



Learn Microservices with Spring Boot

A Practical Approach to RESTful Services
Using an Event-Driven Architecture,
Cloud-Native Patterns, and Containerization

Second Edition

Moisés Macero García

Apress®

Learn Microservices with Spring Boot

A Practical Approach to RESTful
Services Using an Event-Driven
Architecture, Cloud-Native Patterns,
and Containerization

Second Edition

Moisés Macero García

Apress®

Learn Microservices with Spring Boot: A Practical Approach to RESTful Services Using an Event-Driven Architecture, Cloud-Native Patterns, and Containerization

Moisés Macero García
New York, NY, USA

ISBN-13 (pbk): 978-1-4842-6130-9
<https://doi.org/10.1007/978-1-4842-6131-6>

ISBN-13 (electronic): 978-1-4842-6131-6

Copyright © 2020 by Moisés Macero García

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, 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.

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Managing Director, Apress Media LLC: Welmoed Spaehr
Acquisitions Editor: Steve Anglin
Development Editor: Matthew Moodie
Coordinating Editor: Mark Powers

Cover designed by eStudioCalamar

Cover image by Ash from Modern Afflatus on Unsplash (www.unsplash.com)

Distributed to the book trade worldwide by Apress Media, LLC, 1 New York Plaza, New York, NY 10004, U.S.A. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail booktranslations@springernature.com; for reprint, paperback, or audio rights, please e-mail bookpermissions@springernature.com.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this book is available to readers on GitHub via the book's product page, located at www.apress.com/9781484261309. For more detailed information, please visit www.apress.com/source-code.

Printed on acid-free paper

To my family, especially to my mom and my wife. All of you keep teaching me many things, amongst them care, love, and curiosity, the most important learnings in life.

In memory of my dad, an engineer at heart. You encouraged me to keep learning, sparked my curiosity thanks to your inventions, and taught me to be pragmatic to solve the problem at hand.

This book has a bit of all of you.

Table of Contents

About the Author	xiii
About the Technical Reviewers	xv
Chapter 1: Setting the Scene	1
Who Are You?	2
How Is This Book Different from Other Books and Guides?	3
Learning: An Incremental Process	3
Is This a Guide or a Book?	4
From Basics to Advanced Topics.....	4
Skeleton with Spring Boot, the Professional Way.....	5
Test-Driven Development	5
Microservices	6
Event-Driven System	6
Nonfunctional Requirements	6
Online Content	7
Summary.....	7
Chapter 2: Basic Concepts.....	9
Spring.....	9
Spring Boot	10
Lombok and Java.....	12
Testing Basics	15
Test-Driven Development	15
Behavior-Driven Development	15
JUnit	16
Mockito.....	17
AssertJ	20

TABLE OF CONTENTS

Testing in Spring Boot.....	21
Logging	21
Summary and Achievements	23
Chapter 3: A Basic Spring Boot Application	25
Setting Up the Development Environment	26
The Skeleton Web App	27
Spring Boot Autoconfiguration	31
Three-Tier, Three-Layer Architecture	36
Modeling Our Domain	39
Domain Definition and Domain-Driven Design	39
Domain Classes	41
Business Logic	43
What We Need	44
Random Challenges.....	44
Attempt Verification	48
Presentation Layer	52
REST	52
REST APIs with Spring Boot.....	53
Designing Our APIs	55
Our First Controller	56
How Automatic Serialization Works.....	58
Testing Controllers with Spring Boot.....	61
Summary and Achievements	73
Chapter 4: A Minimal Front End with React.....	75
A Quick Intro to React and Node	76
Setting Up the Development Environment	76
The React Skeleton.....	78
A JavaScript Client	80

TABLE OF CONTENTS

The Challenge Component.....	81
The Main Structure of a Component.....	85
Rendering	87
Integration with the App	89
Running Our Front End for the First Time.....	90
Debugging.....	91
Adding CORS Configuration to the Spring Boot App.....	93
Playing with the Application.....	94
Deploying the React App	95
Summary and Achievements	98
Chapter 5: The Data Layer	101
The Data Model.....	102
Choosing a Database	105
SQL vs. NoSQL.....	105
H2, Hibernate, and JPA	106
Spring Boot Data JPA.....	107
Dependencies and Autoconfiguration.....	107
Spring Boot Data JPA Technology Stack.....	111
Data Source (Auto)configuration	113
Entities	115
Repositories	120
Storing Users and Attempts	124
Displaying Last Attempts	130
Service Layer.....	130
Controller Layer	132
User Interface	138
Playing with the New Feature	144
Summary and Achievements	146

TABLE OF CONTENTS

Chapter 6: Starting with Microservices.....	149
The Small Monolith Approach	149
Why a Small Monolith?	150
The Problems with Microservices from Day Zero.....	150
Small Monoliths Are for Small Teams.....	151
Embracing Refactoring.....	152
Planning the Small Monolith for a Future Split	152
New Requirements and Gamification.....	154
Gamification: Points, Badges, and Leaderboards	155
Moving to Microservices.....	156
Independent Workflows.....	157
Horizontal Scalability	158
Fine-Grained Nonfunctional Requirements	160
Other Advantages	160
Disadvantages	161
Architecture Overview.....	163
Designing and Implementing the New Service	164
Interfaces	165
The Spring Boot Skeleton for Gamification.....	165
Domain	166
Service	172
Data	183
Controller.....	186
Configuration.....	188
Changes in Multiplication Microservice	190
UI	195
Playing with the System	201
Fault Tolerance.....	203
The Challenges Ahead.....	205
Tight Coupling.....	206
Synchronous Interfaces vs. Eventual Consistency	206

TABLE OF CONTENTS

Transactions	211
API Exposure.....	213
Summary and Achievements	213
Chapter 7: Event-Driven Architectures.....	215
Core Concepts.....	215
The Message Broker.....	216
Events and Messages.....	218
Thinking in Events	219
Asynchronous Messaging.....	223
Reactive Systems	226
Pros and Cons of Going Event-Driven	226
Messaging Patterns	229
Publish-Subscribe	230
Work Queues	230
Filtering	230
Data Durability	230
Message Broker Protocols, Standards, and Tools	231
AMQP and RabbitMQ	232
Overall Description	233
Exchange Types and Routing.....	234
Message Acknowledgments and Rejection	236
Setting Up RabbitMQ.....	237
Spring AMQP and Spring Boot.....	239
Solution Design	239
Adding the AMQP Starter.....	241
Event Publishing from Multiplication.....	243
Gamification as a Subscriber	250
Analysis of Scenarios.....	259
Happy Flow.....	260
Gamification Becomes Unavailable	266

TABLE OF CONTENTS

The Message Broker Becomes Unavailable	269
Transactionality	270
Scaling Up Microservices	274
Summary and Achievements	279
Chapter 8: Common Patterns in Microservice Architectures	283
Gateway	284
Spring Cloud Gateway	286
The Gateway Microservice	290
Changes in Other Projects	294
Running the Gateway Microservice.....	296
Next Steps	297
Health.....	298
Spring Boot Actuator	299
Including Actuator in Our Microservices.....	302
Service Discovery and Load Balancing.....	304
Consul.....	310
Spring Cloud Consul	312
Spring Cloud Load Balancer	319
Service Discovery and Load Balancing in the Gateway.....	323
Playing with Service Discovery and Load Balancing.....	327
Configuration per Environment	337
Configuration in Consul	340
Spring Cloud Consul Config	341
Implementing Centralized Configuration	343
Centralized Configuration in Practice	347
Centralized Logs	355
Log Aggregation Pattern.....	356
A simple Solution for Log Centralization	357
Consuming Logs and Printing Them.....	361

TABLE OF CONTENTS

Distributed Tracing	368
Spring Cloud Sleuth.....	369
Implementing Distributed Tracing	371
Containerization	373
Docker	376
Spring Boot and Buildpacks	380
Running Our System in Docker.....	381
Dockerizing Microservices	382
Dockerizing the UI	383
Dockerizing the Configuration Importer.....	384
Docker Compose	387
Scaling Up the System with Docker	401
Sharing Docker Images	404
Platforms and Cloud-Native Microservices.....	408
Container Platforms.....	408
Application Platforms	410
Cloud Providers	410
Making a Decision	412
Cloud-Native Microservices	413
Conclusions.....	413
Afterword.....	417
Index.....	419

About the Author



Moisés Macero García has been a software developer since he was a kid, when he started playing around with BASIC on his ZX Spectrum. During his career, Moisés has most often worked in development and architecture for small and large projects, and for his own startups too. He enjoys making software problems simple, and he likes working in teams where he not only coaches others but also learns from them.

Moisés is the author of the blog thepracticaldeveloper.com, where he shares solutions for technical challenges, guides, and his view on ways of working in IT companies. He also organizes workshops for companies that need a practical approach to software engineering. In his free time, he enjoys traveling and hiking. You can follow Moisés or contact him on his Twitter account: [@moises_macero](https://twitter.com/moises_macero).

About the Technical Reviewers



Diego Rosado Fuentes has always enjoyed computers, electronics, and programming. As a child, he liked to disassemble computers and electronic toys. After finishing his studies as a computer science engineer, he started working as a software developer. He worked in several companies, beginning in startups and later in bigger companies. He has been working as a software developer but always somehow doing stuff related to teaching and learning, which he loves.

He likes physics, astronomy, biology, and all sciences, but he also enjoys reading a novel now and then. He considers himself a proactive curious human being trying always to learn something new and comprehend every day a little bit more of all the unrevealed mysteries still to come. You can find him on his Twitter account: @rosado_diego.

Vinay Kumtar is a technology evangelist. He has extensive experience in designing and implementing large-scale projects in enterprise technologies in various consulting and system integration companies. He started working in middleware/integration in 2008 and fell in love with it. His passion helped him achieve certifications in Oracle ADF, Oracle SOA, BPM, WebCenter Portal, and Java. Experience and in-depth knowledge have helped him evolve into a focused domain expert and a well-known technical blogger. He loves to spend his time mentoring, writing professional blogs, publishing white papers, and maintaining a dedicated education channel at YouTube for ADF/Webcenter. Vinay has been contributing to various topics, such as Oracle SOA/BPM/Webcenter/Kafka/microservices/events, for the community by publishing technical articles on his blog Techartifact.com that has 250+ articles. He is a technology leader and currently leading two major digital transformation projects.

ABOUT THE TECHNICAL REVIEWERS



Manuel Jordan Elera is an autodidactic developer and researcher who enjoys learning new technologies for his own experiments, which focus on finding new ways to integrate them.

Manuel won the 2010 Springy Award – Community Champion and Spring Champion 2013. In his little free time, he reads the Bible and composes music on his bass and guitar.

Manuel believes that constant education/training is essential for all developers. You can reach him mostly through his Twitter account: @dr_pompeii.

CHAPTER 1

Setting the Scene

Microservices are becoming more popular and widely used these days. This is not a surprise; this software architecture style has many advantages, such as flexibility and ease of scale. Mapping microservices into small teams in an organization also gives you a lot of efficiency in development. However, going on the adventure of microservices knowing only the benefits is a wrong call. Distributed systems introduce extra complexity, so you need to understand what you are facing and be prepared for that in advance. You can get a lot of knowledge from many books and articles on the Internet, but when you get hands-on with the code, the story changes.

This book covers some of the most important concepts of microservices in a practical way, but not without explaining those concepts. First, we define a use case: an application to build. Then we start with a small monolith, based on some sound reasoning. Once we have the minimal application in place, we evaluate whether it's worthy to move to microservices, and what would be a good way to do so. With the introduction of the second microservice, we analyze the options we have for their communication. Then, we can describe and implement the event-driven architecture pattern to reach loose coupling by informing other parts of the system about *what happened*, instead of explicitly calling others to action. When we reach that point, we notice that a poorly designed distributed system has some flaws that we have to fix with some popular patterns: service discovery, routing, load balancing, traceability, etc. Adding them one by one to our codebase instead of presenting all of them together helps us understand these patterns. We also prepare these microservices for a cloud deployment using Docker and compare the different platform alternatives to run the applications.

The advantage of going step-by-step, pausing when it's needed to settle down the concepts, is that you will understand which problem each tool is trying to solve. That's why the evolving example is an essential part of this book. You can also grasp the concepts without coding one single line since the source code is presented and explained throughout the chapters.

All the code included in this book is available on GitHub in the project Book-Microservices-v2. There are multiple repositories available, divided into chapters and sections, which makes it easier for you to see how the application evolves. The book includes notes with the version that is being covered in each part.

Who Are You?

Let's first start with this: how interesting is this book going to be for you? This book is practical, so let's play this game. If you identify with any of these statements, this book might be good for you:

- “I would like to learn how to build microservices with Spring Boot and how to use the related tools.”
- “Everybody is talking about microservices, but I have no clue what a microservice is yet. I have read only theoretical explanations or just hype-enforcing articles. I can't understand the advantages, even though I work in IT....”
- “I would like to learn how to design and develop Spring Boot applications, but all I find are either quick-start guides with too simple examples or lengthy books that resemble the official documentation. I would like to learn the concepts following a more realistic project-guided approach.”
- “I got a new job, and they're using a microservices architecture. I've been working mainly in big, monolithic projects, so I'd like to have some knowledge and guidance to learn how everything works there, as well as the pros and cons of this architecture.”
- “Every time I go to the cafeteria developers are talking about microservices, gateways, service discovery, containers, resilience patterns, etc. I can't socialize anymore with my colleagues if I don't get what they're saying.” (This one is a joke; don't read this book because of that, especially if you're not interested in programming.)

Regarding the knowledge required to read this book, the following topics should be familiar to you:

- Java (we use Java 14)
- Spring (you don't need strong experience, but you should know at least how dependency injection works)
- Maven (if you know Gradle, you'll be fine as well)

How Is This Book Different from Other Books and Guides?

Software developers and architects read many technical books and guides, either because we're interested in learning new technologies or because our work requires it. We need to do that anyway since it's a constantly changing world. We can find all kinds of books and guides out there. Good ones are usually those from which you learn quickly and that teach you not only how to do stuff but also why you should do it that way. Using new techniques just because they're new is the wrong way to go about it; you need to understand the reasoning behind them so you use them in the best way possible.

This book uses that philosophy: it navigates through the code and design patterns, explaining the reasons to follow one way and not others.

Learning: An Incremental Process

If you look at the guides available on the Internet, you'll notice quickly that they are not real-life examples. Usually, when you apply those cases to more complex scenarios, they don't fit. Guides are too shallow to help you build something real.

Books, on the other hand, are much better at that. There are plenty of good books explaining concepts around an example; they are good because applying theoretical concepts to code is not always easy if you don't see the code. The problem with some of these books is that they're not as practical as guides. You need to read them first to understand the concepts and then code (or see) the example, which is frequently given as a whole piece. It's difficult to put into practice concepts when you see the final version directly. This book stays on the practical side and starts with code that evolves through the chapters so that you grasp the concepts one by one. We cover the problem before exposing the solution.

Because of this incremental way of presenting concepts, this book also allows you to code as you learn and to reflect on the challenges by yourself.

Is This a Guide or a Book?

The pages you have in front of you can't be called a guide: it won't take you 15 or 30 minutes to finish them. Besides, each chapter introduces all the required topics to lay the foundations for the new code additions. But this is not the typical book either, in which you go through isolated concepts illustrated with some scattered code fragments that are made specifically for that situation. Instead, you start with a real-life application that is not yet optimal, and you learn how to evolve it, after learning about the benefits you can extract from that process.

That does not mean you can't just sit down and read it, but it's better if you code at the same time and play with the options and alternatives presented. That's the part of the book that makes it similar to a guide.

In any case, to keep it simple, from here onward we call this a book.

From Basics to Advanced Topics

This book focuses first on some essential concepts to understand the rest of the topics (Chapter 2): Spring Boot, testing, logging, etc. Then, it covers how to design and implement a production-ready Spring Boot application using a well-known layered design, and it dives into how to implement a REST API, the business logic, and database repositories (Chapters 3 and 5). While doing it, you'll see how Spring Boot works internally so it's not magic for you anymore. You'll also learn how to build a basic front-end application with React (Chapter 4), because that will help you visualize how the backend architecture impacts the front end. After that, the book enters into the microservices world with the introduction of a second piece of functionality in a different Spring Boot app. The practical example helps you analyze the factors that you should examine before making the decision of moving to microservices (Chapter 6). Then, you'll get to know the differences between communicating microservices synchronously and asynchronously and how an event-driven architecture can help you keep your system components decoupled (Chapter 7). From there, the book takes you through the journey of tools and frameworks applicable to distributed systems to achieve important

nonfunctional requirements: resilience, scalability, traceability, and deployment to the cloud among others (Chapter 8).

If you are already familiar with Spring Boot applications and how they work, you can go quickly through the first chapters and focus more on the second part of the book. There, we cover more advanced topics such as event-driven design, service discovery, routing, distributed tracing, testing with Cucumber, etc. However, pay attention to the foundations we set up in the first part: test-driven development, the focus on the minimum viable product (MVP), and monolith-first.

Skeleton with Spring Boot, the Professional Way

First, the book guides you through the creation of an application using Spring Boot. All the contents mainly focus on the backend side, but you will create a simple web page with React to demonstrate how to use the exposed functionality as a REST API.

It's important to point out that we don't create "shortcut code" just to showcase Spring Boot features: that's not the objective of this book. We use Spring Boot as a vehicle to teach concepts, but we could use any other framework, and the ideas of this book would still be valid.

