Network DevOps Series:

KAFKA UP AND RUNNING FOR NETWORK DEVOPS

SET YOUR NETWORK DATA IN MOTION

BY ERIC CHOU

Kafka Up and Running for Network DevOps

Set Your Network Data in Motion

Eric Chou

This book is for sale at
http://leanpub.com/network-devops-kafka-up-and-running

This version was published on 2021-12-12

ISBN 978-1-957046-00-6

Tweet This Book!

Please help Eric Chou by spreading the word about this book on Twitter!

The suggested tweet for this book is:

I just bought Kafka Up and Running for Network DevOps, let's keep data flowing!

The suggested hashtag for this book is #NetDevOps.

Find out what other people are saying about the book by clicking on this link to search for this hashtag on Twitter:

#NetDevOps

For my family, you are my 'why' for everything I do.

I would like to thank the open-source software community. My life would be very different without the many dedicated, talented individuals in the open-source community. Thank you all.

Contents

Introduction

Welcome to the world of data!

Unless you have been living under a rock for the last few years, you know data processing, machine learning, and artificial intelligence are taking over the world. Data exists everywhere around us. We can now check real-time traffic information from online cameras before we even leave the house. We can connect to our thermometers remotely to automatically adjust house temperatures. Better yet, the thermometers can also be self-taught so that they can adjust the temperatures all by themselves. Before our family weekend movie nights, my kids love to leverage the WiFi-enabled lights to match the lighting with our mood.

How do these cameras, lights, and thermometers able to take measurements and generate data? It turns out the cost of small sensors and tiny computing units have been coming down steadily since the early days and now can be integrated into everyday items. However, the generated data by one or two devices might not be sufficient enough to yield meaningful results. After all, traffic information on one street might only benefit a tiny fraction of people who travels on that street, but aggregated traffic information on all streets can help everyone. Generally, it is by aggregating all disperse data sets across hundreds of devices; we are able to derive useful information that helps us with our daily lives. The data are constantly flowing between producers and consumers of data.

Have you ever wondered how these data are being exchanged between data producers and consumers? Does each of the devices provide an API (Application Programming Interface) to be queried? Do each of them

have local databases that persist the data? What about data integrity, transmission latency, or scalability?

There are many tools and projects that address these data streaming and exchange issues. One of the most popular open-source tools widely used by companies large and small alike is Apache Kafka[1].

What is Kafka

You might be thinking, "Don't we already have lots of data storage systems? Why do we need yet-another-storage-system?" You are right, and we do have lots of storage solutions such as relational and non-relational databases, cache systems, big data storage clusters, search solutions, and many more. But in most of the data storage cases, the data is entered in once, stored in the database, then retrieved later when needed. For example, when I visited my dentist for the first time, they asked for my personal information, entered them into a database so for my future visits, they could pull up my record. This is very different than the traffic sensor data example that we discussed.

What sets Kafka apart is it was built from the ground up to treat data as continuous flows of information that are constantly being produced, enhanced, manipulated, and consumed. Instead of a focus on holding in data like databases, key-value stores, search indexes, or caches, Kafka architects itself as a system that allows data to be a continually evolving stream of information.

According to the Apache Kafka project page:

> *Apache Kafka is an open-source distributed event streaming platform used by thousands of companies for high-performance*

[1] https://kafka.apache.org/

> *data pipelines, streaming analytics, data integration, and mission-critical applications.*

Companies known for a large amount of data, such as AirBnb, Datadog, Etsy, and many others across different industries, use Kafka to build their data pipeline. These data pipelines use a variety of services that both produce and consume data in a continuous format.

