



Neha Narkhede,
Gwen Shapira & Todd Palino

Kafka: The Definitive Guide: Real-Time Data and Stream Processing at Scale

Neha Narkhede, Gwen Shapira, Todd Palino

[Download now](#)

[Read Online ➔](#)

Kafka: The Definitive Guide: Real-Time Data and Stream Processing at Scale

Neha Narkhede , Gwen Shapira , Todd Palino

Kafka: The Definitive Guide: Real-Time Data and Stream Processing at Scale Neha Narkhede , Gwen Shapira , Todd Palino

Every enterprise application creates data, whether it's log messages, metrics, user activity, outgoing messages, or something else. And how to move all of this data becomes nearly as important as the data itself. If you're an application architect, developer, or production engineer new to Apache Kafka, this practical guide shows you how to use this open source streaming platform to handle real-time data feeds.

Engineers from Confluent and LinkedIn who are responsible for developing Kafka explain how to deploy production Kafka clusters, write reliable event-driven microservices, and build scalable stream-processing applications with this platform. Through detailed examples, you'll learn Kafka's design principles, reliability guarantees, key APIs, and architecture details, including the replication protocol, the controller, and the storage layer.

Understand publish-subscribe messaging and how it fits in the big data ecosystem.

Explore Kafka producers and consumers for writing and reading messages

Understand Kafka patterns and use-case requirements to ensure reliable data delivery

Get best practices for building data pipelines and applications with Kafka

Manage Kafka in production, and learn to perform monitoring, tuning, and maintenance tasks

Learn the most critical metrics among Kafka's operational measurements

Explore how Kafka's stream delivery capabilities make it a perfect source for stream processing systems

Kafka: The Definitive Guide: Real-Time Data and Stream Processing at Scale Details

Date : Published September 29th 2017 by O'Reilly Media

ISBN : 9781491936160

Author : Neha Narkhede , Gwen Shapira , Todd Palino

Format : Paperback 322 pages

Genre : Computer Science, Programming, Science, Technology, Reference, Architecture



[Download Kafka: The Definitive Guide: Real-Time Data and Stream ...pdf](#)



[Read Online Kafka: The Definitive Guide: Real-Time Data and Strea ...pdf](#)

Download and Read Free Online Kafka: The Definitive Guide: Real-Time Data and Stream Processing at Scale Neha Narkhede , Gwen Shapira , Todd Palino

From Reader Review Kafka: The Definitive Guide: Real-Time Data and Stream Processing at Scale for online ebook

Steve says

A solid and informative guide. Kafka is powerful and full of rope; this helps make it more comprehensible.

William Anderson says

Technical and Fantastic. This tome is quite exhaustive and serves as a great cover-to-cover read to understand Kafka thoroughly as well as a terrific guide. If delves into the inner workings as well as architectural applications, and comparisons/advantages.

Xanan says

Acceptable for administrators, bad for programmers, and tons of errors.

The book provides a general coverage of Kafka's architecture and internal working.
It details many configuration parameters that affect clustering, replication, message delivery.
It also offers valuable material for system administrators who need to manage and monitor a running cluster.
It's not very good for programmers: Java API coverage is partial and inadequate.
The chapters are uncoordinated and poorly integrated with some repeated material.
Many errors in the code samples and the text.

Chapters 3, 4, and 11 discuss the programming API.
The first two deal with message producers and consumers, present the Java API for publishing and consuming messages and discuss delivery semantics. They also detail configuration options that can be used to customize message producers and consumers. Security and access control is mentioned but never really discussed. There is not a single full program that can be run but several snippets full of errors.
Chapter 11 offers a tutorial introduction to stream processing: what it is and what problems it solves. Three code examples illustrate Kafka Streams, the Stream framework that comes with Kafka and provides a high level abstraction for manipulating data streams. The chapter gives you a taste of what you can do with Kafka Streams but doesn't do much to teach how to use it.

Chapter 2 gives a tutorial on Kafka installation and discusses several configuration options that may help in tuning a Kafka cluster. Basic ZooKeeper knowledge can help understand.

Chapter 5 delves into the internals of replication, partitions, request processing, and message storage on physical files.