You will learn how to design and implement the application following the well-known three-tier, three-layer pattern. You do this supported by an incremental example, with hands-on code. While writing the applications, we'll also pause a few times to get into the details about how Spring Boot works with so little code (autoconfiguration, starters, etc.).

Test-Driven Development

In the first chapters, we use test-driven development (TDD) to map the prerequisites presented to technical features. This book tries to show this technique in a way that you can see the benefits from the beginning: why it's always a good idea to think about the test cases before writing your code. JUnit 5, AssertJ, and Mockito will serve us to build useful tests efficiently.

The plan is the following: you'll learn how to create the tests first, then make them fail, and finally implement the logic to make them work.

Microservices

Once you have your first application ready, we introduce a second one that will interact with the existing functionality. From that moment on, you'll have a microservices architecture. It doesn't make any sense to try to understand the advantages of microservices if you have only one of them. The real-life scenarios are always distributed systems with functionality split into different services. As usual, to keep it practical, we'll analyze the specific situation for our case study, so you'll see if moving to microservices fits your needs.

The book covers not only the reasons to split the system but also the disadvantages that come with that choice. Once you make the decision to move to microservices, you'll learn which patterns you should use to build a good architecture for the distributed system: service discovery, routing, load balancing, distributed tracing, containerization, and some other supporting mechanisms.

Event-Driven System

An additional concept that does not always need to come with microservices is an *event-driven architecture*. This book uses it since it's a pattern that fits well into a microservice architecture, and you'll make your choice based on good examples. You'll see what the differences are between synchronous and asynchronous communication, as well as their main pros and cons.

This asynchronous way of thinking introduces new ways of designing code, with *eventual consistency* as one of the key changes to embrace. You'll look at it while coding your project, using RabbitMQ to send and receive messages between microservices.

Nonfunctional Requirements

When you build an application in the real world, you have to take into account some requirements that are not directly related to functionalities, but they prepare your system to be more robust, to keep working in the event of failures, or to ensure data integrity, as some examples.

Many of these *nonfunctional requirements* are related to things that can go wrong with your software: network failures that make part of your system unreachable, a high traffic volume that collapses your backend capacity, external services that don't respond, etc.

In this book, you'll learn how to implement and verify patterns to make the system more resilient and scalable. In addition, we'll discuss the importance of data integrity and the tools that help us guarantee it.

The good part about learning how to design and solve all these nonfunctional requirements is that it's knowledge applicable to any system, no matter the programming language and frameworks you're using.

Online Content

For this second edition of the book, I decided to create an online space where you can keep learning new topics related to microservice architectures. On this web page, you'll find new guides that extend the practical use case covering other important aspects of distributed systems. Additionally, new versions of the repositories using up-to-date dependencies will be published there.

The first guide that you'll find online is about testing a distributed system with Cucumber. This framework helps us build human-readable test scripts to make sure our functionalities work end to end.

Visit <https://tpd.io/book-extra> for all the extra content and new updates about the book.

Summary

This chapter introduced the main goals of this book: to teach you the main aspects of a microservice architecture, by starting simple and then growing your knowledge through the development of a sample project.

We also covered briefly the main content of the book: from monolith-first to microservices with Spring Boot, test-driven development, event-driven systems, common architecture patterns, nonfunctional requirements, and end-to-end testing with Cucumber (online).

The next chapter will start with the first step of our learning path: a review of some basic concepts.

CHAPTER 2

Basic Concepts

This book follows a practical approach, so most of the tools covered are introduced as we need them. However, we'll go over some core concepts separately because they're either the foundations of our evolving example or used extensively in the code examples, namely, Spring, Spring Boot, testing libraries, Lombok, and logging. These concepts deserve a separate introduction to avoid long interruptions in our learning path, which is why this chapter gives an overview of them.

Keep in mind that the next sections are not intended to give you a full knowledge base of these frameworks and libraries. The primary objective of this chapter is that either you refresh the concepts in your mind (if you already learned them) or you grasp the basics so that you don't need to consult external references before reading the rest of the chapters.

Spring

The Spring Framework is a vast set of libraries and tools simplifying software development, namely, dependency injection, data access, validation, internationalization, aspect-oriented programming, etc. It's a popular choice for Java projects, and it also works with other JVM-based languages such as Kotlin and Groovy.

One of the reasons why Spring is so popular is that it saves a lot of time by providing built-in implementations for many aspects of software development, such as the following:

- *Spring Data* simplifies data access for relational and NoSQL databases.
- *Spring Batch* provides powerful processing for large volumes of records.

- *Spring Security* is a security framework that abstracts security features to applications.
- *Spring Cloud* provides tools for developers to quickly build some of the common patterns in distributed systems.
- *Spring Integration* is an implementation of enterprise integration patterns. It facilitates integration with other enterprise applications using lightweight messaging and declarative adapters.

As you can see, Spring is divided into different modules. All of the modules are built on top of the core Spring Framework, which establishes a common programming and configuration model for software applications. This model itself is another important reason to choose the framework since it facilitates good programming techniques such as the use of interfaces instead of classes to decouple application layers via dependency injection.

A key topic in Spring is the Inversion of Control (IoC) container, which is supported by the `ApplicationContext` interface. Spring creates this “space” in your application where you, and the framework itself, can put some object instances such as database connection pools, HTTP clients, etc. These objects, called *beans*, can be later used in other parts of your application, commonly through their public interface to abstract your code from specific implementations. The mechanism to reference one of these beans from the application context in other classes is what we call *dependency injection*, and in Spring this is possible via XML configuration or code annotations.

Spring Boot

Spring Boot is a framework that leverages Spring to quickly create stand-alone applications in Java-based languages. It has become a popular tool for building microservices.

Having so many available modules in Spring and other related third-party libraries that can be combined with the framework is powerful for software development. Yet, despite a lot of efforts to make Spring configuration easier, you still need to spend some time to set up everything you need for your application. And, sometimes, you just require the same configuration over and over again. *Bootstrapping* an application, meaning the process to configure your Spring application to have it up and running, is sometimes tedious. The advantage of Spring Boot is that it eliminates most of that

process by providing default configurations and tools that are set up automatically for you. The main disadvantage is that if you rely too much on these defaults, you may lose control and awareness of what's happening. We'll unveil some of the Spring Boot implementations within the book to demonstrate how it works internally so that you can be in control at all times.

Spring Boot provides some predefined *starter packages* that are like collections of Spring modules and some third-party libraries and tools together. As an example, `spring-boot-starter-web` helps you build a stand-alone web application. It groups the Spring Core Web libraries with Jackson (JSON handling), validation, logging, autoconfiguration, and even an embedded Tomcat server, among other tools.

In addition to starters, *autoconfiguration* plays a key role in Spring Boot. This feature makes adding functionality to your application extremely easy. Following the same example, just by including the web starter, you will get an embedded Tomcat server. There's no need to configure anything. This is because the Spring Boot autoconfiguration classes scan your classpath, properties, components, etc., and load some extra beans and behavior based on that.

To be able to manage different configuration options for your Spring Boot application, the framework introduces *profiles*. You can use profiles, for example, to set different values for the host to connect to when using a database in a development environment and a production environment. Additionally, you can use a different profile for tests, where you may need to expose additional functions or mock parts of your application. We'll cover profiles more in detail in Chapter 8.

We'll use the Spring Boot Web and Data starters to quickly build a web application with persistent storage. The Test starter will help us write tests, given that it includes some useful test libraries such as JUnit and AssertJ. Then, we'll add messaging capabilities to our applications by adding the AMQP starter, which includes a message broker integration (RabbitMQ) that we'll use to implement an event-driven architecture. In Chapter 8, we'll include a different type of starters, grouped within the Spring Cloud family. We'll make use of some of these tools to implement common patterns for distributed systems: routing (Spring Cloud Gateway), service discovery (Consul), and load balancing (Spring Cloud Load Balancer), among others. Don't worry about all these new terms for now; they'll be explained in detail while we make progress on the practical example.

The next chapter covers in detail how these starters and Spring Boot autoconfiguration work, based on a practical example.

Lombok and Java

The code examples in this book use Project Lombok, a library that generates Java code based on annotations. The main reason to include Lombok in the book is educational: it keeps the code samples concise, reducing the boilerplate so the reader can focus on what it matters.

Let's use one of the first simple classes as an example. We want to create an immutable multiplication challenge class with two factors. See Listing 2-1.

Listing 2-1. The Challenge Class in Plain Java

```
public final class Challenge {

    // Both factors
    private final int factorA;
    private final int factorB;

    public Challenge(int factorA, int factorB) {
        this.factorA = factorA;
        this.factorB = factorB;
    }

    public int getFactorA() {
        return this.factorA;
    }

    public int getFactorB() {
        return this.factorB;
    }

    public boolean equals(final Object o) {
        if (o == this) return true;
        if (!(o instanceof Challenge)) return false;
        final Challenge other = (Challenge) o;
        if (this.getFactorA() != other.getFactorA()) return false;
        if (this.getFactorB() != other.getFactorB()) return false;
        return true;
    }
}
```

```

public int hashCode() {
    final int PRIME = 59;
    int result = 1;
    result = result * PRIME + this.getFactorA();
    result = result * PRIME + this.getFactorB();
    return result;
}

public String toString() {
    return "Challenge(factorA=" + this.getFactorA() + ", factorB=" + this.
        getFactorB() + ")";
}

}

```

As you can see, the full class has some classic boilerplate code: constructors, getters, and the equals, hashCode, and toString methods. They don't add much to this book, yet we need them for the code to work.

The same class can be reduced with Lombok to its minimum expression. See Listing 2-2.

Listing 2-2. The Challenge Class Using Lombok

```

import lombok.Value;

@Value
public class Challenge {

    // Both factors
    int factorA;
    int factorB;
}

```

The @Value annotation provided by Lombok groups some other annotations in this library that we could also use separately. Each of the following annotations instructs Lombok to generate code blocks before the Java build phase:

- @AllArgsConstructor creates a constructor with all the existing fields.
- @FieldDefaults makes our fields private and final.
- @Getter generates getters for factorA and factorB.

- `@ToString` includes a simple implementation concatenating fields.
- `@EqualsAndHashCode` generates basic `equals()` and `hashCode()` methods using all fields by default, but we could also customize it.

Not only does Lombok reduce our code to the minimum, but it also helps when you need to modify these classes. Adding a new field to the `Challenge` class in Lombok means adding one line (excluding usages of the class). If we would use the plain Java version, we would need to add the new argument to the constructor, add `equals` and `hashCode` methods, and add a new getter. Not only it means extra work, but it's also error-prone: in case we forget the extra field in the `equals` method, for example, we would introduce a bug in our application.

Like many tools, Lombok has also detractors. The main reason not to like Lombok is that since it's easy to add code to your classes, you might end up adding code that you don't really need (e.g., setters or extra constructors). Besides, you could argue that having a good IDE with code generation and a refactoring assistant can help more or less at the same level. Keep in mind that, to use Lombok properly, you need your IDE to provide support for it. This may happen either natively or, typically, via a plugin. For example, in IntelliJ, you have to download and install the Lombok plugin. All the developers in the project have to adapt their IDE to Lombok, so even though it's easy to do, you could see that as an extra inconvenience.

In the coming chapters, we'll use mainly these Lombok features:

- We annotate with `@Value` the immutable classes.
- For the data entities, we use separately some of the annotations described earlier.
- We add the `@Slf4j` annotation for Lombok to create a logger using the standard Simple Logging Facade for Java API (SLF4J). The section “Logging” in this chapter gives more background about these concepts.

In any case, we'll describe what these annotations do when we look at the code examples, so you don't need to dive into more details on how they work.

If you prefer plain Java code, just use the Lombok's code annotations in this book as a reference to know what extra code you need to include in your classes.

Java Records

Starting with JDK 14, the Java records feature has been made available in preview mode. If we would use this feature, we could write our Challenge class in pure Java in a concise way as well.

```
public record Challenge(int factorA, int factorB) {}
```

However, there isn't yet a complete integration of this feature with other libraries and frameworks at the time of writing this book. In addition, Lombok adds some extra features and better granularity of options when compared to the Java records. For these reasons, we won't use records in this book.

Testing Basics

In this section, we'll go through some important testing approaches and libraries. We'll put them into practice in the next chapters, so it's good to learn (or review) the basic concepts first.

Test-Driven Development

The first practical chapters in this book encourage you to use *test-driven development* (TDD). This technique helps you by putting the focus first on what you need and what your expectations are and later moving to the implementation. It makes that you, as a developer, think about what the code should do under certain situations or use cases. In real life, TDD also helps you clarify vague requirements and discard invalid ones.

Given that this book is driven by a practical case, you'll find that TDD fits quite well within the main purpose.

Behavior-Driven Development

As an addition to the idea of writing your tests before your logic, *behavior-driven development* (BDD) brings some better structure and readability to your tests.

In BDD, we write the tests according to a *Given-When-Then* structure. This removes the gap between developers and business analysts when mapping use cases to tests. Analysts can just read the code and identify what is being tested there.

Keep in mind that BDD, like TDD, is a development process by itself and not simply a way of writing tests. Its main goal is to facilitate conversations to improve the definition of requirements and their test cases. In this book, our focus regarding BDD will be on the test structure. See Listing 2-3 for an example of how these tests look.

Listing 2-3. An Example of a BDD Test Case Using a Given-When-Then Structure

```

@Test
public void getRandomMultiplicationTest() throws Exception {
    // given
    given(challengeGeneratorService.randomChallenge())
        .willReturn(new Challenge(70, 20));

    // when
    MockHttpServletResponse response = mvc.perform(
        get("/multiplications/random")
            .accept(MediaType.APPLICATION_JSON))
        .andReturn().getResponse();

    // then
    then(response.getStatus()).isEqualTo(HttpStatus.OK.value());
    then(response.getContentAsString())
        .isEqualTo(json.writeValueAsString(new Challenge(70, 20)));
}

```

JUnit

The code in this book uses JUnit 5 for the unit tests. The Spring Boot Test starter includes these libraries, so we won't need to include it in our dependencies.

In general, the idea behind unit tests is that you can verify the behavior of your classes (units) separately. In this book, we'll write unit tests for every class where we put logic.

Among all the features in JUnit 5, we will use mainly the basic ones, listed here:

- `@BeforeEach` and `@AfterEach` for code that should be executed before and after each test, respectively.
- `@Test` for every method that represents a test we want to execute.
- `@ExtendsWith` at the class level to add JUnit 5 extensions. We will use this to add the Mockito extension and the Spring extension to our tests.

Mockito

Mockito is a mocking framework for unit tests in Java. When you *mock* a class, you are overriding the real behavior of that class with some predefined instructions of what their methods should return or do for their arguments. This is an important requirement to write unit tests since you want to validate the behavior of one class only and simulate all its interactions.

The easiest way to mock a class with Mockito is to use the `@Mock` annotation in a field combined with `MockitoExtension` for JUnit 5. See Listing 2-4.

Listing 2-4. MockitoExtension and Mock Annotation Usage

```
@ExtendWith(MockitoExtension.class)
public class MultiplicationServiceImplTest {

    @Mock
    private ChallengeAttemptRepository attemptRepository;

    // [...] -> tests
}
```

Then, we could use the static method `Mockito.when` to define custom behavior. See Listing 2-5.

Listing 2-5. Defining Custom Behavior with Mockito's when

```
import static org.mockito.Mockito.when;
// ...
when(attemptRepository.methodThatReturnsSomething())
    .thenReturn(predefinedResponse);
```

However, we will use the alternative methods from `BDDMockito`, also included in the Mockito dependency. This gives us a more readable, BDD-style way of writing unit tests. See Listing 2-6.

Listing 2-6. Using given to Define Custom Behavior

```
import static org.mockito.BDDMockito.given;
// ...
given(attemptRepository.methodThatReturnsSomething())
    .willReturn(predefinedResponse);
```

In some cases, we will also need to check that an expected call to a mocked class was invoked. With Mockito, we use `verify()` for that. See Listing 2-7.

Listing 2-7. Verifying an Expected Call

```
import static org.mockito.Mockito.verify;
// ...
verify(attemptRepository).save(attempt);
```

As some extra background, it's good to know that there is also a BDD variation of `verify()`, called `then()`. Unfortunately, this replacement may be confusing when we combine BDDMockito with BDDAssertions from AssertJ (covered in the next section). Since we will use in this book assertions more extensively than verifications, we will opt for `verify` to better distinguish them.

Listing 2-8 shows a full example of a test using JUnit 5 and Mockito based on a class that we'll implement later in the book. For now, you can ignore the `then` assertion; we'll get there soon.

