repository for the Book's Project Code	49
Summary	50

PART II: LOG EXTRACTION AND MANAGEMENT 51

4

ENDPOINT AND NETWORK DATA 53

Collecting Logs with Filebeat	54
Installation	55

Enabling TLS	55
Creating a Configuration File	55
Generating Certificate Signing Requests.	56
Configuration	57
Input Sources.	59
Reading Local Files	59
Applying Parsers and New Fields	60
Listening to the Network.	63
Connecting to External Systems.	64
Processors	65
Controlling Processors with Conditionals	66
Screening Out Data.	68
Modules	69
Sending Outputs	71
Publishing to Kafka	72
Publishing to Redis	74
Writing Output Data to a File	75
Sending Data to Logstash.	76
Pruning and Privatizing Data	77
Filebeat for Windows	78
Summary	80

5 WINDOWS LOGS 81

Collecting Logs with Winlogbeat	82
Event Log Types	82
Application and System Logs	82
The Security Log and Sysmon	83
Other Log Sources.	83
Installation.	84
Enabling TLS	85
Creating a Configuration File	85
Generating a Key Pair and Client Certificate	86
Configuration	87
Starting the Service	89
Collecting Logs from Sysmon and PowerShell	90
Sysmon	90
PowerShell Script Blocks	91
PowerShell Modules	93
Exploring Available Event Logs	93
Working with Processors	94
Summary	96

6 INTEGRATING AND STORING DATA 97

An Elastic Agent Architecture.	98
Setting Up the Environment	99
Enabling TLS.	99
Configuring DNS and Firewall Rules	102
Installing Tools on Multiple Servers	103

Viewing Events in Kibana	114
Configuring Different Technology Integrations	116
Network Metadata	116
Iptables Firewall Logs	117
Integration Guidelines	117
Stand-Alone Agents	118
Preparing the Virtual Machine	119
Receiving a Logstash API Key	119
Configuring a Stand-Alone Agent Policy.	120
Configuring Logstash	121
Elasticsearch Ingest Pipelines	123
Creating a Custom Ingest Pipeline.	123
Testing the Pipeline	124
Summary	125

7

WORKING WITH SYSLOG DATA 127

Logging Priorities	128
Collecting Data with Rsyslog	129
Installation.	129
Enabling TLS	130
Custom Rsyslog Configurations	132
Basic vs. Advanced Formats	132
Adding and Testing New Configurations	133
Advanced Format Syntax	133
Global Properties	133
Input Modules.	134
Output Modules	136
Templates.	137
Rulesets	141
Example Configurations	142
A Client	144
A Server Relay	144
Summary	146

PART III: DATA TRANSFORMATION AND STANDARDIZATION 147

8

DATA MANIPULATION PIPELINES 149

Logstash Pipeline Components	150
Logstash Installation and Setup	150
Enabling TLS.	151
Configuring Java Virtual Machine Options	152
Installing Codecs.	152
Inputs and Outputs	153
HTTP POST Requests	153
API Requests.	156
Reading and Writing Files	159

Kafka	161
Redis	162
Amazon S3 and MinIO	164
Logstash to Logstash	167
Null Output	170
Virtual Addressing in Pipelines	170
Summary	172

9

TRANSFORMATION FILTERS 173

Logstash Filters for Data Transformation	174
Extracting Data from Messy Inputs	175
Extracting Unstructured Data with grok	176
Extracting Reliably Ordered Data with dissect	178
Handling Inaccurate Syslog Timestamps	178
Enriching Data	181
Network Identification with the CIDR Filter	181
Key-Value Lookups with translate	182
Adding Directional Tags	185
Field Comparison	187
Equality and Inequality	188
Field Existence	188
Membership	188
Regular Expressions	190
Parsing Key-Value Pairs from Structured Syslog Data	191
Splitting Fields and Avoiding Backslashes	192
Input	192
Mutate	193
Output	194
Writing Filters with Ruby Scripting	195
Init and Code	195
Array Comparisons	196
Dropping Data to Save CPU	198
Summary	199

PART IV: DATA CENTRALIZATION, AUTOMATION, AND ENRICHMENT 201

10

CENTRALIZING SECURITY DATA 203

Kafka Fundamentals	204
Producers, Consumers, and Consumer Groups	204
Brokers and Controllers	205
Topics	205
Partitions	205
Data Replication	206
Listeners and Metadata	207
Streams and Connectors	207

Creating a Kafka Cluster	207
Setting Up the Network	208
Creating Users and Directories	209
Installing Kafka and a Java Development Kit	209
Enabling TLS	210
Configurations	211
Configuring Server Properties	212
Defining the Connection Settings	214
Creating Cluster IDs	214
Using Systemd for Automatic Startup	215
Running Kafka	216
Starting Brokers	216
Creating Topics	217
Publishing Messages	217
Subscribing to Topics	218
Creating Tool-Specific Topics	218
Connecting to External Tools	219
Rsyslog	219
Filebeat	221
Logstash	223
Kafka Pipeline Design Considerations	225
Summary	226

11

AUTOMATING TOOL CONFIGURATIONS

227

Working with Ansible	228
Inventories	228
Commands, Playbooks, and Roles	228
Gathering System Facts	229
Templates	229
Temporary Files	229
SSH Requirements	230
Installation	230
Creating Inventories	231
Configuration Files	232
Running Ad Hoc Commands	233
Enabling Check Mode	233
Distributing Files	234
Executing Commands	235
Updating Packages	236
Managing Services	236
Rebooting Servers	237
Playbooks for Task Execution	237
Reusable Tasks	240
Delegating Tasks	241
Performing Tasks Locally	241
Writing Dynamic Data to Files	242
Modifying hosts Files	242
Automatically Generating Encryption Keys	243
Appending Text	244
Summary	245

12	ANSIBLE TASKS AND PLAYBOOKS	247
Allowing Firewall Traffic		247
Automatic File Transfers		249
Distributing the Kafka Package		249
Downloading Files from the Internet		250
Cloning Git Repositories		250
Creating Kafka Configuration Templates		251
Installing GPG Keys		253
Responding to Prompts		254
Accessing APIs and Web Services		255
TLS Automation		260
Setting Up the Directory		260
Creating the Root CA		261
Creating the Intermediate CA		262
Generating the CA Chain		263
Creating a Flex Certificate		264
Adding Container Files		265
Creating the TLS Playbook		266
Troubleshooting with debug		268
Summary		270
13	CACHING THREAT INTELLIGENCE DATA	271
Speeding Up Lookups with Caching		272
Installing Redis and Memcached		272
Dividing the Project Servers into Nodes		272
Enabling TLS		273
Configuring Redis		275
Configuring Memcached		276
Working with Redis		277
Working with Memcached		279
Populating and Replicating Redis for Threat Lookups		280
Creating the Leader Cache		281
Receiving Threat Indicators with Logstash		282
Defining Inputs		283
Processing Indicators		283
Handling Custom Uploads		284
Populating Redis with Key-Value Pairs		286
Creating Custom API Keys		288
Outputting Data to Elasticsearch		288
Reading Threat Intelligence		289
Creating Redis Followers		291
Configuring a Logstash Getter		292
Deduplicating and Enriching Data		294
Distributing Threat Intelligence with Memcached		295
Configuring the Cyber Threat Intelligence Node		296
Configuring the Data Node		297
Preventing Drift		303