Chapter 6 discusses data delivery guarantees. It revisits producer and consumer issues related to message delivery, and how to configure brokers and topics. It also explains how "at least once" delivery is easily achievable while "exactly once" delivery is not.

Chapter 7 briefly explores the Kafka Connect architecture: a producer/consumer alternative to exchange data

between Kafka and another data storage system.

Chapters 8-10 have a more sysadmin-oriented content.

Chapter 8 explores cross cluster data mirroring, why you may need it, available alternative architectures/models and issues of lost or duplicated data you may come across. It also introduces Kafka's own cluster mirroring tool MirrorMaker, its configuration and tuning.

Chapter 9 covers command line tools to create and manage topics and partitions.

Chapter 10 is on monitoring a Kafka cluster and explores JMX metrics exposed by brokers, producers and consumers that can help in monitoring and detecting problems. Basic JMX knowledge is required to follow along.

Greg says

Really good book for Kafka newcomers. Covers all of the topics which developers or application architects would be interested in while choosing a use case specific integration solution. I find the book being detailed enough, so one will not get bored reading it. Of course it is important to remember that the book covers Kafka itself and not 'will it fit my case?'. However, the authors focus on explaining Kafka in a way that it makes easy to understand if Kafka is the right solution for your needs.

Alex says

A fantastic book about Kafka principles and paradigms. Helped me a lot to understand Kafka architecture.

Randy Goering says

Great book

I read this book to start my understanding of Kafka and its usage. I found the book very helpful. It's a good book for anyone needing to setup and use Kafka.

Yin Yunpeng says

excellent introduction about Apache Kafka, especially the tuning tips of each components and detailed explanation on configuration options, very useful. I would recommend it to anyone who I know is using Kafka.

Oleg Prozorov says

This book is an interesting reading. As a person who has never used Kafka I've found enough information

about when and how to use it. One minor inconvinience for me was that a lot of low level configuration and production concerns were described in the second chapter, at this moment I wasn't able to understand them.

Rinat Sharipov says

This is the first thing you should read before started to work with Kafka

Luca says

A great book. The concepts, while hard, are explained carefully. The pace of the book is nice as well.

I would consider this book a 5-star book without the unreasonable number of occurrences of the word "simply". It was unnerving and unnecessary. But if you care about Apache Kafka, this is a required reading.

Marek Skrobacci says

Very interesting introduction to Kafka with just right amount of details. Plenty of practical examples and advice which is quite helpful when you are running Kafka in production. I especially enjoyed chapter about Kafka internals.

The only "bad" think about the book is ton of typos and formatting errors (partially converted markdown, typos, etc), but given that it was pre-release I assume it is going to be edited appropriately.

MaKeDa says

A really clear and concise introduction to Apache Kafka, to really master this tool you would still require to take a good dive into the official documentation but this book really helps to get the things going,

Rod Hilton says

Pretty much what you expect from a "The Definitive Guide" book - it takes you through installing and using Kafka, how to work with it in production, how the internals work, and a laundry list of operations you might want to perform as an admin and how to perform those operations. It's thorough and complete and yet not overly long. It's more than you get from the documentation on the web, a bit better-written and more comprehensive. Plus, hey, it's free <https://www.confluent.io/resources/ka...>

Definitely recommended if you're getting into Kafka. My team was considering using Kafka for a product and I quickly came to feel like I knew what I was talking about during discussions just from reading the first half of this book. Unfortunately, we decided not to go with Kafka for the project but I finished the book anyway and I'd still recommend it.

Kamil says

Excellent introduction into the Apache Kafka. Provides examples of usages and configuration. Explains how internals works. It goes further then just messaging and introduces data pipelines and stream processing with Kafka.

Alex Ott says

Real rating is about 7/10...

Quite good overview of the Apache Kafka, covering most of the aspects, including new additions, like, Kafka Streams.

The cons of the book are:

- Errors, typos, etc. (list will be sent to authors) - as I see, they mostly are from the fact that work on the book started long time ago
- Some topics are covered not very deep, and reader should reference the official documentation. For example, authentication is just mentioned, without saying how to setup it & how to use from clients.