Listing 2-8. A Complete Unit Test with JUnit5 and Mockito

```
package microservices.book.multiplication.challenge;

import java.util.Optional;

import org.junit.jupiter.api.BeforeEach;
import org.junit.jupiter.api.Test;
import org.junit.jupiter.api.extension.ExtendWith;
import org.mockito.Mock;
import org.mockito.junit.jupiter.MockitoExtension;

import microservices.book.multiplication.event.ChallengeSolvedEvent;
import microservices.book.multiplication.event.EventDispatcher;
import microservices.book.multiplication.user.User;
import microservices.book.multiplication.user.UserRepository;

import static org.assertj.core.api.BDDAssertions.then;
import static org.mockito.BDDMockito.*;

@ExtendWith(MockitoExtension.class)
public class ChallengeServiceImplTest {
```

```
private ChallengeServiceImpl challengeServiceImpl;  
  
@Mock  
private ChallengeAttemptRepository attemptRepository;  
  
@Mock  
private UserRepository userRepository;  
  
@Mock  
private EventDispatcher eventDispatcher;  
  
@BeforeEach  
public void setUp() {  
    challengeServiceImpl = new ChallengeServiceImpl(attemptRepository,  
        userRepository, eventDispatcher);  
}  
  
@Test  
public void checkCorrectAttemptTest() {  
    // given  
    long userId = 9L, attemptId = 1L;  
    User user = new User("john_doe");  
    User savedUser = new User(userId, "john_doe");  
    ChallengeAttemptDTO attemptDTO =  
        new ChallengeAttemptDTO(50, 60, "john_doe", 3000);  
    ChallengeAttempt attempt =  
        new ChallengeAttempt(null, savedUser, 50, 60, 3000, true);  
    ChallengeAttempt storedAttempt =  
        new ChallengeAttempt(attemptId, savedUser, 50, 60, 3000, true);  
    ChallengeSolvedEvent event = new ChallengeSolvedEvent(attemptId, true,  
        attempt.getFactorA(), attempt.getFactorB(), userId,  
        attempt.getUser().getAlias());  
    // user does not exist, should be created  
    given(userRepository.findByAlias("john_doe"))  
        .willReturn(Optional.empty());  
    given(userRepository.save(user))  
        .willReturn(savedUser);  
    given(attemptRepository.save(attempt))  
        .willReturn(storedAttempt);
```

```
// when
ChallengeAttempt resultAttempt =
    challengeServiceImpl.checkAttempt(attemptDTO);

// then
then(resultAttempt.isCorrect()).isTrue();
verify(userRepository).save(user);
verify(attemptRepository).save(attempt);
verify(eventDispatcher).send(event);
}

}
```

AssertJ

The standard way to verify expected results with JUnit 5 is using assertions.

```
assertEquals("Hello, World!", actualGreeting);
```

There are not only assertions for equality of all kinds of objects but also to verify true/false, null, execution before a timeout, throwing an exception, etc. You can find them all in the Assertions Javadoc (<https://tpd.io/junit-assert-docs>).

Even though JUnit assertions are enough in most cases, they are not as easy to use and readable as the ones provided by AssertJ. This library implements a fluent way of writing assertions and provides extra functionality so you can write more concise tests.

In its standard form, the previous example looks like this:

```
assertThat(actualGreeting).isEqualTo("Hello, World!");
```

However, as we mentioned in previous sections, we want to make use of a BDD language approach. Therefore, we will use the `BDDAssertions` class included in AssertJ. This class contains method equivalencies for all the `assertThat` cases, renamed as `then`.

```
then(actualGreeting).isEqualTo("Hello, World!");
```

In the book, we will mostly some basic assertions from AssertJ. If you're interested in extending your knowledge about AssertJ, you can start with the official documentation page (<https://tpd.io/assertj>).

Testing in Spring Boot

Both JUnit 5 and AssertJ are included in `spring-boot-starter-test`, so we only need to include this dependency in our Spring Boot application to make use of them. Then, we can use different testing strategies.

One of the most popular ways to write tests in Spring Boot is to make use of the `@SpringBootTest` annotation. It will start a Spring context and make all your configured beans available for the test. If you're running integration tests and want to verify how different parts of your application work together, this approach is convenient.

When testing specific slices or individual classes of your application, it's better to use plain unit tests (without Spring at all) or more fine-grained annotations like `@WebMvcTest`, focused on controller-layer tests. This is the approach we'll use in the book, so we'll explain it more in detail when we get there.

For now, let's focus only on the integration between the libraries and frameworks we've been describing in this chapter.

- The Spring Test libraries (included via Spring Boot Test starter) come with a `SpringExtension` so you can integrate Spring in your JUnit 5 tests via the `@ExtendWith` annotation.
- The Spring Boot Test package introduces the `@MockBean` annotation that we can use to replace or add a bean in the Spring context, in a similar way to how Mockito's `@Mock` annotation can replace the behavior of a given class. This is helpful to test the application layers separately so you don't need to bring all your real class behaviors in your Spring context together. We'll see a practical example when testing our application controllers.

Logging

In Java, we can log messages to the console just by using the `System.out` and `System.err` print streams.

```
System.out.println("Hello, standard output stream!");
```

This is considered simply good enough for a 12-factor app (<https://tpd.io/12-logs>), a popular set of best practices for writing cloud-native applications. The reason is that, eventually, some other tool will collect them from the standard outputs at the system level and aggregate them in an external framework.

Therefore, we'll write our logs to the standard and the error output. But that doesn't mean that we have to stick to the plain, ugly `System.out` variants in Java.

Most professional Java apps use a logger implementation such as LogBack. And, given that there are multiple logging frameworks for Java, it's even better to choose a generic abstraction such as SLF4J.

The good news is that Spring Boot comes with all this logging configuration already set up for us. The default implementation is LogBack, and Spring Boot's preconfigured message format is as follows:

```
2020-03-22 10:19:59.556 INFO 93532 --- [main]
o.s.b.w.embedded.tomcat.TomcatWebServer : Tomcat started on
port(s): 8080 (http) with context path ''
```

SLF4J loggers are also supported. To use a logger, we create it via the `LoggerFactory`. The only argument it needs is a name. By default, it is common to use the factory method that takes the class itself and gets the logger name from it. See Listing 2-9.

Listing 2-9. Creating and Using a Logger with SLF4J

```
import org.slf4j.Logger;
import org.slf4j.LoggerFactory;

class ChallengeServiceImpl {

    private static final Logger log = LoggerFactory.getLogger(ChallengeServiceImpl.class);

    public void dummyMethod() {
        var name = "John";
        log.info("Hello, {}!", name);
    }
}
```

As you see in the example, loggers support parameter replacement via the curly-braces placeholder.

Given that we are using Lombok in this book, we can replace that line to create a logger in our class with a simple annotation: `@Slf4j`. This helps to keep our code concise. By default, Lombok creates a static variable named `log`. See Listing 2-10.

Listing 2-10. Using a Logger with Lombok

```
import lombok.extern.slf4j.Slf4j;

@Slf4j
class ChallengeServiceImpl {

    public void dummyMethod() {
        var name = "John";
        log.info("Hello, {}!", name);
    }
}
```

Summary and Achievements

In this chapter, we reviewed some basic libraries and concepts that we'll use in the book: Spring Boot, Lombok, tests with JUnit and AssertJ, and logging. These are only a few of what you'll learn during the journey, but they were introduced separately to avoid long pauses in the main learning path. All the other topics, more related to our evolving architecture, are explained in detail as we navigate through the book pages.

Do not worry if you still feel like you have some knowledge gaps. The practical code examples in the next chapters will help you understand these concepts by providing extra context.

Chapter's Achievements:

- You reviewed the core ideas about Spring and Spring Boot.
- You understood how we'll use Lombok in the book to reduce boilerplate code.
- You learned how to use tools like JUnit, Mockito, and AssertJ to implement test-driven development and how these tools are integrated in Spring Boot.
- You reviewed some logging basics and how to use a logger with Lombok.

CHAPTER 3

A Basic Spring Boot Application

We could start writing code directly, but that, even while being pragmatic, would be far from being a real case. Instead, we'll define a product that we want to build, and we'll split it into small chunks. This requirements-oriented approach is used throughout the book to make it more practical. In real life, you'll always have these business requirements.

We want a web application to encourage users to exercise their brains every day. To begin with, we will present users with two-digit multiplications, one every time they access the page. They will type their alias (a short name) and their guess for the result of the operation. The idea is that they should use only mental calculations. After they send the data, the web page will indicate the user if the guess was correct or not.

Also, we'd like to keep user motivation as high as possible. To achieve that, we will use some gamification. For each correct guess, we give points to the user, and they will see their score in a ranking, so they can compete with other people.

This is the main idea of the complete application we will build, our product vision. But we won't build it all at once. This book will emulate an agile way of working in which we split requirements into user stories, small chunks of functionality that give value by themselves. We'll follow this methodology to keep this book as close as possible to real life since a vast majority of IT companies use agile.

Let's start simple and focus first on the multiplication solving logic. Consider the first user story here.

USER STORY 1

As a user of the application, I want to solve a random multiplication problem using mental calculation so I exercise my brain.

To make this work, we need to build a minimal skeleton of our web application. Therefore, we'll split the user story into several subtasks.

1. Create a basic service with business logic.
2. Create a basic API to access this service (REST API).
3. Create a basic web page to ask the users to solve that calculation.

In this chapter, we'll focus on 1 and 2. After creating the skeleton of our first Spring Boot application, we'll use test-driven development to build the main logic of this component: generating multiplication challenges and verifying attempts from the users to solve those challenges. Then, we'll add the controller layer implementing the REST API. You'll learn what the advantages are of this layered design.

Our learning path includes some reasoning about one of the most important features in Spring Boot: autoconfiguration. We'll use our practical case to see how, for example, the application includes its own embedded web server, only because we added a specific dependency to our project.

Setting Up the Development Environment

We'll use Java 14 in this book. Make sure you get at least that version of the JDK from the official downloads page (<https://tpd.io/jdk14>). Install it following the instructions for your OS.

A good IDE is also convenient to develop Java code. If you have a preferred one, just use it. Otherwise, you can download for example the community version of IntelliJ IDEA or Eclipse.

In this book, we'll also use HTTPie to quickly test our web applications. It's a command-line tool that allows us to interact with an HTTP server. You can follow the instructions at <https://tpd.io/httpie-install> to download it for Linux, Mac, or Windows. Alternatively, if you are a curl user, you can also map this book's http commands to curl commands easily.

The Skeleton Web App

It's time to write some code! Spring offers a fantastic way to build the skeleton of an application: the Spring Initializr. This is a web page that allows us to select what components and libraries we want to include in our Spring Boot project, and it generates the structure and the dependency configuration into a zip file that we can download. We'll use the Initializr a few times in the book since it saves time over creating the project from scratch, but you can also create the project yourself if you prefer that option.

Source Code: chapter03

You can find all the source code for this chapter on GitHub, in the `chapter03` repository.

See <https://github.com/Book-Microservices-v2/chapter03>.

Let's navigate to <https://start.spring.io/> and fill in some data, as shown in Figure 3-1.

The screenshot shows the Spring Initializr web application interface. At the top, there's a logo and a search bar. Below the search bar, there are sections for Project type (Maven Project selected), Language (Java selected), and Spring Boot version (2.3.3 selected). The 'Project Metadata' section includes fields for Group (microservices.book), Artifact (multiplication), Name (multiplication), Description (Multiplication Application), Package name (microservices.book.multiplication), and Packaging (Jar selected). Under 'Dependencies', 'Lombok' and 'Developer Tools' are listed with a note about reducing boilerplate code. 'Spring Web' and 'WEB' are selected under 'Dependencies'. The 'Validation' section includes 'JSR-303 validation with Hibernate validator'. At the bottom, there are three buttons: 'GENERATE' (with a keyboard shortcut of ⌘ + ↵), 'EXPLORE' (with a keyboard shortcut of CTRL + SPACE), and 'SHARE...'.

Figure 3-1. Creating a Spring Boot project with the Spring Initializr

All the code in this book uses Maven, Java, and the Spring Boot version 2.3.3, so let's stick to them. If that Spring Boot version is not available, you can select a more recent one. In that case, remember to change it later in the generated `pom.xml` file if you want to use the same version as in the book. You may also go ahead with other Java and Spring Boot versions, but then some of the code examples in this book may not work for you. Check the online book resources (<https://tpd.io/book-extra>) for the latest news about compatibility and upgrades.

Give some values to the group (`microservices.book`) and the artifact (`multiplication`). Select Java 14. Do not forget to add the dependencies Spring Web, Validation, and Lombok from the list or search tool. You already know what Lombok is intended for, and you'll see what the other two dependencies do in this chapter. That's all we need for now.

Generate the project and extract the ZIP contents. The `multiplication` folder contains everything you need to run the application. You can now open it with your favorite IDE, usually by selecting the `pom.xml` file.

These are the main elements we'll find in the autogenerated package:

- There is Maven's `pom.xml` file with the application metadata, configuration, and dependencies. This is the main file used by Maven to build the application. We'll examine separately some of the dependencies added by Spring Boot. Inside this file, you can also find the configuration to build the application using Spring Boot's Maven plugin, which also knows how to package all its dependencies in a stand-alone `.jar` file and how to run these applications from the command line.
- There is the Maven wrapper. This is a stand-alone version of Maven so you don't need to install it to build your app. These are the `.mvn` folder and the `mvnw` executables for Windows and UNIX-based systems.
- We'll find a `HELP.md` file with some links to Spring Boot's documentation.
- Assuming we'll use Git as a version control system, the included `.gitignore` has some predefined exclusions, so we don't commit the compiled classes or any IDE-generated files to the repository.
- The `src` folder follows the standard Maven structure that divides the code into the subfolders `main` and `test`. Both folders may contain their respective `java` and `resources` children. In this case, there is a source folder for both our main code and tests, and a resource folder for the main code.

- There is a class created by default in our main sources, `MultiplicationApplication`. It's already annotated with `@SpringBootApplication` and contains the `main` method that boots up the application. This is the standard way of defining the main class for a Spring Boot application, as detailed in the reference documentation (<https://tpd.io/sb-annotation>). We'll take a look at this class later.
- Within the `resources` folder, we find two empty subfolders: `static` and `templates`. You can safely delete them since they're intended to include static resources and HTML templates, which we won't use.
- The `application.properties` file is where we can configure our Spring Boot application. We'll add later some configuration parameters here.

Now that we understand the different pieces of this skeleton, let's try to *make it walk*. To run this app, you can either use your IDE interface or use this command from the project's root folder:

```
multiplication $ ./mvnw spring-boot:run
```

Running Commands from the Terminal

In this book, we use the `$` character to represent the command prompt. Everything after that character is the command itself. Sometimes, it's important to highlight that you have to run the command within a given folder in the workspace. In that case, you'll find the folder name before the `$` character (e.g. `multiplication $`). Of course, the specific location of your workspace may be different.

Note also that some commands may vary depending on whether you're using a UNIX-based operating system (like Linux or Mac) or in case you use Windows. All the commands shown in this book use a UNIX-based system version.

When we run that command, we're using the Maven wrapper included in the project's main folder (`mvnw`) with the goal (what's next to the Maven executable) `spring-boot:run`. This goal is provided by Spring Boot's Maven plugin, included also in the `pom.xml` file generated by the Initializr web page. The Spring Boot application should start successfully. The last line in the logs should look like this:

```
INFO 4139 --- [main] m.b.m.MultiplicationApplication: Started  
MultiplicationApplication in 6.599 seconds (JVM running for 6.912)
```

Great! We have a first Spring Boot app running without writing a single line of code! However, there isn't much we can do with it yet. What's this application doing? We'll figure that out soon.

Spring Boot Autoconfiguration

In the logs of our skeleton app, you can also find this log line:

```
INFO 30593 --- [main] o.s.b.w.embedded.tomcat.TomcatWebServer: Tomcat initialized  
with port(s): 8080 (http)
```

What we get when we add the web dependency is an independently deployable web application that uses Tomcat, thanks to the autoconfiguration feature in Spring.

As we introduced in the previous chapter, Spring Boot sets up libraries and default configuration automatically. This saves us a lot of time when we rely on all these defaults. One of those conventions is to add a ready-to-use Tomcat server when we add the Web starter to our project.

To learn more about Spring Boot autoconfiguration, let's go through how this specific case works, step-by-step. Use also Figure 3-2 for some useful visual help.

The Spring Boot application we generated automatically has a main class annotated with `@SpringBootApplication`. This is a *shortcut* annotation because it groups several others, among them `@EnableAutoConfiguration`. As its name suggests, with this one we're enabling the autoconfiguration feature. Therefore, Spring activates this smart mechanism and finds and processes classes annotated with the `@Configuration` annotation, from your own code but also from your dependencies.

CHAPTER 3 A BASIC SPRING BOOT APPLICATION

Our project includes the dependency `spring-boot-starter-web`. This is one of the main Spring Boot components, which has tooling to build a web application. Inside this artifact's dependencies, the Spring Boot developers added another starter, `spring-boot-starter-tomcat`. See Listing 3-1 or the online sources (<https://tpd.io/starter-web-deps>).

Listing 3-1. Web Starter Dependencies

```
plugins {  
    id "org.springframework.boot.starter"  
}  
  
description = "Starter for building web, including RESTful, applications using  
Spring MVC. Uses Tomcat as the default embedded container"  
  
dependencies {  
    api(project(":spring-boot-project:spring-boot-starters:spring-boot-starter"))  
    api(project(":spring-boot-project:spring-boot-starters:spring-boot-  
    starter-json"))  
    api(project(":spring-boot-project:spring-boot-starters:spring-boot-  
    starter-tomcat"))  
    api("org.springframework:spring-web")  
    api("org.springframework:spring-webmvc")  
}
```

As you can see, Spring Boot artifacts use Gradle (since version 2.3), but you don't need to know the specific syntax to understand what the dependencies are. If we now check the dependencies of the `spring-boot-starter-tomcat` artifact (in Listing 3-2 or the online sources, at <https://tpd.io/tomcat-starter-deps>), we see that it contains a library that doesn't belong to the Spring family, `tomcat-embed-core`. This is an Apache library that we can use to start a Tomcat embedded server. Its main logic is included in a class named `Tomcat`.

