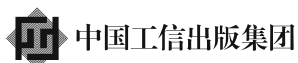


# Cloud-Native Application Architecture

FreeWheel Biz-UI Team

# Cloud-Native Application Architecture

Microservice Development Best Practice



FreeWheel Biz-UI Team  
Biz-UI  
FreeWheel  
Beijing, Beijing, China

ISBN 978-981-19-9781-5                      ISBN 978-981-19-9782-2 (eBook)  
<https://doi.org/10.1007/978-981-19-9782-2>

Jointly published with Publishing House of Electronics Industry, Beijing, PR China  
The print edition is not for sale in China (Mainland). Customers from China (Mainland) please order the print book from: Publishing House of Electronics Industry.  
ISBN of the Co-Publisher's edition: 978-7-121-42274-4

© Publishing House of Electronics Industry 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publishers, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publishers nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publishers remain neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.  
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

# Foreword

As a veteran in the technology field, I have witnessed the dramatic evolution of the software industry in terms of architecture in the past 20 years. I clearly remember when I first joined IBM China Software R&D Center in 2004, the whole team was focused on building projects based on SOA (Service Oriented Architecture). At that time, SOA was still very heavy, and the implementation relied strongly on J2EE, Web Service, SOAP, and other technologies, which had high learning costs for developers and complex implementation and adoptions; customers also needed to take more time to learn how to use the system. In addition, the technology stack was relatively lacking in support for non-functional requirements (such as resilience, performance, and scalability) that needed to be considered in the design phase and eventually implemented, which was a heavy task. Nevertheless, this brought a lot of inspiration and thinking in the system architecture design of the software industry. As a part of the architectural evolution process, SOA has been an important catalyst for the emergence of new technologies and architectural design styles dedicated to decoupling the complex software systems.

Nowadays, microservices and cloud-native technologies have become the mainstream of the software development industry. Many companies, organizations, and teams are migrating or planning to adopt their business to the cloud and refactor their system architecture based on microservices. In fact, regardless of SOA or microservices, the goal is to decouple the complex system and define and wrap the functional modules in a “service,” to realize the independence of system design, development, and maintenance.

But there are two sides to everything. While we benefit from new technologies, we also need to face lots of challenges introduced by them, such as distributed transactions, networking issues, and service discovery. These are only a small part of the problems faced in building microservices architecture. The main process of software design is trade-off under various constraints.

As the leader of global online video advertising technology and innovation, FreeWheel has strong experience of building microservice application. Since 2016, the team has been committed to refactoring the architecture, and after

4 years hard work, the core business system has been rebuilt based on the microservices and migrated to the cloud. In the process of upgrading the technical architecture, we inevitably encountered many unexpected technical challenges, and the team keeps exploring and moving forward all the way, accumulating a lot of valuable experience. This book is derived from our best practice; the colleagues who participated in writing the book are all responsible for development in the frontline and have in-depth understanding of technical details.

Therefore, I highly recommend that the developers interested in microservices and cloud-native technologies read this book, as it will certainly provide a valuable reference for your work.

FreeWheel, Beijing, China

Qiang Wang

# Preface

Practice makes real knowledge. As engineers in the front line, the author team knows the value of real cases to readers and the importance of practice for technical learning. Therefore, most of the cases described in the book come from user scenarios. This is why the book is named “Best Practices.” Of course, we should not underestimate the importance of theoretical learning, and the author team tried to explain accurately the technical concepts based on lots of research, so that the practice can be justified and followed.

The content of the book covers full development lifecycle. No matter what architecture style is used to build a system, it is bound to go through the complete process from design to deployment, and the same is true for microservice applications. Especially with the addition of cloud-native technologies, the development approach and design mindset will be different in many aspects such as technology selection, implementation, and deployment. Therefore, we do not talk about these technologies in a discrete way, but introduce the knowledges of the software development lifecycle step by step based on the development process, in order to bring readers a reasonable and smooth reading experience. From technology selection to service splitting; from agile development to code management; from service governance to quality assurance, the corresponding technologies and practices are shown to readers how to integrate cloud-native technologies into each part of the software development lifecycle.