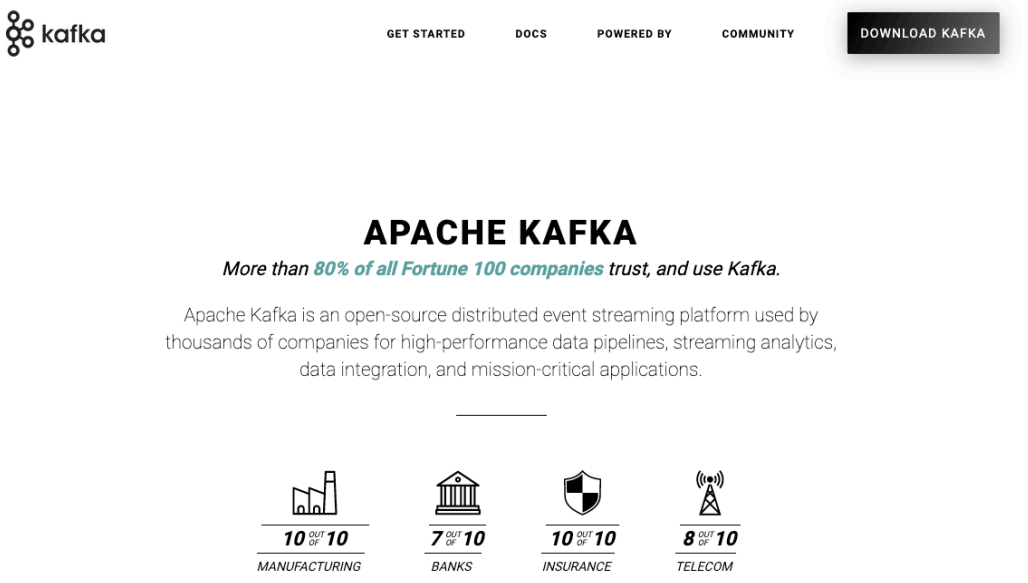

Figure Intro. 1: Powered by Apache Kafka (https://kafka.apache.org/powered-by)

Don't worry if you have not heard of Kafka before or are not sure how, as network DevOps engineers, this tool can help us. We will go a lot deeper into Kafka in this book.

Why do we need Kafka

As a general overview, there are many uses cases for Kafka in network engineers:

- We can use Kafka to stream data, such as logs and NetFlow data, once and be consumed by multiple receivers. Kafka takes care of the ordering of messages, acknowledging receipt to producers, delivery confirmation to consumers, and balancing the data between different recipients.
- We can separate data into logical partitions called *Topics* in a single Kafka cluster. This allows subscribers to only receive the data they are interested in, so the log receiver will not need to receive flow data.
- Kafka allows for an event-driven architecture, such as triggering events based on different types of events. For example, a log receiver can page an on-call engineer if it notices a BGP neighbor of a core device going down.
- Kafka allows us to build a centralized pipeline for network data processing instead of having dispersed teams process bits and pieces of data separately.

These are just some of the use cases of Kafka. By the end of this book, I am sure we will be able to find much more creative use cases.

Prerequisites for this book

Basic knowledge of Linux command line is required to make the most out of this book. We would use command-line tools such as using `cd` for

changing directories, `ls` for listing directories contents, and `pwd` to know where in the directory tree you are currently operating from.

We will be using Python 3 as the programming language in this book. Python is a popular language amongst network engineers with a large ecosystem of tools and libraries. We will use Python to create Kafka producers, consumers and interface with public cloud providers. However, I do not believe you need to be an expert in Python 3 to understand the scripts in this book. If you need a refresher on Python, a good place to go would be the official Python Tutorial[2].

Who this book is for

This book is ideal for IT professionals and engineers who want to take advantage of Kafka's distributed, fault-tolerant streaming data platform. This book can also be used by management to gain a general understanding of Kafka and how it fits into the general IT infrastructure.

What this book covers

Chapter 1. Kafka Introduction, In this chapter, we will cover the general concepts of Kafka. The core architecture, components, and tools. The idea behind Kafka, how it was built, and how the components can help maintain data streams at scale.

Chapter 2. Kafka Installation and Testing, In this chapter, we will install Zookeeper and Kafka on a single Virtual Machine and configure both components. We will also prepare our network lab to be used for future

[2] https://docs.python.org/3/tutorial/

examples. After installation, we will work on a few producer-consumer examples using Kafka command-line tools.

Chapter 3. Kafka Concepts and Examples, In this chapter, we will provide examples of Kafka usage for Producers and Consumers. The producers will write messages to a Topic with consumers receiving the messages. We will look at examples of offset, commit, and acknowledgment for data in the topics.

Chapter 4. Hosted Kafka Services, When we want to move Kafka from our lab setup into production, we can use the Kafka-hosting-as-a-service provided by various cloud providers, such as Amazon AWS or Confluent Cloud. In this chapter, we will provide a step-by-step guide to launch our Kafka cluster using Amazon Managed Streaming for Apache Kafka.

Chapter 5. Cloud Providers Messaging Services, If we are not ready for a managed Kafka cluster, the top public cloud providers, Amazon AWS, Microsoft Azure, and Google Cloud, offer their adopted version of message streaming service. The messaging services have various degrees of Kafka compatibility. In this chapter, we will look at examples of AWS Kinesis, Azure Event Hub, and Google Pub/Sub.