Listing 3-2. Tomcat Starter Dependencies

```
plugins {  
    id "org.springframework.boot.starter"  
}  
  
description = "Starter for using Tomcat as the embedded servlet container. Default  
servlet container starter used by spring-boot-starter-web"
```

```

dependencies {
    api("jakarta.annotation:jakarta.annotation-api")
    api("org.apache.tomcat.embed:tomcat-embed-core") {
        exclude group: "org.apache.tomcat", module: "tomcat-annotations-api"
    }
    api("org.glassfish:jakarta.el")
    api("org.apache.tomcat.embed:tomcat-embed-websocket") {
        exclude group: "org.apache.tomcat", module: "tomcat-annotations-api"
    }
}

```

Coming back to the hierarchy of dependencies, the `spring-boot-starter-web` also depends on `spring-boot-starter` (see Listing 3-1 and Figure 3-2 for some contextual help). That's the *core* Spring Boot starter, which includes the artifact `spring-boot-autoconfigure` (see Listing 3-3 or the online sources, at <https://tpd.io/sb-starter>). That Spring Boot artifact has a whole set of classes annotated with `@Configuration`, which are responsible for a big part of the whole Spring Boot magic. There are classes intended to configure web servers, message brokers, error handlers, databases, and many more. Check the complete list of packages at <https://tpd.io/auto-conf-packages> to get a better idea of the supported tools.

Listing 3-3. Spring Boot's Main Starter

```

plugins {
    id "org.springframework.boot.starter"
}

description = "Core starter, including auto-configuration support, logging
and YAML"

dependencies {
    api(project(":spring-boot-project:spring-boot"))
    api(project(":spring-boot-project:spring-boot-autoconfigure"))
    api(project(":spring-boot-project:spring-boot-starters:spring-boot-starter-
logging"))
    api("jakarta.annotation:jakarta.annotation-api")
    api("org.springframework:spring-core")
    api("org.yaml:snakeyaml")
}

```

For us, the relevant class that takes care of the embedded Tomcat server autoconfiguration is `ServletWebServerFactoryConfiguration`. See Listing 3-4 showing its most relevant code fragment, or see the complete source code available online (<https://tpd.io/swsfc-source>).

Listing 3-4. `ServletWebServerFactoryConfiguration` Fragment

```
@Configuration(proxyBeanMethods = false)
class ServletWebServerFactoryConfiguration {

    @Configuration(proxyBeanMethods = false)
    @ConditionalOnClass({ Servlet.class, Tomcat.class, UpgradeProtocol.class })
    @ConditionalOnMissingBean(value = ServletWebServerFactory.class, search =
        SearchStrategy.CURRENT)
    static class EmbeddedTomcat {

        @Bean
        TomcatServletWebServerFactory tomcatServletWebServerFactory(
            ObjectProvider<TomcatConnectorCustomizer> connectorCustomizers,
            ObjectProvider<TomcatContextCustomizer> contextCustomizers,
            ObjectProvider<TomcatProtocolHandlerCustomizer<?>>
                protocolHandlerCustomizers) {
            TomcatServletWebServerFactory factory = new
            TomcatServletWebServerFactory();
            factory.getTomcatConnectorCustomizers()
                .addAll(connectorCustomizers.orderedStream().
                    collect(Collectors.toList()));
            factory.getTomcatContextCustomizers()
                .addAll(contextCustomizers.orderedStream().collect(Collectors.
                    toList()));
            factory.getTomcatProtocolHandlerCustomizers()
                .addAll(protocolHandlerCustomizers.orderedStream().
                    collect(Collectors.toList()));
        return factory;
    }

}
// ...
}
```

This class defines some inner classes, one of them being `EmbeddedTomcat`. As you can see, that one is annotated with this:

```
@ConditionalOnClass({ Servlet.class, Tomcat.class, UpgradeProtocol.class })
```

Spring processes the `@ConditionalOnClass` annotation, which is used to load beans in the context if the linked class can be found in the classpath. In this case, the condition matches since we already saw how the `Tomcat` class got into our classpath via the starter hierarchy. Therefore, Spring loads the bean declared in `EmbeddedTomcat`, which turns out to be a `TomcatServletWebServerFactory`.

That factory is contained inside Spring Boot's core artifact (`spring-boot`, a dependency included in `spring-boot-starter`). It sets up a Tomcat embedded server with some default configuration. This is where the logic to create an embedded web server finally lives.

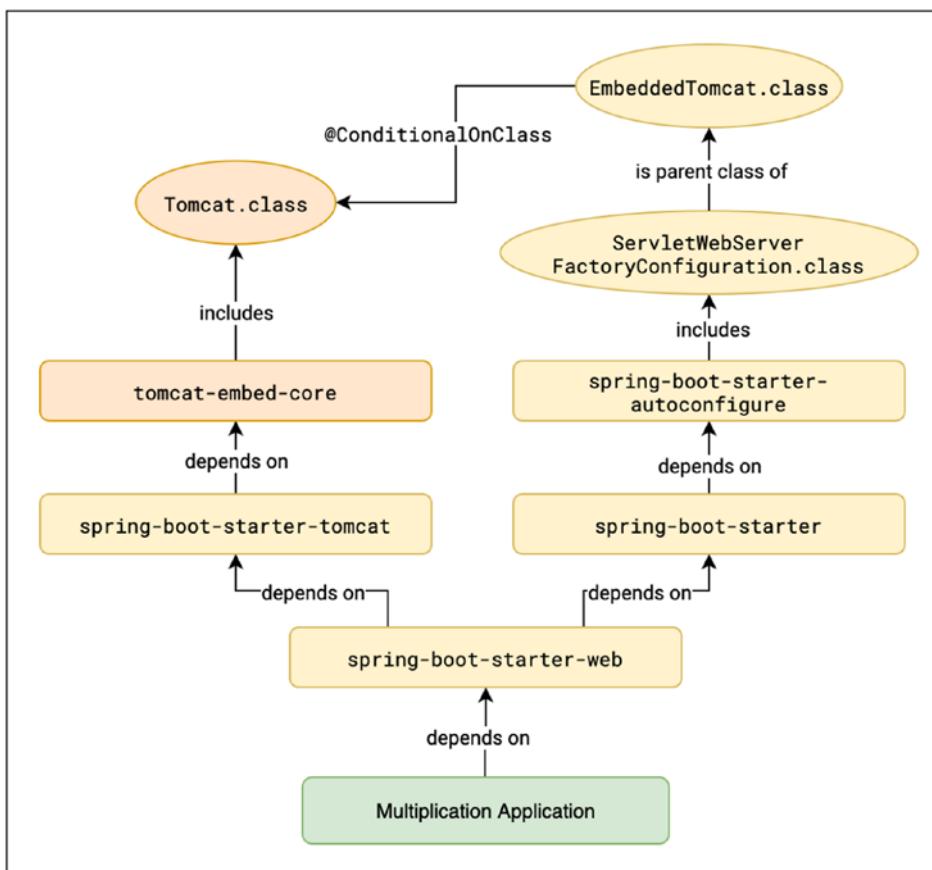


Figure 3-2. Autoconfiguration example: embedded Tomcat

Once more, just to recap, Spring scans all our classes and, given that the condition stated in the `EmbeddedTomcat` is fulfilled (the Tomcat library *is* an included dependency), it loads a `TomcatServletWebServerFactory` bean in the context. This Spring Boot class starts an embedded Tomcat server with the default configuration, exposing an HTTP interface on the port 8080.

As you can imagine, this same mechanism applies to many other libraries for databases, web servers, message brokers, cloud-native patterns, security, etc. In Spring Boot, you can find multiple starters that you can add as dependencies. When you do that, the autoconfiguration mechanism comes into play, and you get additional behavior out of the box. Many configuration classes are conditional on the presence of other classes like the ones we analyzed, but there are other condition types, for example, parameter values in the `application.properties` file.

Autoconfiguration is a key concept in Spring Boot. Once you understand it, the features that many developers consider magic are no longer a secret for you. We navigated through the details because it's important that you know this mechanism so you can configure it according to your needs and avoid getting a lot of behavior that you don't want or simply don't need. As a good practice, read carefully the documentation of the Spring Boot modules you're using, and familiarize yourself with the configuration options they allow.

Don't worry if you didn't understand this concept fully; we'll come back to the autoconfiguration mechanism a few times in this book. The reason is that we'll add extra features to our application, and, for that, we'll need to add extra dependencies to our project and analyze the new behavior they introduce.

Three-Tier, Three-Layer Architecture

The next step in our practical journey is designing how to structure our application and model our business logic in different classes.

A multitier architecture will provide our application with a more production-ready look. Most of the real-world applications follow this architecture pattern. Among web applications, the *three-tier design* is the most popular one and widely extended. These three tiers are as follows:

- *Client tier*: This tier is responsible for the user interface. Typically, this is what we call the *front end*.

- *Application tier:* This contains all the business logic together with the interfaces to interact with it and the data interfaces for persistence. This maps to what we call the *back end*.
- *Data store tier:* It's the database, file system, etc., that persists the application's data.

In this book, we're mainly focused on the application tier, although we'll use the other two as well. If now we zoom in, that application tier is commonly designed using three layers.

- *Business layer:* This includes the classes that model our domain and the business specifics. It's where the intelligence of the application resides. Sometimes this layer is divided into two parts: domains (entities) and applications (services) providing business logic.
- *Presentation layer:* In our case, it will be represented by the Controller classes, which will provide the functionality to the web client. Our REST API implementation will reside here.
- *Data layer:* This layer will be responsible for persisting our entities in a data storage, usually a database. It can typically include *data access object* (DAO) classes, which work with objects that map directly to rows in a database, or *repository* classes, which are domain-centric, so they may need to translate from domain representations to the database structure.

Our goal is now to apply this pattern to the Multiplication web application, as shown in Figure 3-3.

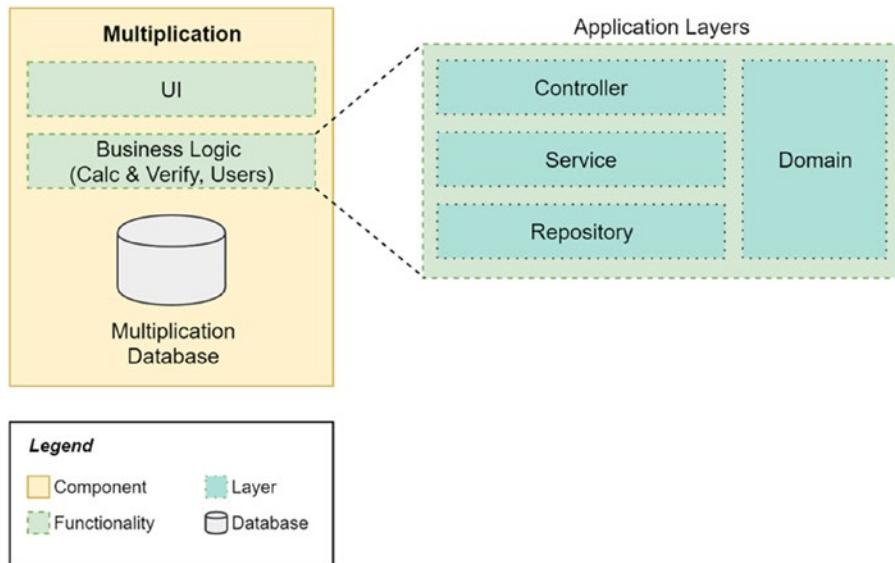


Figure 3-3. Three-tier, three-layer architecture applied to our Spring Boot project

The advantages of using this software architecture are all related to achieving loose coupling.

- All layers are interchangeable (such as, for instance, changing the database for a file storage solution or changing from a REST API to any other interface). This is a key asset because it makes easier to evolve the codebase. Additionally, you can replace complete layers by test mocks, which keeps your tests simple, as we'll see later in this chapter.
- The domain part is isolated and independent of everything else. It's not mixed with interface or database specifics.
- There is a clear separation of responsibilities: a class to handle database storage of the objects, a separate class for the REST API implementation, and another class for the business logic.

Spring is an excellent option to build this type of architecture, with many out-of-the-box features that will help us easily create a production-ready three-tier application. It provides three *stereotype* annotations for our classes that map to each of this design's layers, so we can use them to implement our architecture.

- The `@Controller` annotation is for the presentation layer. In our case, we'll implement a REST interface using controllers.

- The `@Service` annotation is for classes implementing business logic.
- The `@Repository` annotation is for the data layer, namely, the classes that interact with the database.

When we annotate classes with these variants, they become Spring-managed *components*. When initializing the web context, Spring scans your packages, finds these classes, and loads them as beans in the context. Then, we can use dependency injection to wire (or *inject*) these beans and, for example, use services from our presentation layer (controllers). We'll see this in practice soon.

Modeling Our Domain

Let's start by modeling our business domain, because this will help us structure our project.

Domain Definition and Domain-Driven Design

Our first web application takes care of generating multiplication challenges and verifying the subsequent attempts from the user. Let's define these three business entities.

- *Challenge*: Contains the two factors of a multiplication challenge
- *User*: Identifies the person who will try to solve a Challenge
- *Challenge Attempt*: Represents the attempt from a User to solve the operation from a Challenge

We can model these domain objects and their relationship as shown in Figure 3-4.

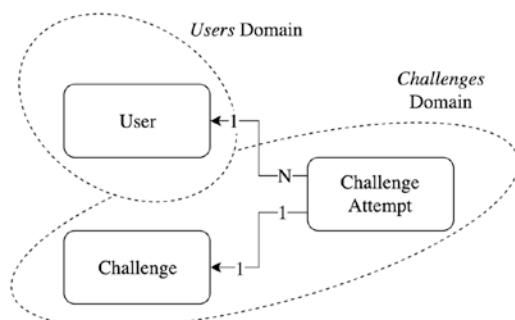


Figure 3-4. Business model

The relations between these objects are as follows:

- Users and Challenges are independent entities. They don't keep any references.
- *Challenge Attempts* are always for a given user and a given Challenge. Conceptually, there could be many attempts for the same Challenge, given that there is a limited number of generated challenges. Also, the same user may create many attempts since they can use the web application as many times as they want.

In Figure 3-4, you can also see how we split these three objects into two different domains: Users and Challenges. Finding domain boundaries (also known as bounded contexts; see <https://tpd.io/bounded-ctx>) and defining relations between your objects are essential tasks of designing software. This design approach based on domains is called *domain-driven design* (DDD). It helps you build an architecture that is modular, scalable, and loosely coupled. In our example, Users and Challenges are completely different concepts. Challenges, and their attempts, are related to users, but they together have enough relevance to belong to their own domain.

To make DDD clearer, we could think of an evolved version of this small system where other domains relate to Users or Challenges. For instance, we could introduce social network features by creating the domain *Friends* and modeling relationships and interactions between users. If we had mixed up the domains Users and Challenges, this evolution would be much harder to accomplish since the new domain has nothing to do with challenges.

For extra reading about DDD, you can get Eric Evans' book (<https://tpd.io/ddd-book>) or download the free InfoQ minibook (<https://tpd.io/ddd-quickly>).

Microservices and Domain-Driven Design

A common mistake when designing microservices is thinking that each domain has to be immediately split into a different microservice. This, however, may lead to premature optimization and an exponential complexity increase from the beginning of a software project.

We'll dive into more details about microservices and the monolith-first approach. For now, the important takeaway is that modeling domains is a crucial task, but splitting domains doesn't require splitting the code into microservices. In our first application, we'll include both domains together, but not mixed up. We'll use a simple strategy for the split: root-level packages.

Domain Classes

It's time to create the classes `Challenge`, `ChallengeAttempt`, and `User`. First, we divide our root package (`microservices.book.multiplication`) in two: `users` and `challenges`, following the domains that we identified for our `Multiplication` application. Then, we create three empty classes with the chosen names in these two packages. See Listing 3-5.

Listing 3-5. Splitting Domains by Creating Different Root Packages

```
+-- microservices.book.multiplication.user
|   \- User.java
+- microservices.book.multiplication.challenge
|   \- Challenge.java
|   \- ChallengeAttempt.java
```

Since we added Lombok as a dependency when we created the skeleton app, we can use it to keep our domain classes very small, as described in the previous chapter. Remember that you may need to add a plugin to your IDE to get full integration with Lombok; otherwise, you may get errors from the linter. As an example, in IntelliJ, you can install the official Lombok plugin by selecting Preferences ➤ Plugins and searching for *Lombok*.

The `Challenge` class holds both factors of the multiplication. We add getters, a constructor with all fields, and the `toString()`, `equals()`, and `hashCode()` methods. See Listing 3-6.

Listing 3-6. The Challenge Class

```
package microservices.book.multiplication.challenge;

import lombok.*;

/**
 * This class represents a Challenge to solve a Multiplication (a * b).
 */
@Getter
@ToString
@EqualsAndHashCode
@AllArgsConstructor
public class Challenge {
    private int factorA;
    private int factorB;
}
```

The User class has the same Lombok annotations, an identifier for the user, and a friendly alias (e.g., the user's first name). See Listing 3-7.