Sharing customized case. Another feature of this book is that it summarizes some customized practices and tools based on the team’s own experience, such as a low-code development platform for building serverless, a virtual team of middle platform, and an interesting bug bush activity. We believe these will give readers a new feeling. We also hope that these special practices will help you improve your own R&D tools.

This book can be used as a professional book or reference book for industrial practitioners who wish to use the concept of cloud-native technology in practical applications; it should also be useful to senior undergraduate and postgraduate students who intend to access the industry, researchers who work on cloud-native

application architecture and microservices and may also gain inspiration from this book.

Beijing, China  
Nov 2022

FreeWheel Biz-UI Team

# Introduction to the Author Team

## **About FreeWheel Biz-UI Team**

FreeWheel Biz-UI team started to build cloud-native application based on microservice architecture from 2017 and has accumulated rich experience. In addition to developing enterprise-level SaaS, they also lead internal innovative projects such as microservice business middle platform, service governance platform, and serverless low-code development platform.

The team is committed to the popularization and promotion of cloud-native technologies and is keen to share their years of engineering practice. They have established a technical column on microservices practice in InfoQ and published many technical articles. Members of the team have published several technical patents based on their development practices, which have passed the audit of the USPTO. The team also has many experts in cloud-native domain, published books and shared their practices in QCon, AWS Summit, CNCF Webinar, and other offline and online technical conferences.

The authors of this book all come from the team, and they are all experts in the relevant technologies of the chapters they are responsible for, with many years of development experience and a love for technology sharing. The book on microservices best practices covers several cloud-native technology hotspots and provides a detailed summary of the whole development lifecycle practices, which can provide readers with valuable reference.

## **About FreeWheel**

FreeWheel, A Comcast Company is a provider of comprehensive ad platforms for publishers, advertisers, and media buyers. Powered by premium video content, robust data, and advanced technology, the company is revolutionizing the way publishers and marketers transact across all screens, data types, and sales channels.

## List of Authors

Ruofei Ma, Lucas Wang, James Wang, Yu Cao, Kan Xu, Fan Yang, Qi Zhang, Yuming Song, Yanmei Guo, Hanyu Xue and Na Yang

# Acknowledgements

We would like to thank Rufe Ma, the software architect of FreeWheel’s BizUI team, as the first author of this book, he led the team to write and polish the chapters and contents of the book. Thanks to the author team for their contributions, they are: Ruofei Ma, Lucas Wang, James Wang, Yu Cao, Kan Xu, Fan Yang, Qi Zhang, Yuming Song, Yanmei Guo, Hanyu Xue and Na Yang. We would also like to thank all the colleagues who participated in building microservices and cloud adoption for their hard work. We also want to thank our families, whose support made it possible for us to finish the book. In addition, we would like to thank the editor of Springer, whose professionalism is the key to improve the quality of the book.

This book was born in the beginning of the New Year, which signifies the renewal of everything for a year. We also hope this book can help more readers to build a scalable and maintainable cloud-native application from scratch in a brand new way, and wish everyone can easily build enterprise applications with microservices technology, and really enjoy the convenience brought by microservices and cloud-native technology efficiently!

# About the Book

This book introduces the engineering practice of building microservice applications based on cloud-native technology, and there are nine chapters in the book, each chapter is briefly described as follows.

## **Chapter 1: Microservices in the Cloud-Native Era**

This chapter starts from the characteristics of microservices and provides an in-depth analysis of the concepts and core technologies of cloud-native, and what changes microservice applications need to make in the cloud-native era, to complete the development journey from traditional microservices to cloud-native applications.

## **Chapter 2: Microservice Application Design**

In this chapter, we will discuss how to design a microservice application based on the team's practice. The author will talk about the architectural selection of the application and introduce the solutions of architecture, presentation layer, persistence layer, and business layer and will also analyze how to adopt microservice from monolithic systems.