Chapter 6. Network Operations with Kafka, In this chapter, we will explore examples of Kafka in network engineering. We will look at data feeds, data enhancement, and Kafka Connect. The Kafka Connect reuses code provided by the community. We will look at the File and Elasticsearch Kafka connect plugins.

Chapter 7. Other Kafka Considerations and Looking Ahead, In this chapter, we will discuss other Kafka considerations, such as hardware requirements, Broker and Topic configuration, Schema registry, and many more. This chapter will provide additional resources for readers to explore Kafka.

Download the example code files

The code examples used in this book can be downloaded from GitHub at https://github.com/ericchou1/network-devops-kafka-up-and-running[3].

Conventions used

There are a number of text conventions used in this book to help organize the flow. Information in **bold** and *italic* are used to indicate important or special terms.

Code blocks are shown below:

```
print('hello world')
```

Command-line input or output will be shown as follows:

```
$ touch my_script.py

$ ls /
bin   cdrom  etc   lib    lib64   lost+found  mnt  proc  run   snap  sw\
apfile  tmp  var
boot  dev    home  lib32  libx32  media       opt  root  sbin  srv   sy\
s       usr

$ python
Python 3.8.10 (default, Jun  2 2021, 10:49:15)
[GCC 9.4.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> print('hello world')
```

[3] https://github.com/ericchou1/network-devops-kafka-up-and-running

```
14 hello world
15 >>> exit()
```

Warning, tips, and information will be specified in their own special block:

This is a tip section. It will include useful tips and tricks in relation to the topic discussed at hand.

This is an information section. It will provide additional information to help you explore the topic further.

This is a warning blurb. Please pay special attention to this section when they appear, as they will contain important warnings.

Get in touch

Feedbacks from our readers is always welcome and appreciated. Please consider leaving a review on various platforms. They can really help others to discover the book.

All feedback can be submitted to:

book-feedback@networkautomationnerds.com.

Chapter 1. Kafka Introduction

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

History of Kafka

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Kafka Use Cases

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Disadvantages of Kafka

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Kafka Concepts

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Zookeepers

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Brokers

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Topics, Partitions, and Offsets

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Producers and Consumers

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Other Elements in the Ecosystem

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Conclusion

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Chapter 2. Kafka Installation and Testing

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Network Lab Setup

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Kafka Installation Overview

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Install Java

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Download Kafka

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Configure Zookeeper

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Configure Kafka

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Start Zookper and Kafka manually

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Test the Kafka operations

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Configure System Services

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Conclusion

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Chapter 3. Kafka Concepts and Examples

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Producers: Writing Messages

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Producer Example

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Consumers: Receiving Messages

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Consumer Groups

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Consumer Group Example

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Consumer Group with Python

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Offsets in Action

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Kafka Topic Administration

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Replication

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Conclusion

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Chapter 4. Hosted Kafka Services

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

AWS Managed Kafka Service

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Amazon MSK Costs

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Launch Amazon MSK Cluster

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Creating VPC

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Create High Availability subnets

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Create an MSK Cluster.

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Create a Client machine

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Client Setup

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Produce and Consume Data

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Conclusion

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Chapter 5. Cloud Provider Messaging Services

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Amazon Kinesis

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Amazon Kinesis Example

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Azure Event Hub

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Azure Event Hub Example

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Google Cloud Pub/Sub

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

GCP Pub/Sub Python Example

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Conclusion

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Chapter 6. Network Operations with Kafka

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Install Docker

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Install Elasticsearch

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Install Kibana

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Network Data Feed

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Network Data Pipeline

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Network Log as a Service

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Kafka Connect

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Kafka File Connector Example

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-

and-running.

Kafka Elasticsearch Connector Example

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Install Elasticsearch connector

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Conclusion

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Chapter 7. Other Kafka Considerations and Looking Ahead

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Hardware Considerations

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Disk

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Memory

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Network and CPU

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Kafka Broker and Topic Configurations

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Schema Registry

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Kafka Stream Processing

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Cross-Cluster Data Mirroring

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Additional Resources

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Conclusion

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.

Appendix A. Installing Lab Instance in Public Cloud

This content is not available in the sample book. The book can be purchased on Leanpub at http://leanpub.com/network-devops-kafka-up-and-running.