Listing 3-7. The User Class

```
package microservices.book.multiplication.user;

import lombok.*;

/**
 * Stores information to identify the user.
 */
@Getter
@ToString
@EqualsAndHashCode
@AllArgsConstructor
public class User {
    private Long id;
    private String alias;
}
```

Attempts also have an id, the value input by the user (resultAttempt), and whether it's correct or not. See Listing 3-8. We link it to the user via userId. Note that we also have here both challenge factors. We do this to avoid having a reference to a challenge via challengeId because we can simply generate new challenges "on the fly" and copy them here to keep our data structures simple. Therefore, as you can see, we have multiple options to implement the business model we depicted in Figure 3-4. To model the relationship with users, we use a reference; to model challenges, we embed the data inside the attempt. We'll analyze this decision in more detail in Chapter 5 when we cover data persistence.

Listing 3-8. The ChallengeAttempt Class

```
package microservices.book.multiplication.challenge;

import lombok.*;
import microservices.book.multiplication.user.User;

/**
 * Identifies the attempt from a {@link User} to solve a challenge.
 */
@Getter
@ToString
@EqualsAndHashCode
@AllArgsConstructor
public class ChallengeAttempt {

    private Long id;
    private Long userId;
    private int factorA;
    private int factorB;
    private int resultAttempt;
    private boolean correct;
}
```

Business Logic

Once we have the domain model defined, it's time to think about the other part of the business logic: the *application services*.

What We Need

Having looked at our requirements, we need the following:

- A way of generating a mid-complexity multiplication problem. Let's make all factors between 11 and 99.
- Some functionality to check whether an attempt is correct or not.

Random Challenges

Let's put test-driven development into practice for our business logic. First, we write a basic interface to generate random challenges. See Listing 3-9.

Listing 3-9. The ChallengeGeneratorService Interface

```
package microservices.book.multiplication.challenge;

public interface ChallengeGeneratorService {

    /**
     * @return a randomly-generated challenge with factors between 11 and 99
     */
    Challenge randomChallenge();

}
```

We place this interface also in the challenge package. Now, we write an empty implementation of this interface that wraps a Java's Random. See Listing 3-10. Besides the no-args constructor, we make our class testable by having a second constructor that accepts the random object.

Listing 3-10. An Empty Implementation of the ChallengeGeneratorService Interface

```
package microservices.book.multiplication.challenge;

import org.springframework.stereotype.Service;

import java.util.Random;
```

```

@Service
public class ChallengeGeneratorServiceImpl implements ChallengeGeneratorService {

    private final Random random;

    ChallengeGeneratorServiceImpl() {
        this.random = new Random();
    }

    protected ChallengeGeneratorServiceImpl(final Random random) {
        this.random = random;
    }

    @Override
    public Challenge randomChallenge() {
        return null;
    }
}

```

To instruct Spring to load this service implementation in the context, we annotate the class with `@Service`. We can later inject this service into other layers by using the interface and not the implementation. This way, we keep loose coupling since we could swap the implementation without needing to change anything in other layers. We'll put dependency injection into practice soon. For now, let's focus on TDD and leave empty the `randomChallenge()` implementation.

The next step is to write a test for this. We create a class in the same package but this time inside the test source folder. See Listing 3-11.

Listing 3-11. Creating the Unit Test Before the Real Implementation

```

package microservices.book.multiplication.challenge;

import org.junit.jupiter.api.BeforeEach;
import org.junit.jupiter.api.Test;
import org.junit.jupiter.api.extension.ExtendWith;
import org.mockito.Spy;
import org.mockito.junit.jupiter.MockitoExtension;

import java.util.Random;

import static org.assertj.core.api.BDDAssertions.then;
import static org.mockito.BDDMockito.given;

```

CHAPTER 3 A BASIC SPRING BOOT APPLICATION

```
@ExtendWith(MockitoExtension.class)
public class ChallengeGeneratorServiceTest {

    private ChallengeGeneratorService challengeGeneratorService;

    @Spy
    private Random random;

    @BeforeEach
    public void setUp() {
        challengeGeneratorService = new ChallengeGeneratorServiceImpl(random);
    }

    @Test
    public void generateRandomFactorIsBetweenExpectedLimits() {
        // 89 is max - min range
        given(random.nextInt(89)).willReturn(20, 30);

        // when we generate a challenge
        Challenge challenge = challengeGeneratorService.randomChallenge();

        // then the challenge contains factors as expected
        then(challenge).isEqualTo(new Challenge(31, 41));
    }
}
```

In the previous chapter, we reviewed how we can use Mockito to replace the behavior of a given class with the `@Mock` annotation and the `MockitoExtension` class for JUnit 5. In this test, we need to replace the behavior of an object, not a class. We use `@Spy` to stub an object. The Mockito extension will help to create a `Random` instance using the empty constructor and stubbing it for us to override the behavior. This is the simplest way to get our test to work since the basic Java classes implementing random generators do not work on interfaces (which we could then simply *mock* instead of *spy*).

Normally, we initialize what we need for all our tests in a method annotated with `@BeforeEach` so this happens before each test starts. Here we construct the service implementation passing this stub object.

The only test method sets up the preconditions with `given()` following a BDD style. The way to generate random numbers between 11 and 99 is to get a random number between 0 and 89 and add 11 to it. Therefore, we know that `random` should be called with 89 to generate numbers in the range 11, 100, so we override that call to return 20 when

it's called first time and 30 for the second time. Then, when we call `randomChallenge()`, we expect it to get 20 and 30 as random numbers from `random` (our stubbed object) and therefore return a `Challenge` object with 31 and 41.

So, we made a test that obviously fails when you run it. Let's try it; you can use your IDE or a Maven command from the project's root folder.

```
multiplication$ ./mvnw test
```

As expected, the test will fail. See the result in Listing 3-12.

Listing 3-12. Error Output After Running the Test for the First Time

Expecting:

```
<null>
to be equal to:
<Challenge(factorA=20, factorB=30)>
but was not.
Expected :Challenge(factorA=20, factorB=30)
Actual   :null
```

Now, we only need to make the test pass. In our case, the solution is quite simple, and we needed to figure it out while implementing the test. Later, we'll see more valuable cases of TDD, but this one already helps to get started with this way of working. See Listing 3-13.

Listing 3-13. Implementing a Valid Logic to Generate Challenges

```
@Service
public class ChallengeGeneratorServiceImpl implements ChallengeGeneratorService {

    private final static int MINIMUM_FACTOR = 11;
    private final static int MAXIMUM_FACTOR = 100;

    // ...

    private int next() {
        return random.nextInt(MAXIMUM_FACTOR - MINIMUM_FACTOR) + MINIMUM_FACTOR;
    }
}
```

```

@Override
public Challenge randomChallenge() {
    return new Challenge(next(), next());
}
}

```

Now, we run the test again, and it passes this time:

```
[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0
```

Test-driven development is just this simple. First, you design the tests, which will fail at the beginning. Then, you implement your logic to make them pass. In real life, you get the most of it when you get help to build the test cases from the people who define the requirements. You can write better tests and, therefore, a better implementation of what you really want to build.

Attempt Verification

To cover the second part of our business requirements, we implement an interface to verify attempts from users. See Listing 3-14.

Listing 3-14. The ChallengeService Interface

```

package microservices.book.multiplication.challenge;

public interface ChallengeService {

    /**
     * Verifies if an attempt coming from the presentation layer is correct or not.
     *
     * @return the resulting ChallengeAttempt object
     */
    ChallengeAttempt verifyAttempt(ChallengeAttemptDTO resultAttempt);

}

```

As you see in the code, we're passing a `ChallengeAttemptDTO` object to the `verifyAttempt` method. This class doesn't exist yet. *Data transfer objects* (DTOs) carry data between different parts of the system. In this case, we use a DTO to model the data

needed from the presentation layer to create an attempt. See Listing 3-15. An attempt from the user doesn't have the field `correct` and does not need to know about the user's ID. We can also use DTOs to validate data, as we'll see when we build the controllers.

Listing 3-15. The ChallengeAttemptDTO Class

```
package microservices.book.multiplication.challenge;

import lombok.Value;

/**
 * Attempt coming from the user
 */
@Value
public class ChallengeAttemptDTO {

    int factorA, factorB;
    String userAlias;
    int guess;

}
```

This time we use Lombok's `@Value`, a shortcut annotation to create an immutable class with an all-args-constructor and `toString`, `equals`, and `hashCode` methods. It'll also set our fields to be `private final`; that's why we didn't need to add that. Continuing with a TDD approach, we create do-nothing logic in the `ChallengeServiceImpl` interface implementation. See Listing 3-16.

Listing 3-16. An Empty ChallengeService Interface Implementation

```
package microservices.book.multiplication.challenge;

import org.springframework.stereotype.Service;

@Service
public class ChallengeServiceImpl implements ChallengeService {

    @Override
    public ChallengeAttempt verifyAttempt(ChallengeAttemptDTO attemptDTO) {
        return null;
    }
}
```

And now, we write a unit test for this class, so we verify that it works for both correct and wrong attempts. See Listing 3-17.

Listing 3-17. Writing the Test to Verify Challenge Attempts

```
package microservices.book.multiplication.challenge;

import org.junit.jupiter.api.BeforeEach;
import org.junit.jupiter.api.Test;

import static org.assertj.core.api.BDDAssertions.then;

public class ChallengeServiceTest {

    private ChallengeService challengeService;

    @BeforeEach
    public void setUp() {
        challengeService = new ChallengeServiceImpl();
    }

    @Test
    public void checkCorrectAttemptTest() {
        // given
        ChallengeAttemptDTO attemptDTO =
            new ChallengeAttemptDTO(50, 60, "john_doe", 3000);

        // when
        ChallengeAttempt resultAttempt =
            challengeService.verifyAttempt(attemptDTO);

        // then
        then(resultAttempt.isCorrect()).isTrue();
    }

    @Test
    public void checkWrongAttemptTest() {
        // given
        ChallengeAttemptDTO attemptDTO =
            new ChallengeAttemptDTO(50, 60, "john_doe", 5000);
```

```

// when
ChallengeAttempt resultAttempt =
    challengeService.verifyAttempt(attemptDTO);

// then
then(resultAttempt.isCorrect()).isFalse();
}
}

```

The result of multiplying 50 and 60 is 3,000, so the first test case's assertion expects the correct field to be true, whereas the second test expects false for a wrong guess (5,000).

Let's execute now the tests. You can use your IDE, or you can use a Maven command where you specify the name of the test to run.

```
multiplication$ ./mvnw -Dtest=ChallengeServiceTest test
```

You'll see an output similar to this:

```

[INFO] Results:
[INFO]
[ERROR] Errors:
[ERROR]   ChallengeServiceTest.checkCorrectAttemptTest:28 NullPointer
[ERROR]   ChallengeServiceTest.checkWrongAttemptTest:42 NullPointer
[INFO]
[ERROR] Tests run: 2, Failures: 0, Errors: 2, Skipped: 0

```

As foreseen, both tests will throw a null pointer exception for now.

Then, we go back to the service implementation, and we make it work. See Listing 3-18.

Listing 3-18. Implementing the Logic to Verify Attempts

```

@Override
public ChallengeAttempt verifyAttempt(ChallengeAttemptDTO attemptDTO) {
    // Check if the attempt is correct
    boolean isCorrect = attemptDTO.getGuess() ==
        attemptDTO.getFactorA() * attemptDTO.getFactorB();

    // We don't use identifiers for now
    User user = new User(null, attemptDTO.getUserAlias());

```

```
// Builds the domain object. Null id for now.  
ChallengeAttempt checkedAttempt = new ChallengeAttempt(null,  
    user,  
    attemptDTO.getFactorA(),  
    attemptDTO.getFactorB(),  
    attemptDTO.getGuess(),  
    isCorrect  
);  
  
return checkedAttempt;  
}
```

We keep it simple for now. Later, this implementation should take care of more tasks. We will need to create a user or find an existing one, connect that user to the new attempt, and store it in a database.

Now, run the test again to verify that it's passing:

```
[INFO] Tests run: 2, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.083 s -  
in microservices.book.multiplication.challenge.ChallengeServiceTest
```

Again, we used TDD successfully to build the logic to verify the challenge attempts.

The Users domain doesn't need any business logic within the scope of the first user story, so let's move to the next layer.

Presentation Layer

This section covers the presentation layer.

REST

Instead of building HTML from the server, we decided to approach the presentation layer as it's normally done in real software projects: with an API layer in between. By doing so, not only we can expose our functionality to other back-end services, but also we keep the back end and the front end completely isolated. This way, we could start, for example, with a simple HTML page and plain JavaScript and later move to a full front-end framework without changing the back-end code.

Among all the possible API alternatives, the most popular now is REpresentational State Transfer (REST). It's normally built on top of HTTP, so it uses HTTP verbs to perform the API operations: GET, POST, PUT, DELETE, etc. We'll build RESTful web services in this book, which are simply web services that conform to the REST architectural style. Therefore, we follow some conventions for URLs and HTTP verbs that have become the *de facto* standard. See Table 3-1.

Table 3-1. Conventions for Our REST APIs

HTTP Verb	Operation on Collection, e.g., /challenges	Operation on Item, e.g., challenges/3
GET	Gets the full list of items	Gets the item
POST	Creates a new item	Not applicable
PUT	Not applicable	Updates the item
DELETE	Deletes the full collection	Deletes the item

There are a few different styles for writing REST APIs. Table 3-1 shows the most basic operations with some convention choices made for this book. There are also multiple aspects of the contents transferred via the API: pagination, null handling, format (e.g., JSON), security, versioning, etc. If you are curious about how detailed these conventions may become for a real organization, you can take a look at Zalando's API Guidelines (<https://tpd.io/api-zalando>).

REST APIs with Spring Boot

Building a REST API with Spring is a simple task. There is a specialization of the `@Controller` stereotype that is intended for building REST controllers called, unsurprisingly, `@RestController`.

To model resources and mappings for different HTTP verbs, we use the `@RequestMapping` annotation. It's applicable to the class level and method level, so we can build our API contexts in a simple manner. To make it even simpler, Spring provides variants like `@PostMapping`, `@GetMapping`, etc., so we don't even need to specify the HTTP verb.

CHAPTER 3 A BASIC SPRING BOOT APPLICATION

Whenever we want to pass the body of a request to our method, we use the `@RequestBody` annotation. If we use a custom class, Spring Boot will try to deserialize it, using the type passed to the method. Spring Boot uses a JSON serialization format by default, although it also supports other formats when specified via the `Accept` HTTP header. In our web applications, we'll use all the Spring Boot defaults.

We can also customize our API with request parameters and read values from the request path. Let's take this request as an example:

```
GET http://ourhost.com/challenges/5?factorA=40
```

These are its different parts:

- GET is the HTTP verb.
- `http://ourhost.com/` is the host where the web server is running. In this example, the application is serving from the *root context*, `/`.
- `/challenges/` is an API context created by the application, to provide functionalities around this domain.
- `/5` is called a *path variable*. In this case, it represents the `Challenge` object with identifier 5.
- `factorA=40` is a request parameter and its value.

To process this request, we could create a controller that would get 5 as a path variable `challengeId` and 40 as a request parameter called `factorA`. See Listing 3-19.

Listing 3-19. An Example of Using Annotations to Map REST API URLs

```
@RestController
@RequestMapping("/challenges")
class ChallengeAttemptController {

    @GetMapping("/{challengeId}")
    public Challenge getChallengeWithParam(@PathVariable("challengeId") Long
        challengeId,
                                            @RequestParam("factorA") int factorA)
    {...}
}
```

The offered functionality doesn't stop there. We can also validate the requests given that REST controllers integrate with the `javax.validation` API. This means we can annotate the classes used during deserialization to avoid empty values or force numbers to be within a given range when we get requests from the client, just as examples.

Don't worry about the number of new concepts introduced. We'll cover them with practical examples over the following sections.

Designing Our APIs

We can use the requirements to design what functionalities we need to expose in our REST API.

- An interface to get a random, medium complexity multiplication
- An endpoint to send a guess for a given multiplication from a given user's alias

These are a read operation for challenges and an action to create attempts. Keeping in mind that multiplication challenges and attempts are different resources, we split our API in two and assign the corresponding verbs to these actions:

- GET `/challenges/random` will return a randomly generated challenge.
- POST `/attempts/` will be our endpoint to send an attempt to solve a challenge.

Both resources belong to the challenges domain. Eventually, we will also need a `/users` mapping to perform operations with our users, but we're leaving that for later since we don't need it to complete the first requirements (user story).

API-First Approach

It's normally a good practice to define and discuss the API contract within your organization before implementing it. You should include the endpoints, HTTP verbs, allowed parameters, and request and response body examples. This way, other developers and clients can verify if the exposed functionality is what they need and give feedback before you waste time implementing the wrong solution. This strategy is known as *API First*, and there are industry standards to write the API specifications, like OpenAPI.

If you want to know more about *API First* and *OpenAPI*, see <https://tpd.io/apifirst>, from Swagger, the original creators of the specification.

Our First Controller

Let's create a controller that generates a random challenge. We already have that operation in the service layer, so we only need to use that method from the controller. That's what we should do in the presentation layer: keep it isolated from any business logic. We'll use it only to model the API and validate the passed data. See Listing 3-20.

Listing 3-20. The ChallengeController Class

```
package microservices.book.multiplication.challenge;

import lombok.RequiredArgsConstructor;
import lombok.extern.slf4j.Slf4j;
import org.springframework.web.bind.annotation.*;

/**
 * This class implements a REST API to get random challenges
 */
@Slf4j
@RequiredArgsConstructor
@RestController
@RequestMapping("/challenges")
class ChallengeController {

    private final ChallengeGeneratorService challengeGeneratorService;

    @GetMapping("/random")
    Challenge getRandomChallenge() {
        Challenge challenge = challengeGeneratorService.randomChallenge();
        log.info("Generating a random challenge: {}", challenge);
        return challenge;
    }
}
```

The `@RestController` annotation tells Spring that this is a specialized component modeling a REST controller. It's a combination of `@Controller` and `@ResponseBody`, which instructs Spring to put the result of this method as the HTTP response body. As a default in Spring Boot and if not instructed otherwise, the response will be serialized as JSON and included in the response body.