## **Chapter 3: Service Development and Operation**

This chapter will introduce how to go through the whole development process by Scrum agile development method based on the team's engineering practice and introduce the service management and operation and maintenance platform we built to improve the development efficiency.

## **Chapter 4: Microservice Traffic Management**

Service mesh is the preferred solution for traffic management in the cloud-native era. With declarative configuration, you can give your application the ability to control traffic with transparency. This chapter details how to use it to provide traffic control for your microservice applications, based on our practice with service mesh.

## **Chapter 5: Distributed Transactions**

As software systems move from monolithic applications to microservices and cloud-native, and the trend of database decentralization, local transactions on monolithic

applications will transform into distributed transactions, posing a challenge to the need for data consistency. This chapter will introduce our team's practice of using Saga pattern to implement distributed transactions.

### **Chapter 6: Serverless Architecture**

The advantages of building elastic and scalable applications through serverless computing are becoming more and more obvious. As an emerging application architecture, what are its core concepts, what are its features that distinguish it from traditional architectures, its advantages and application scenarios, and what changes it can bring to the building of applications? This chapter will explain each of these issues.

### **Chapter 7: Service Observability**

There are many challenges when using microservice architecture and migrating to the cloud, especially how to know the status and behavior of the application, how to quickly find and solve online problems, and how to monitor the invocation chain between services, all of which will have to be faced. Building observable applications is an important factor in ensuring service quality. This chapter will introduce the definition and application of service observability.

### **Chapter 8: Quality Assurance Practices**

In this chapter, we will introduce some of the practical experiences related to quality assurance accumulated in the process of building microservice applications, and tell how the team can build a quality assurance system in the cloud-native era through sound testing and chaos engineering.

### **Chapter 9: Continuous Integration and Continuous Deployment**

Continuous integration and continuous deployment are necessary to build cloud-native applications. In this chapter, we will talk about the automation triggering, differential execution, and unified product archiving of continuous integration and introduce the product release planning and cloud-native-based deployment framework after microservicization, as well as the full lifecycle support of continuous deployment for microservice applications.

# Contents

<b>1</b>	<b>Microservices in the Cloud Native Era</b>	<b>1</b>
1.1	Starting with Microservices	1
1.1.1	Characteristics of a Microservice Architecture	2
1.1.2	Microservice Trade-Offs	7
1.2	Cloud-Native Applications	10
1.2.1	What Is Cloud-Native	11
1.2.2	Cloud-Native Technologies	14
1.2.3	Characteristics of Cloud-Native Applications	18
1.3	From Microservices to Cloud-Native	20
1.3.1	Adjustment of Nonfunctional Requirements	20
1.3.2	Change in Governance	21
1.3.3	Deployment and Release Changes	22
1.3.4	From Microservice Applications to Cloud-Native Applications	23
1.4	Summary	25
<b>2</b>	<b>Microservice Application Design</b>	<b>27</b>
2.1	Application Architecture Design	27
2.1.1	Service Architecture Selection	27
2.1.2	Service Communication Strategy	34
2.1.3	Storage Layer Design and Selection	42
2.2	Legacy System Modification	45
2.2.1	Greenfield and Brownfield	46
2.2.2	Strangler Pattern	47
2.3	Business Logic Design	52
2.3.1	Splitting Services	53
2.3.2	Design API	60
2.4	Summary	65