We also added a `@RequestMapping("/challenges")` at the class level, so all mapping methods will have this added as a prefix.

There are also two Lombok annotations in our controller.

- `@RequiredArgsConstructor` creates a constructor with a `ChallengeGeneratorService` as the argument since the field is private and final, which Lombok understands as required. Spring uses dependency injection, so it'll try to find a bean implementing this interface, and it'll wire it to the controller. In this case, it'll take the only candidate, the service `ChallengeGeneratorServiceImpl`.
- `Slf4j` creates a logger named `log`. We use it to print a message to console with the generated challenge.

The method `getRandomChallenge()` has the `@GetMapping("/random")` annotation. It means that this method will handle GET requests to the context `/challenges/random`, the first part coming from the class-level annotation. It simply returns a `Challenge` object.

Let's now run our web application again and do a quick API test. From your IDE, run the `MultiplicationApplication` class or, from the console, use `mvnw spring-boot:run`.

Using `HTTPie` (see Chapter 2), we try our new endpoint by doing a simple GET request to `localhost` (our machine) on port 8080 (Spring Boot's default). See Listing 3-21.

Listing 3-21. Making a Request to the Newly Created API

```
$ http localhost:8080/challenges/random
HTTP/1.1 200
Connection: keep-alive
Content-Type: application/json
Date: Sun, 29 Mar 2020 07:59:00 GMT
Keep-Alive: timeout=60
Transfer-Encoding: chunked
```

```
{
    "factorA": 39,
    "factorB": 36
}
```

We got an HTTP response with its header and its body, which is a nicely serialized JSON representation of a challenge object. We did it! Our application is finally doing something.

How Automatic Serialization Works

When covering how automatic configuration works in Spring Boot, we had a look at the example of the Tomcat embedded server, and we mentioned that there are many more autoconfiguration classes included as part of the `spring-boot-autoconfigure` dependency. Therefore, this other piece of *magic* taking care of serializing a Challenge into a proper JSON HTTP response should be no longer a mystery for you. In any case, let's take a look at how this works since it's a core concept of the web module in Spring Boot. Also, it's quite common to customize this configuration in real life.

A lot of important logic and defaults for the Spring Boot Web module live in the `WebMvcAutoConfiguration` class (see <https://tpd.io/mvcauto-source>). This class collects all available HTTP message converters in the context together for later use. See a fragment of this class in Listing 3-22.

Listing 3-22. A Fragment of `WebMvcAutoConfiguration` Class Provided by Spring Web

```
@Override
public void configureMessageConverters(List<HttpMessageConverter<?>> converters) {
    this.messageConvertersProvider
        .ifAvailable((customConverters) -> converters.addAll(customConverters.
            getConverters()));
}
```

The `HttpMessageConverter` interface (<https://tpd.io/hmc-source>) is included in the core `spring-web` artifact and defines what media types are supported by the converter, what classes can convert to and from, and the `read` and `write` methods to do conversions.

Where are these converters coming from? More autoconfiguration classes.

Spring Boot includes a `JacksonHttpMessageConvertersConfiguration` class (<https://tpd.io/jhmcc-source>) that has some logic to load a bean of type `MappingJackson2HttpMessageConverter`. This logic is conditional on the presence of the class `ObjectMapper` in the classpath. That one is a core class of the Jackson libraries, the most popular implementation of JSON serialization for Java. The `ObjectMapper` is included in the `jackson-databind` dependency. The class is in the classpath because its artifact is a dependency included in `spring-boot-starter-json`, which is itself included in the `spring-boot-starter-web`.

Again, it's better to understand all this with a diagram. See Figure 3-5.

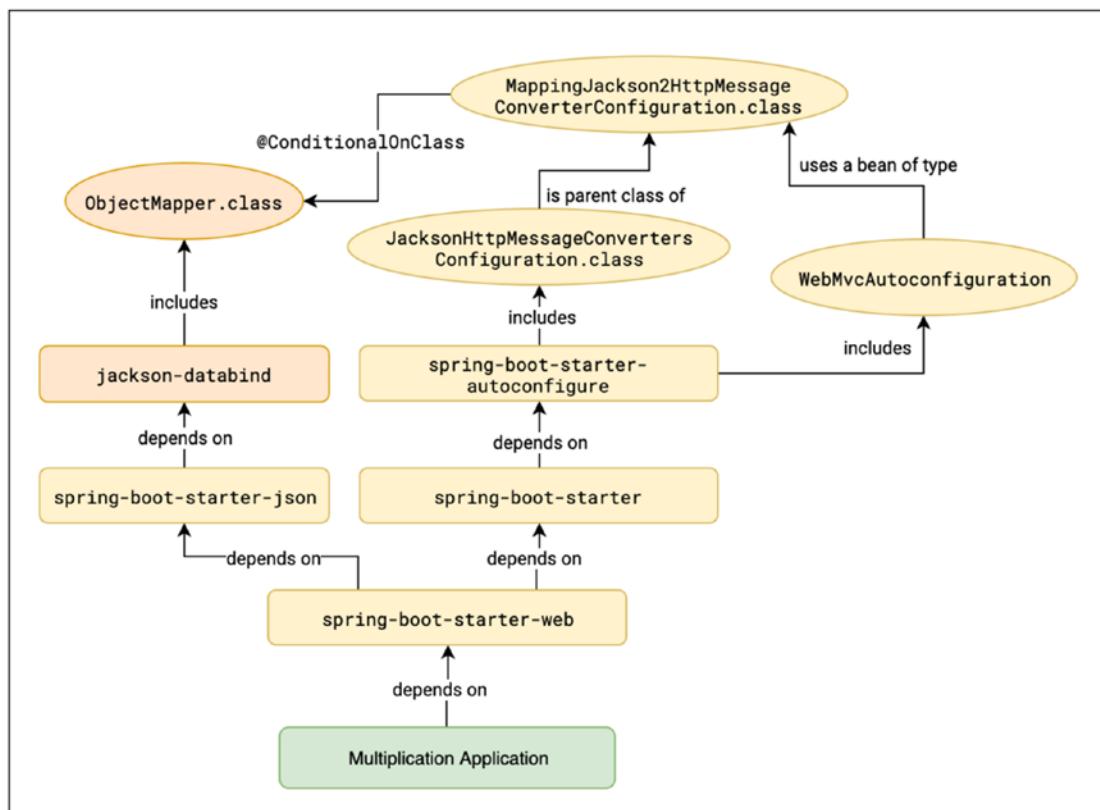


Figure 3-5. Spring Boot Web JSON autoconfiguration

CHAPTER 3 A BASIC SPRING BOOT APPLICATION

The default ObjectMapper bean is configured in the class JacksonAutoConfiguration (<https://tpd.io/jac-source>). Everything there is set up in a flexible way. If we want to customize a specific feature, we don't need to take into account this whole hierarchy. Normally, it's just a matter of overriding default beans.

For instance, if we want to change the JSON property naming to be snake-case instead of camel-case, we can declare a custom ObjectMapper in our app configuration that will be loaded instead of the default one. That's what we do in Listing 3-23.

Listing 3-23. Injecting Beans in the Context to Override Defaults in Spring Boot

```
@SpringBootApplication
public class MultiplicationApplication {

    public static void main(String[] args) {
        SpringApplication.run(MultiplicationApplication.class, args);
    }

    @Bean
    public ObjectMapper objectMapper() {
        var om = new ObjectMapper();
        om.setPropertyNamingStrategy(PropertyNamingStrategy.SNAKE_CASE);
        return om;
    }
}
```

Normally, we would add this bean declaration into a separated class annotated with @Configuration, but this piece of code is good enough for this quick example. If you run again the app and call the endpoint, you'll get the factor properties in snake-case. See Listing 3-24.

Listing 3-24. Verifying Spring Boot Configuration Changes with a New Request

```
$ http localhost:8080/challenges/random
HTTP/1.1 200
Connection: keep-alive
Content-Type: application/json
Date: Sun, 29 Mar 2020 10:05:00 GMT
Keep-Alive: timeout=60
Transfer-Encoding: chunked
```

```
{
    "factor_a": 39,
    "factor_b": 36
}
```

As you see, it's really easy to customize Spring Boot configuration by overriding beans. This specific case works because the default ObjectMapper is annotated with @ConditionalOnMissingBean, which makes Spring Boot load the bean only if there is no other bean of the same type defined in the context. Remember to remove this custom ObjectMapper since we'll use just Spring Boot defaults for now.

You might be missing already the TDD approach for these controllers. The reason why we introduced first a simple controller implementation is that it's easier for you to grasp the concepts about how controllers work in Spring Boot before diving into the testing strategies.

Testing Controllers with Spring Boot

Our second controller will implement the REST API to receive attempts to solve challenges from the front end. For this one, it's time to go back to a test-driven approach. First, we create an empty shell of the new controller. See Listing 3-25.

Listing 3-25. An Empty Implementation of the ChallengeAttemptController

```
package microservices.book.multiplication.challenge;

import lombok.RequiredArgsConstructor;
import lombok.extern.slf4j.Slf4j;
import org.springframework.web.bind.annotation.RequestMapping;
import org.springframework.web.bind.annotation.RestController;

/**
 * This class provides a REST API to POST the attempts from users.
 */
@Slf4j
@RequiredArgsConstructor
@RestController
```

CHAPTER 3 A BASIC SPRING BOOT APPLICATION

```
@RequestMapping("/attempts")
class ChallengeAttemptController {

    private final ChallengeService challengeService;

}
```

Similarly to the previous implementation, we use Lombok to add a constructor with the service interface. Spring will inject the corresponding bean `ChallengeServiceImpl`.

Now let's write a test with the expected logic. Keep in mind that testing a controller requires a slightly different approach since there is a web layer in between. Sometimes we want to verify features such as validation, request mapping, or error handling, which are configured by us but provided by Spring Boot. Therefore, we normally want a unit test that covers not only the class itself but also all these features around it.

In Spring Boot, there are multiple ways of implementing a controller test:

- Without running the embedded server. We can use `@SpringBootTest` without parameters or, even better, `@WebMvcTest` to instruct Spring to selectively load only the required configuration instead of the whole application context. Then, we simulate requests with a dedicated tool included in the Spring Test module, `MockMvc`.
- Running the embedded server. In this case, we use `@SpringBootTest` with its `webEnvironment` parameter set to `RANDOM_PORT` or `DEFINED_PORT`. Then, we have to make real HTTP calls to the server. Spring Boot includes a class `TestRestTemplate` with some useful features to perform these test requests. This option is good when you want to test some web server configuration you may have customized (e.g., custom Tomcat configuration).

The best option is usually the first one and choosing a fine-grained configuration with `@WebMvcTest`. We get all the configuration surrounding our controller without taking extra time to boot up the server for each test. If you want to get extra knowledge about all these different options, check out <https://tpd.io/sb-test-guide>.

We could write a test for one valid request and an invalid one, as shown in Listing 3-26.

Listing 3-26. Testing the Expected ChallengeAttemptController Logic

```
package microservices.book.multiplication.challenge;

import microservices.book.multiplication.user.User;
import org.junit.jupiter.api.Test;
import org.junit.jupiter.api.extension.ExtendWith;
import org.springframework.beans.factory.annotation.Autowired;
import org.springframework.boot.test.autoconfigure.json.AutoConfigureJsonTesters;
import org.springframework.boot.test.autoconfigure.web.servlet.WebMvcTest;
import org.springframework.boot.test.json.JacksonTester;
import org.springframework.boot.test.mock.mockito.MockBean;
import org.springframework.http.HttpStatus;
import org.springframework.http.MediaType;
import org.springframework.mock.web.MockHttpServletResponse;
import org.springframework.test.context.junit.jupiter.SpringExtension;
import org.springframework.test.web.servlet.MockMvc;

import static org.assertj.core.api.BDDAssertions.then;
import static org.mockito.ArgumentMatchers.eq;
import static org.mockito.BDDMockito.given;
import static org.springframework.test.web.servlet.request.MockMvcRequestBuilders.post;

@ExtendWith(SpringExtension.class)
@AutoConfigureJsonTesters
@WebMvcTest(ChallengeAttemptController.class)
class ChallengeAttemptControllerTest {

    @MockBean
    private ChallengeService challengeService;

    @Autowired
    private MockMvc mvc;

    @Autowired
    private JacksonTester<ChallengeAttemptDTO> jsonRequestAttempt;
    @Autowired
    private JacksonTester<ChallengeAttempt> jsonResultAttempt;
```

CHAPTER 3 A BASIC SPRING BOOT APPLICATION

```
@Test
void postValidResult() throws Exception {
    // given
    User user = new User(1L, "john");
    long attemptId = 5L;
    ChallengeAttemptDTO attemptDTO = new ChallengeAttemptDTO(50, 70, "john", 3500);
    ChallengeAttempt expectedResponse = new ChallengeAttempt(attemptId, user,
      50, 70, 3500, true);
    given(challengeService
        .verifyAttempt(eq(attemptDTO)))
        .willReturn(expectedResponse);

    // when
    MockHttpServletResponse response = mvc.perform(
        post("/attempts").contentType(MediaType.APPLICATION_JSON)
            .content(jsonRequestAttempt.write(attemptDTO).getJson()))
        .andReturn().getResponse();

    // then
    then(response.getStatus()).isEqualTo(HttpStatus.OK.value());
    then(response.getContentAsString()).isEqualTo(
        jsonResultAttempt.write(
            expectedResponse
        ).getJson());
}

@Test
void postInvalidResult() throws Exception {
    // given an attempt with invalid input data
    ChallengeAttemptDTO attemptDTO = new ChallengeAttemptDTO(2000, -70,
      "john", 1);

    // when
    MockHttpServletResponse response = mvc.perform(
        post("/attempts").contentType(MediaType.APPLICATION_JSON)
            .content(jsonRequestAttempt.write(attemptDTO).getJson()))
        .andReturn().getResponse();
```

```
// then
then(response.getStatus()).isEqualTo(HttpStatus.BAD_REQUEST.value());
}

}
```

There are a few new annotations and helper classes in this code. Let's review them one by one.

- `@ExtendWith(SpringExtension.class)` makes sure that our JUnit 5 test loads the extensions for Spring so we can use a test context.
- `@AutoConfigureJsonTesters` tells Spring to configure beans of type `JacksonTester` for some fields we declare in the test. In our case, we use `@Autowired` to inject two `JacksonTester` beans from the test context. Spring Boot, when instructed via this annotation, takes care of building these utility classes. A `JacksonTester` may be used to serialize and deserialize objects using the same configuration (i.e., `ObjectMapper`) as the app would do in runtime.
- `@WebMvcTest`, with the controller class as a parameter, makes Spring treat this as a presentation layer test. Thus, it'll load only the relevant configuration around the controller: validation, serializers, security, error handlers, etc. (see <https://tpd.io/test-autoconf> for a full list of included auto-configuration classes).
- `@MockBean` comes with the Spring Boot Test module and helps you develop proper unit tests by allowing you to mock other layers and beans you're not testing. In our case, we replace the service bean in the context by a mock. We set the expected return values within the test methods, using BDDMockito's `given()`.
- `@Autowired` might be familiar to you. It's a basic annotation in Spring to make it inject (or wire) a bean in the context to the field. It used to be common in all classes using Spring, but since version 4.3, it can be omitted from fields if they are initialized in a constructor and the class has only one constructor.
- The `MockMvc` class is what we use in Spring to simulate requests to the presentation layer when we make a test that doesn't load a real server. It's provided by the test context so we can just inject it in our test.

Valid Attempt Test

Now we can focus on the test cases and how to make them pass. The first test sets up the scenario for a valid attempt. It creates the DTO that acts as the data sent from the API client with a valid result. It uses BDDMockito's given() to specify that, when the service (a mocked bean) is called with an argument equal (Mockito's eq) to the DTO, it should return the expected ChallengeAttempt response.

We build the POST request with the static method post included in the helper class MockMvcRequestBuilders. Our target is the expected path /attempts. The content type is set to application/json, and its body is the serialized DTO in JSON format. We use the wired JacksonTester for serialization. Then, mvc does the request via perform(), and we get the response calling to .andReturn(). We could also call instead the method andExpect() if we would use MockMvc also for assertions, but it's better to do them separately with a dedicated assertions library like AssertJ.

In the last part of the test, we verify that the HTTP status code should be 200 OK and that the result must be a serialized version of the expected response. Again, we use a JacksonTester object for this.

This test fails with a 404 NOT FOUND when we execute it. See Listing 3-27. There is no implementation for that request, so the server can't simply find a logic to map that POST mapping.

Listing 3-27. The ChallengeAttemptControllerTest Fails

```
Expecting:  
<404>  
to be equal to:  
<200>  
but was not.
```

Then, we go back to the ChallengeAttemptController and implement this mapping. See Listing 3-28.