- 3 Service Development and Operation . . . . . 67**
  - 3.1 Agile Software Development . . . . . 67
    - 3.1.1 From Waterfall Model to Agile Development . . . . . 68
    - 3.1.2 Scrum in Practice . . . . . 70
  - 3.2 Runtime Environment . . . . . 76
    - 3.2.1 Development Environment . . . . . 76
    - 3.2.2 Test Environment . . . . . 78
    - 3.2.3 Staging Environment . . . . . 79
    - 3.2.4 Production Environment . . . . . 79
  - 3.3 Code Management . . . . . 80
    - 3.3.1 Git Branch Management . . . . . 80
    - 3.3.2 Code Inspection Based on Sonar . . . . . 84
    - 3.3.3 Code Review . . . . . 87
    - 3.3.4 Commit and Merge . . . . . 89
  - 3.4 Low-Code Development Platform . . . . . 90
    - 3.4.1 Low Code and Development Platform . . . . . 91
    - 3.4.2 Low-Code Development Platform Practices . . . . . 91
  - 3.5 Service Operation and Maintenance Platform . . . . . 96
    - 3.5.1 Problems to Be Solved by the Platform . . . . . 96
    - 3.5.2 Platform Architecture . . . . . 97
    - 3.5.3 Platform Modules . . . . . 97
  - 3.6 Service Middle Platform . . . . . 101
    - 3.6.1 What Is a Middle Platform . . . . . 101
    - 3.6.2 The Road to Building a Middle Platform . . . . . 103
  - 3.7 Summary . . . . . 108
- 4 Microservice Traffic Management . . . . . 109**
  - 4.1 Traffic Management in the Cloud-Native Era . . . . . 109
    - 4.1.1 Flow Type . . . . . 110
    - 4.1.2 Service Mesh . . . . . 111
  - 4.2 Service Discovery . . . . . 113
    - 4.2.1 Traditional Services Discovery Problems on the Cloud . . . . . 114
    - 4.2.2 Kubernetes’ Service Discovery Mechanism . . . . . 115
  - 4.3 Using the Istio Service Mesh for Traffic Management . . . . . 117
    - 4.3.1 Core CRDs . . . . . 118
    - 4.3.2 Istio-Based Traffic Management Practices . . . . . 128
    - 4.3.3 Common Adoption Problems and Debugging . . . . . 136
  - 4.4 Improving Application Fault Tolerance with Istio . . . . . 145
    - 4.4.1 Circuit Breaking . . . . . 145
    - 4.4.2 Timeouts and Retries . . . . . 149
  - 4.5 Summary . . . . . 152

- 5 Distributed Transactions** . . . . . 153
  - 5.1 Theoretical Foundations . . . . . 153
    - 5.1.1 Background . . . . . 153
    - 5.1.2 ACID: Transaction Constraints in the Traditional Sense . . . 156
    - 5.1.3 CAP: The Challenge of Distributed Systems . . . . . 157
    - 5.1.4 BASE: The Cost of High Availability . . . . . 158
    - 5.1.5 Write Order . . . . . 158
  - 5.2 Solution Selection for Distributed Transaction Framework . . . . . 159
    - 5.2.1 Existing Research and Practice . . . . . 159
    - 5.2.2 Design Goals of Distributed Transaction Framework . . . . . 164
    - 5.2.3 Choosing Saga . . . . . 165
    - 5.2.4 Introducing Kafka . . . . . 166
    - 5.2.5 System Architecture . . . . . 169
    - 5.2.6 Business Process . . . . . 170
  - 5.3 Distributed Transactions Based on Saga and Kafka in Practice . . . . 171
    - 5.3.1 Improvements to Kafka’s Parallel Consumption Model . . . 171
    - 5.3.2 Deployment Details . . . . . 173
    - 5.3.3 System Availability Analysis . . . . . 173
    - 5.3.4 Production Issues and Handling . . . . . 174
  - 5.4 Chapter Summary . . . . . 177
- 6 Serverless Architecture** . . . . . 179
  - 6.1 What Is Serverless Architecture . . . . . 179
    - 6.1.1 Definition of Serverless Architecture . . . . . 179
    - 6.1.2 Development of Serverless Architecture . . . . . 181
    - 6.1.3 Advantages of Serverless Architecture . . . . . 182
    - 6.1.4 Shortcomings of Serverless Architecture . . . . . 183
  - 6.2 Serverless Architecture Applications . . . . . 185
    - 6.2.1 Building Web API Backend Services . . . . . 185
    - 6.2.2 Building the Data Orchestrator . . . . . 187
    - 6.2.3 Building Timed Tasks . . . . . 188
    - 6.2.4 Building Real-Time Stream Processing Services . . . . . 189
  - 6.3 Serverless Architecture in Practice . . . . . 192
    - 6.3.1 Why AWS Lambda . . . . . 192
    - 6.3.2 Import and Processing of Large Amounts of Data . . . . . 193
    - 6.3.3 Log Collection and Processing . . . . . 201
  - 6.4 Summary of This Chapter . . . . . 211
- 7 Service Observability** . . . . . 213
  - 7.1 What Is Observability . . . . . 213
    - 7.1.1 Definition of Observability . . . . . 213
    - 7.1.2 Three Pillars of Observability . . . . . 214
    - 7.1.3 The Difference and Relation Between Observability and Monitoring . . . . . 215
    - 7.1.4 Community Products and Technology Selection . . . . . 217