Listing 3-28. Adding the Working Implementation to ChallengeAttempt Controller

```
@Slf4j  
 @RequiredArgsConstructor  
 @RestController
```

```

@RequestMapping("/attempts")
class ChallengeAttemptController {

    private final ChallengeService challengeService;

    @PostMapping
    ResponseEntity<ChallengeAttempt> postResult(@RequestBody ChallengeAttemptDTO
challengeAttemptDTO) {
        return ResponseEntity.ok(challengeService.verifyAttempt
        (challengeAttemptDTO));
    }
}

```

It's a simple logic that just calls the service layer. Our method is annotated with @PostMapping without parameters so it will handle a POST request to the context path already set at the class level. Note that here we're using a ResponseEntity as the return type instead of using the ChallengeAttempt directly. That other option would also work. We're using this new way here to show that there are ways to build different types of responses with the ResponseEntity static builder.

That's it! The first test case will pass now.

Validating Data in Controllers

The second test case, postInvalidResult(), checks that an attempt with negative or out-of-range numbers shouldn't be accepted by the application. It expects our logic to return a 400 BAD REQUEST, which is a good practice when the error is on the client side, like this one. See Listing 3-29.

Listing 3-29. Verifying That the Client Gets a BAD REQUEST Status Code

```

// then
then(response.getStatus()).isEqualTo(HttpStatus.BAD_REQUEST.value());

```

If you run it before implementing our POST mapping in the controller, it fails with a NOT FOUND status code. With the implementation in place, it also fails. However, in this case, the result is even worse. See Listing 3-30.

Listing 3-30. Posting an Invalid Request Returns a 200 OK Status Code

```
org.opentest4j.AssertionFailedError:  
Expecting:  
<200>  
to be equal to:  
<400>  
but was not.
```

Our application is just accepting the invalid attempt and returning an OK status. This is wrong; we should not pass this attempt to the service layer but reject it in the presentation layer. To accomplish this, we're going to use the Java Bean Validation API (<https://tpd.io/bean-validation>) integrated with Spring.

In our DTO class, we add some Java Validation annotations to indicate what are valid inputs. See Listing 3-31. All these annotations are implemented in the jakarta.validation-api library, available in our classpath via spring-boot-starter-validation. This starter is included as part of the Spring Boot Web starter (spring-boot-starter-web).

Listing 3-31. Adding Validation Constraints to Our DTO Class

```
package microservices.book.multiplication.challenge;  
  
import lombok.Value;  
  
import javax.validation.constraints.*;  
  
/**  
 * Attempt coming from the user  
 */  
@Value  
public class ChallengeAttemptDTO {  
  
    @Min(1) @Max(99)  
    int factorA, factorB;  
    @NotBlank  
    String userAlias;  
    @Positive  
    int guess;  
}
```

There are a lot of available constraints within that package (<https://tpd.io/constraints-source>). We use `@Min` and `@Max` to define the range of allowed values for the multiplication factors, `@NotBlank` to make sure we always get an alias, and `@Positive` for the guess since we know we're handling only positive results (we could also use a predefined range here).

An important step to make these constraints work is to integrate them with Spring via the `@Valid` annotation in the controller's method argument. See Listing 3-32. Only if we add this, Spring Boot will analyze the constraints and throw an exception if they don't match.

Listing 3-32. Using the `@Valid` Annotation to Validate Requests

```
@PostMapping
 ResponseEntity<ChallengeAttempt> postResult(
     @RequestBody @Valid ChallengeAttemptDTO challengeAttemptDTO) {
    return ResponseEntity.ok(challengeService.verifyAttempt(challengeAttemptDTO));
}
```

As you may have guessed, there is autoconfiguration to handle the errors and build a predefined response when the object is not valid. By default, the error handler constructs a response with a `400 BAD_REQUEST` status code.

Starting with Spring Boot version 2.3, the validation messages are no longer included in the error response by default. This might be confusing for the callers since they don't know exactly what's wrong with the request. The reason to not include them is that these messages could potentially expose information to a malicious API client. For our educational goal, we want to enable validation messages, so we'll add two settings to our `application.properties` file. See Listing 3-33. These properties are listed in the official Spring Boot docs (<https://tpd.io/server-props>), and we'll see what they do soon.

Listing 3-33. Adding Validation Logging Configuration to the application.properties File

```
server.error.include-message=always
server.error.include-binding-errors=always
```

To verify all our validation configuration, let's now run the test again. This time it'll pass, and you'll see some extra logs, as shown in Listing 3-34.

Listing 3-34. An Invalid Request Causes Now the Expected Result

```
[Field error in object 'challengeAttemptDTO' on field 'factorB': rejected value [-70];  
[...]  
[Field error in object 'challengeAttemptDTO' on field 'factorA': rejected value  
[2000];  
[...]
```

The controller handling REST API calls for users to send attempts is working now. If we start the application again, we can play with this new endpoint via the HTTPPie command. First, we ask for a random challenge as before. Then, we post an attempt to solve it. See Listing 3-35.

Listing 3-35. Running a Standard Use Case for the Application Using HTTPPie Commands

```
$ http -b :8080/challenges/random  
{  
    "factorA": 58,  
    "factorB": 92  
}  
$ http POST :8080/attempts factorA=58 factorB=92 userAlias=moises guess=5400  
HTTP/1.1 200  
Connection: keep-alive  
Content-Type: application/json  
Date: Fri, 03 Apr 2020 04:49:51 GMT  
Keep-Alive: timeout=60  
Transfer-Encoding: chunked  
  
{  
    "correct": false,  
    "factorA": 58,  
    "factorB": 92,  
    "id": null,  
    "resultAttempt": 5400,  
    "user": {  
        "alias": "moises",  
        "id": null
```

```
    }
}
```

The first command uses the parameter `-b` to print only the body of the response. As you see, we can also omit `localhost`, and `HTTPie` will use it as default.

To send the attempt, we use the `POST` argument before the URL. `JSON` is the default content type in `HTTPie`, so we can simply pass `key=value` parameters, and this tool will convert it to proper `JSON`. As expected, we got a serialized `ChallengeAttempt` object indicating that the result is not correct.

We can also try an invalid request to see exactly how Spring Boot handles the validation errors. See Listing 3-36.

Listing 3-36. Error Response Including Validation Messages

```
$ http POST :8080/attempts factorA=58 factorB=92 userAlias=moises guess=-400
HTTP/1.1 400
Connection: close
Content-Type: application/json
Date: Sun, 16 Aug 2020 07:30:10 GMT
Transfer-Encoding: chunked

{
  "error": "Bad Request",
  "errors": [
    {
      "arguments": [
        {
          "arguments": null,
          "code": "guess",
          "codes": [
            "challengeAttemptDTO.guess",
            "guess"
          ],
          "defaultMessage": "guess"
        }
      ],
      "bindingFailure": false,
      "code": "Positive",
```

```

    "codes": [
        "Positive.challengeAttemptDTO.guess",
        "Positive.guess",
        "Positive.int",
        "Positive"
    ],
    "defaultMessage": "must be greater than 0",
    "field": "guess",
    "objectName": "challengeAttemptDTO",
    "rejectedValue": -400
}
],
"message": "Validation failed for object='challengeAttemptDTO'. Error count: 1",
"path": "/attempts",
"status": 400,
"timestamp": "2020-08-16T07:30:10.212+00:00"
}

```

It's quite a verbose response. The main reason is that all the *binding errors* (those caused by the validation constraints) are added to the error response. This is what we switched on with `server.error.include-binding-errors=always`. Besides, the `message` field also gives the client an overall description of what went wrong. This description is omitted by default, but we enabled it with the property `server.error.include-message=always`.

If this response goes to a user interface, you need to parse that JSON response in the front end, get the fields that are invalid, and maybe display the `defaultMessage` fields. Changing this default message is really simple since you can just override it with the constraint annotations. Let's modify this annotation in `ChallengeAttemptDTO` and try again with the same invalid request. See Listing 3-37.

Listing 3-37. Changing the Validation Message

```
@Positive(message = "How could you possibly get a negative result here? Try again.")
int guess;
```

What Spring Boot does in this case to handle the errors is to sneakily add a `@Controller` to your context: the `BasicErrorController` (see <https://tpd.io/bec-source>). This one uses the class `DefaultErrorAttributes` (<https://tpd.io/dea-source>) to compose the error response. If you want to dive into more details about how to customizing this behavior, you can have a look at <https://tpd.io/cust-err-handling>.

Summary and Achievements

We started this chapter with the requirements of the application we'll build in this book. Then, we sliced the scope and took the first item for development: the functionality to generate a random challenge and allow users to guess the result.

You learned how to create the skeleton of a Spring Boot application and some best practices regarding software design and architecture: three-tier and three-layer architecture, domain-driven design, test-driven/behavior-driven development, basic unit tests with JUnit 5, and REST API design. Within this chapter, you focused on the application tier and implemented the domain objects, the business layer, and the presentation layer in the form of a REST API. See Figure 3-6.

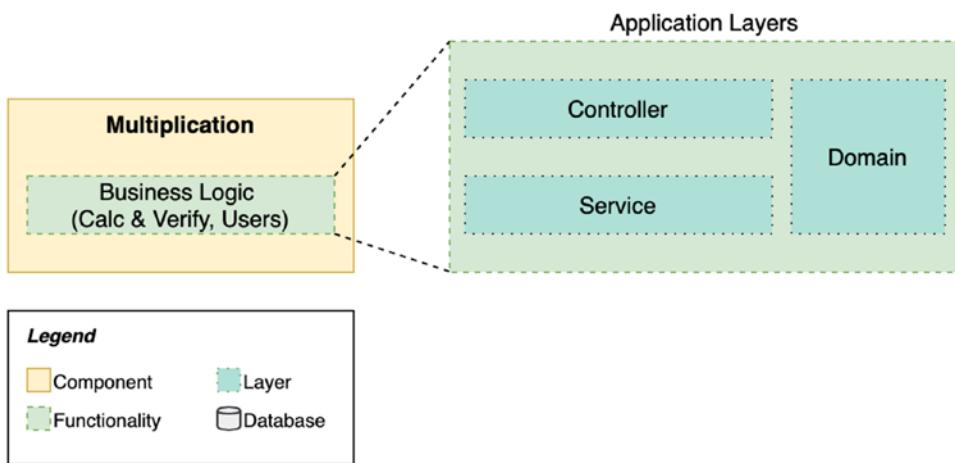


Figure 3-6. Application status after Chapter 3

A core concept in Spring Boot was also covered in this chapter: autoconfiguration. Now you know where a big part of the Spring Boot magic lives. In the future, you should be able to find your way through the reference documentation to override other default behaviors in any other configuration class.

CHAPTER 3 A BASIC SPRING BOOT APPLICATION

We also went through other features in Spring Boot such as implementing `@Service` and `@Controller` components, testing controllers with `MockMvc`, and validating input via the Java Bean Validation API.

To complete our first web application we need to build a user interface. Later, we'll also cover the data layer to make sure we can persist users and attempts.

Chapter's Achievements:

- You learned how to build a properly structured Spring Boot application, following a three-layered design.
- You understood how Spring Boot's autoconfiguration works, the key to unveil its magic, based on two practical examples with supporting diagrams: the Tomcat embedded server and the JSON serialization defaults.
- You modeled an example business case following domain-driven design techniques.
- You developed two of the three layers of the first application (*service*, *controller*) using a test-driven development approach.
- You used the most important Spring MVC annotations to implement a REST API with Spring Boot.
- You learned how to test the controller layer in Spring using `MockMVC`.
- You added validation constraints to your API to protect it against invalid input.

CHAPTER 4

A Minimal Front End with React

A book about microservices that claims to be practical has to provide a front end too. In real life, users don't interact with applications via REST APIs.

Since this book focuses on popular technologies used in real life, we'll build our front end in React. This JavaScript framework allows us to easily develop web pages based on reusable components and services. According to the 2020 StackOverflow's Developer Survey (<https://tpd.io/js-fw-list>), React is the most popular framework when compared to other similar alternatives like Angular or Vue.js. That makes it already a good choice. On top of that, it's a framework that I consider Java-developer-friendly: you can use TypeScript, an extension of JavaScript that adds types to this programming language, which makes everything easier for people used to them. Besides, React's programming style allows us to create classes to build components and services, and this makes the structure of a React project familiar for a Java developer.

We'll also use Node, a JavaScript runtime that comes with `npm`, its tool to manage JavaScript dependencies. This way you can get some practical experience with UI technologies and, why not, become a full stack developer if you aren't one yet.

In any case, keep in mind this important disclaimer: we won't dive into the details of how to build a web app with React. We want to keep the attention on microservices with Spring Boot. Therefore, don't feel bad if you don't fully grasp all the concepts in this chapter, especially if you've never seen JavaScript code or CSS.

Taking into account that you have all the source code available in the GitHub repository (<https://github.com/Book-Microservices-v2/chapter04>), you can approach this chapter in several ways.

- Read it *as is*. You'll get some basic knowledge and will be experimenting with some important concepts in React.

- Pause for a bit to read the Main Concepts Guide (<https://tpd.io/react-mc>) on the official website and then come back to this chapter. This way you will have more background knowledge about what we're going to build.
- Skip this chapter completely and use the sources from the repository in case you're not interested in front-end technologies at all. It's safe to jump to the next chapter and continue with the evolving-application approach.

A Quick Intro to React and Node

React is a JavaScript library for building user interfaces. Developed by Facebook, it has become popular among front-end developers. It's widely used in many organizations, which leads also to an active job market.

Like the other libraries, React is based on components. This is an advantage for back-end developers given that the concept of a piece of code that you write once and reuse everywhere sounds familiar.

Instead of writing HTML and JavaScript source code in separate files, in React you can use JSX, an extension of the JavaScript syntax that allows us to combine these languages. This is useful since you can write components in individual files and isolate them by functionality, keeping all behavior and rendering logic together.

Setting Up the Development Environment

First, you need to install Node.js using one of the available installer packages located at the nodejs.org site. In this book, we use Node v13.10 and npm 6.13.7. After the installation finishes, verify it with the command-line tools, as indicated in Listing 4-1.

Listing 4-1. Getting the Version of Node.js and npm

```
$ node --version  
v13.10.1  
$ npm --version  
6.13.7
```

Now you can use `npx`, a tool included with `npm`, to create the React's front-end project. Make sure you run this command from your workspace root and not inside the Multiplication service.

```
$ npx create-react-app challenges-frontend
```

Source Code

You can find all the source code for this chapter on GitHub, in the `chapter04` repository.

See <https://github.com/Book-Microservices-v2/chapter04>.

After some time downloading and installing dependencies, you'll get an output like the one shown in Listing 4-2.

Listing 4-2. Console Output After Creating the React Project

```
Success! Created challenges-frontend at /Users/moises/workspace/learn-microservices/challenges-frontend
```

Inside that directory, you can run several commands:

```
[...]
```

We suggest that you begin by typing:

```
cd challenges-frontend  
npm start
```

If you follow the suggestion and run `npm start`, a node server will start at `http://localhost:3000`, and you may even get a browser window opened and showing a predefined web page included in the application we just generated. In case you don't, you can have a quick look at this page if you navigate to `http://localhost:3000` from your browser.

The React Skeleton

The next task is to load the React project into our workspace. For example, in IntelliJ, you can use the option File > New > Module from existing sources to load the front-end folder as a separate module. As you'll see, we got a lot of files already created by the `create-react-app` tool. See Figure 4-1.

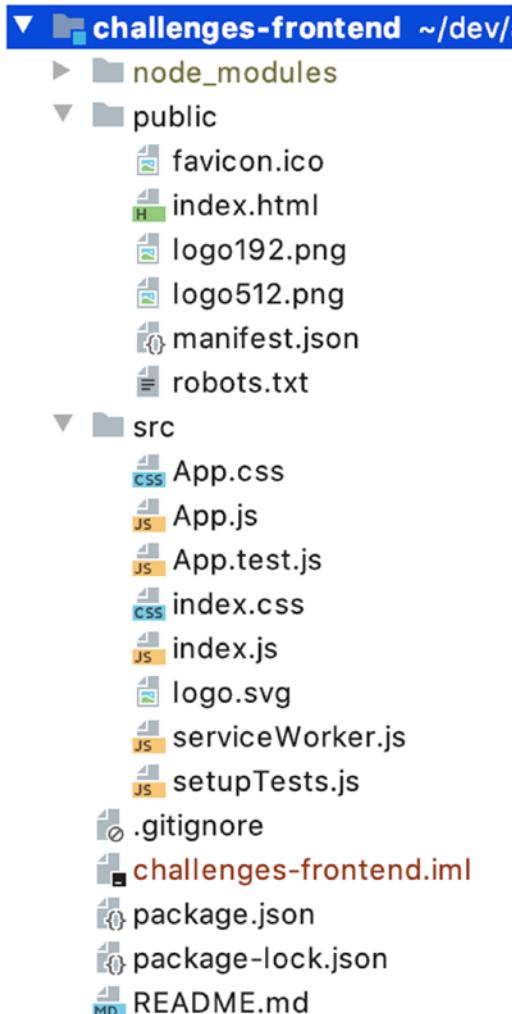


Figure 4-1. React project skeleton

- The package.json and package-lock.json are npm files. They contain basic information about the project, and they also list its dependencies. Those dependencies are stored in the node_modules folder.
- The public folder is where you can keep all the static files that will remain untouched after the build. The only exception is index.html, which will be processed to include the resulting JavaScript sources.
- All your React sources and their related resources are included in the src folder. In this skeleton app, you can find the main entrypoint file index.js and a React component, App. This sample component comes with its own stylesheet App.css and a test, App.test.js. When you build a React project, all these files end up merged into bigger files, but this naming convention and structure are helpful for development.