7.2	Logging Solutions Under Cloud-Native . . . . .	218
7.2.1	Log Classification and Design . . . . .	218
7.2.2	Evolution of Cloud-Native Log Collection Scheme . . . . .	227
7.2.3	Displaying Logs with Kibana . . . . .	234
7.3	Distributed Tracing . . . . .	246
7.3.1	Core Concepts of Distributed Tracing System . . . . .	246
7.3.2	Jaeger-Based Tracing Solution . . . . .	247
7.4	Metrics . . . . .	255
7.4.1	Collecting Metrics with Prometheus . . . . .	256
7.4.2	Displaying Metrics with Grafana . . . . .	264
7.5	Monitoring and Alerting . . . . .	266
7.5.1	Monitoring Platform . . . . .	266
7.5.2	Alert System . . . . .	277
7.6	Summary . . . . .	281
<b>8</b>	<b>Quality Assurance Practices . . . . .</b>	<b>283</b>
8.1	Quality Assurance System . . . . .	283
8.1.1	Quality Challenges . . . . .	284
8.1.2	Testing Strategy . . . . .	285
8.1.3	Building a Quality Assurance System . . . . .	287
8.2	Testing Practices . . . . .	290
8.2.1	Unit Testing and Mock Practice . . . . .	290
8.2.2	Godog-Based Integration Testing Practices . . . . .	296
8.2.3	Cypress-Based End-to-End Testing Practices . . . . .	302
8.2.4	Test Automation . . . . .	305
8.3	Chaos Engineering . . . . .	309
8.3.1	The Core Concept of Chaos Engineering . . . . .	309
8.3.2	How to Run Chaos Experiments . . . . .	316
8.3.3	Fault Injection Experiments with System Resources . . . . .	323
8.3.4	Service Mesh-Based Network Traffic Fault Injection Method . . . . .	333
8.4	Quality Assurance of Production-Like Environments . . . . .	338
8.4.1	Monitoring and Analysis of Online Services . . . . .	339
8.4.2	Bug Bash Practice . . . . .	341
8.4.3	Post-release Check Practice . . . . .	344
8.4.4	Disaster Recovery Strategies and Practices . . . . .	346
8.5	Summary . . . . .	350
<b>9</b>	<b>Continuous Integration and Continuous Deployment . . . . .</b>	<b>351</b>
9.1	Git-Based Continuous Integration . . . . .	351
9.1.1	Auto-Trigger Pipeline . . . . .	352
9.1.2	Pipeline Differentiation and Unified Collaboration . . . . .	359
9.1.3	Pipeline Product Storage Planning . . . . .	363
9.2	Helm-Based Continuous Deployment . . . . .	365
9.2.1	Deployment Planning . . . . .	366

- 9.2.2 Framework for the Multiple Clusters Deployment in Different Environments . . . . . 368
- 9.2.3 Cloud-Native Support and Task Maintenance . . . . . 374
- 9.3 Continuous Deployment Practices Based on Kubernetes . . . . . 377
  - 9.3.1 Pod Resource Quotas and Horizontal Autoscaling . . . . . 377
  - 9.3.2 Service Online/Offline Workflow and Fault Analysis . . . . . 379
- 9.4 Summary . . . . . 382