How do these files relate to each other in React? Let's start with index.html. See Listing 4-3 for the content of the body tag after removing the comment lines.

Listing 4-3. The root Div in HTML

```
<body>
  <noscript>You need to enable JavaScript to run this app.</noscript>
  <div id="root"></div>
</body>
```

Listing 4-4 shows a fragment of the content of the index.js file.

Listing 4-4. The Entrypoint to Render the React Content

```
ReactDOM.render(
  <React.StrictMode>
    <App />
  </React.StrictMode>,
  document.getElementById('root')
);
```

This code shows how to render a React element into the Document Object Model (DOM), a tree representation of the HTML elements. This piece of code renders the elements `React.StrictMode` and its child `App` component into the HTML. More specifically, they get rendered into the element with the ID `root`, the `div` tag inserted in `index.html`. Since `App` is a component and it may contain other components, it ends up processing and rendering the whole React application.

A JavaScript Client

Before creating our first component, let's make sure we have a way to retrieve data from the REST API we created in the previous chapter. We're going to use a JavaScript class for this. As you'll see in the rest of the chapter, we'll keep a Java-ish programming style for building our front end, using classes and types.

Classes in JavaScript are similar to Java classes. For our specific case, we can create a utility class with two static methods. See Listing 4-5.

Listing 4-5. The `ApiClient` Class

```
class ApiClient {

    static SERVER_URL = 'http://localhost:8080';
    static GET_CHALLENGE = '/challenges/random';
    static POST_RESULT = '/attempts';

    static challenge(): Promise<Response> {
        return fetch(ApiClient.SERVER_URL + ApiClient.GET_CHALLENGE);
    }

    static sendGuess(user: string,
                    a: number,
                    b: number,
                    guess: number): Promise<Response> {
        return fetch(ApiClient.SERVER_URL + ApiClient.POST_RESULT,
                    {
                        method: 'POST',
                        headers: {
                            'Content-Type': 'application/json'
                        },

```

```

        body: JSON.stringify(
          {
            userAlias: user,
            factorA: a,
            factorB: b,
            guess: guess
          }
        )
      });
    }
}

export default ApiClient;

```

Both methods return promises. A promise in JavaScript is comparable to a Java's Future class: it represents the result of an asynchronous operation. Our functions call `fetch` (see <https://tpd.io/fetch-api>), a function in JavaScript that we can use to interact with an HTTP server.

The first method, `challenge()`, uses the `fetch` function in its basic form since it'll default to a GET operation to the passed URL. This method returns a promise of a Response object (<https://tpd.io/js-response>).

The `sendGuess` method accepts the parameters we need to build the request to solve a challenge. This time, we use `fetch` with a second argument: an object defining the HTTP method (POST), the content-type of the body in our request (JSON), and the body. To build the JSON request, we use the utility method `JSON.stringify`, which serializes an object.

Last, to make our class publicly accessible, we add `export default ApiClient` at the end of the file. This makes it possible to import the complete class in other components and classes.

The Challenge Component

Let's build our first React component. We will follow modularization also in the front end, and that means this component will take care of the Challenges domain. For now, this implies the following:

- Rendering the challenge retrieved from the back end
- Displaying a form for the user to send the guess

See Listing 4-6 for the complete source code of the ChallengeComponent class. In the following sections, we'll dissect this code, and we'll use it to learn how we can structure components in React and some of its basic concepts.

Listing 4-6. Our First React Component: ChallengeComponent

```
import * as React from "react";
import ApiClient from "../services/ApiClient";

class ChallengeComponent extends React.Component {

  constructor(props) {
    super(props);
    this.state = {
      a: '',
      b: '',
      user: '',
      message: '',
      guess: 0
    };
    this.handleSubmitResult = this.handleSubmitResult.bind(this);
    this.handleChange = this.handleChange.bind(this);
  }

  componentDidMount(): void {
    ApiClient.challenge().then(
      res => {
        if (res.ok) {
          res.json().then(json => {
            this.setState({
              a: json.factorA,
              b: json.factorB
            });
          });
        } else {
          this.updateMessage("Can't reach the server");
        }
      }
    );
  }
}
```

```
handleChange(event) {
  const name = event.target.name;
  this.setState({
    [name]: event.target.value
  });
}

handleSubmitResult(event) {
  event.preventDefault();
  ApiClient.sendGuess(this.state.user,
    this.state.a, this.state.b,
    this.state.guess)
  .then(res => {
    if (res.ok) {
      res.json().then(json => {
        if (json.correct) {
          this.updateMessage("Congratulations! Your guess is
            correct");
        } else {
          this.updateMessage("Oops! Your guess " + json.
            resultAttempt +
            " is wrong, but keep playing!");
        }
      });
    } else {
      this.updateMessage("Error: server error or not available");
    }
  });
}

updateMessage(m: string) {
  this.setState({
    message: m
  });
}

render() {
  return (
    <div>
```

CHAPTER 4 A MINIMAL FRONT END WITH REACT

```
<div>
  <h3>Your new challenge is</h3>
  <h1>
    {this.state.a} x {this.state.b}
  </h1>
</div>
<form onSubmit={this.handleSubmitResult}>
  <label>
    Your alias:
    <input type="text" maxLength="12"
           name="user"
           value={this.state.user}
           onChange={this.handleChange} />
  </label>
  <br/>
  <label>
    Your guess:
    <input type="number" min="0"
           name="guess"
           value={this.state.guess}
           onChange={this.handleChange} />
  </label>
  <br/>
  <input type="submit" value="Submit"/>
</form>
<h4>{this.state.message}</h4>
</div>
);
}
}

export default ChallengeComponent;
```

The Main Structure of a Component

Our class extends `React.Component`, and this is how you create components in React. The only required method you need to implement is `render()`, which must return DOM elements to display in the browser. In our case, we build these elements using JSX (<https://tpd.io/jsx>). See Listing 4-7, which shows the main structure of our component class.

Listing 4-7. Main Structure of a Component in React

```
class ChallengeComponent extends React.Component {

  constructor(props) {
    super(props);
    this.state = {
      a: '',
      b: '',
      user: '',
      message: '',
      guess: 0
    };
    this.handleSubmitResult = this.handleSubmitResult.bind(this);
    this.handleChange = this.handleChange.bind(this);
  }

  componentDidMount(): void {
    // ... Component initialization
  }

  render() {
    return (
      // ... HTML as JSX ...
    )
  }
}
```

Typically, we also need a constructor to initialize properties, and the component's `state` (in case it's needed). In `ChallengeComponent`, we create a state to hold the retrieved challenge and the data the user is entering to solve an attempt. The argument `props` is the input passed to your component as an HTML attribute.

<`ChallengeComponent` prop1="value"/>

We don't need props for our component, yet we need to accept it as an argument and pass it to the parent constructor as it's expected if we use a constructor.

Inside the constructor two lines bind the class methods. This is required if we want to use `this` in the event handlers, which are the functions we need to implement to work with the user's input data. See Handling Events (<https://tpd.io/react-events>) if you want to know more details. We'll describe these functions later in this chapter.

The function `componentDidMount` is a lifecycle method that we can implement in React to execute logic right after the component is rendered for the first time. See Listing 4-8.

Listing 4-8. Running Logic After Rendering the Component

```
componentDidMount(): void {
    ApiClient.challenge().then(
        res => {
            if (res.ok) {
                res.json().then(json => {
                    this.setState({
                        a: json.factorA,
                        b: json.factorB
                    });
                });
            } else {
                this.updateMessage("Can't reach the server");
            }
        );
    );
}
```

What we do is call the server to retrieve a challenge, using the `ApiClient` utility class we built before. Given that the function returns a promise, we use `then()` to specify what to do when we obtain the response. The inner logic is also simple: if the response is `ok` (meaning a `2xx` status code), we parse the body as `json()`. That's also an asynchronous method, so we resolve the promise again with `then()` and pass the expected `factorA` and `factorB` from the REST API response to `setState()`.

In React, the `setState` function reloads partially the DOM. That means the browser will render again the part of the HTML that changed, so we'll see our multiplication factors on the page right after we got the response from the server. In our application,

that should be a matter of milliseconds since we're calling our own local server. In a real-life web page, you could set up a spinner, for example, to improve the user experience in the case of having a slow connection.

Rendering

JSX allows us to mix HTML and JavaScript. This is powerful since you can benefit from the simplicity of the HTML language, but you can add placeholders and JavaScript logic as well. See the complete source of the `render()` method in Listing 4-9, and its subsequent explanation.

Listing 4-9. Using `render()` with JSX to Display the Component's Elements

```
render() {
  return (
    <div>
      <div>
        <h3>Your new challenge is</h3>
        <h1>
          {this.state.a} x {this.state.b}
        </h1>
      </div>
      <form onSubmit={this.handleSubmitResult}>
        <label>
          Your alias:
          <input type="text" maxLength="12"
                 name="user"
                 value={this.state.user}
                 onChange={this.handleChange}/>
        </label>
        <br/>
        <label>
          Your guess:
          <input type="number" min="0"
                 name="guess"
                 value={this.state.guess}
                 onChange={this.handleChange}/>
        </label>
      </form>
    </div>
  )
}
```

```

        <br/>
        <input type="submit" value="Submit"/>
    </form>
    <h4>{this.state.message}</h4>
</div>
);
}

```

The Challenge component has a root div element with three main blocks. The first one displays the challenge by showing both factors included in the state. At rendering time they'll be undefined, but immediately after they'll be reloaded once we get the response from the server (the logic inside componentDidMount). A similar block is the last one; it displays the message state property, which we set when we get the response for a sent attempt request.

For users to enter their guess, we add a form that calls handleSubmitResult when submitted. This form has two inputs: a field for the user's alias and another one for the guess. Both follow the same approach: their value is a property of the state object, and they call the same function handleChange on every keystroke. This function uses the name attribute of our inputs to find the corresponding property in the component state to update. Note that event.target points to the HTML element where the event happened. See Listing 4-10 for the source code of these handler functions.

Listing 4-10. Handling User's Input

```

handleChange(event) {
    const name = event.target.name;
    this.setState({
        [name]: event.target.value
    });
}

handleSubmitResult(event) {
    event.preventDefault();
    ApiClient.sendGuess(this.state.user,
        this.state.a, this.state.b,
        this.state.guess)
        .then(res => {
            if (res.ok) {
                res.json().then(json => {

```

```

        if (json.correct) {
            this.updateMessage("Congratulations! Your guess is correct");
        } else {
            this.updateMessage("Oops! Your guess " + json.
                resultAttempt +
                " is wrong, but keep playing!");
        }
    });
} else {
    this.updateMessage("Error: server error or not available");
}
});
}

```

On form submission, we call the server's API to send a guess. When we get the response, we check if it's OK, parse the JSON, and then update the message in the state. Then, that part of the HTML DOM is rendered again.

Integration with the App

Now that we have finished the code for the component, we can use it in our application. To do so, let's modify the `App.js` file, which is the main (or root) component in our React codebase. See Listing 4-11.

Listing 4-11. Adding Our Component as a Child of `App.js`, the Root Component

```

import React from 'react';
import './App.css';
import ChallengeComponent from './components/ChallengeComponent';

function App() {
    return (
        <div className="App">
            <header className="App-header">
                <ChallengeComponent/>
            </header>
        </div>
    );
}

export default App;

```

As described earlier, the skeleton app uses this App component in the `index.js` file. When we build the code, the resulting scripts are included in the `index.html` file.

We should also either adapt the test included in `App.test.js` or simply delete it. We won't dive into details about React testing, so you can delete it for now. If you want to learn more about writing tests for React components, check the Testing chapter (<https://tpd.io/r-testing>) in the official guide.

Running Our Front End for the First Time

We modified the skeleton application built with `create-react-app` to include our custom React component. Note that we didn't get rid of other files such as stylesheets, which we could customize too. We are actually reusing some of those classes, as you can see in the code in `App.js`.

It's time to verify if our front end and our back end work together. Make sure you run the Spring Boot application first and then execute the React front end using `npm` from the front-end app's root folder.

```
$ npm start
```

After a successful compilation, this command-line tool should open your default browser and display the page located at `localhost:3000`. This is where the development server lives. See Figure 4-2 showing the web page rendered when we visit that URL from our browser.



Figure 4-2. App with blank factors

Something goes wrong there. The factors are blank, but our code retrieves them after the component rendering. Let's see how to debug this problem.

Debugging

Sometimes things don't go as expected and your app simply doesn't work. You're running the app on a browser, so how do you figure out what happens? The good news is that most of the popular browsers come with powerful tools for developers. In Chrome, you can use the Chrome DevTools (see <https://tpd.io/devtools>). Use Ctrl+May+I (Windows) or Cmd+Opt+I (Mac) to open an area in your browser with several tabs and sections showing network activity, the JavaScript console, etc.

Open the Development Mode and refresh your browser. One of the functionalities you can check is if your front end is interacting properly with the server. Click the Network tab and, within the list, you'll see a failing HTTP request to `http://localhost:8080/challenges/random`, as displayed in Figure 4-3.

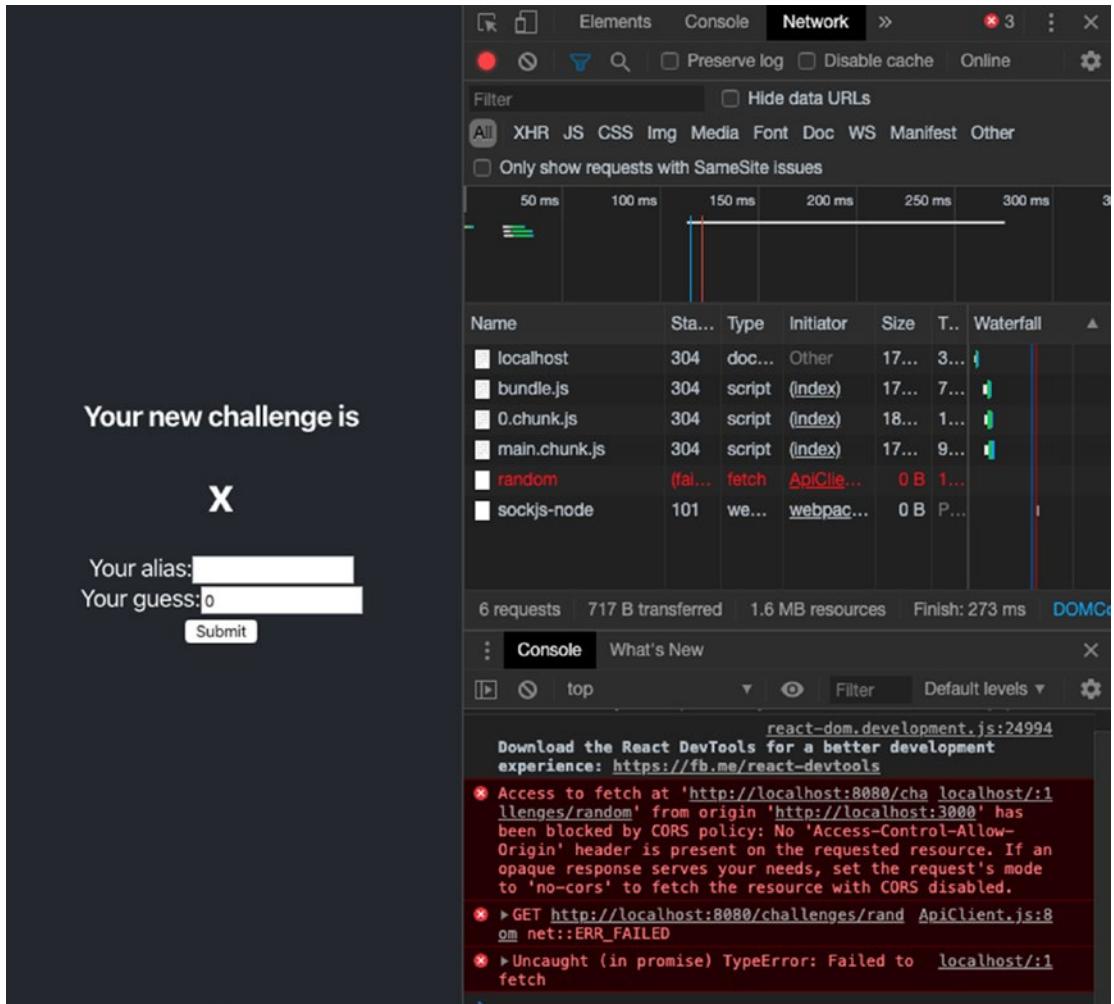


Figure 4-3. Chrome DevTools

This console also shows a descriptive message:

“Access to fetch at ‘http://localhost:8080/challenges/random’ from origin ‘http://localhost:3000’ has been blocked by CORS policy: No ‘Access-Control-Allow-Origin’ header is present on the requested resource [...]”.

By default, your browser blocks requests that try to access resources in a different domain than the one in which your front end is located. This is to avoid that a malicious page in your browser has access to data in a different page, and it's called the *same-origin policy*. In our case, we're running both the front end and the back end in localhost, but they run on different ports, so they are considered different *origins*.

There are multiple options to fix this. In our case, we're going to enable cross-origin resource sharing (CORS), a security policy that can be enabled on the server side to allow our front end to work with our REST API from a different origin.

Adding CORS Configuration to the Spring Boot App

We go back to the back-end codebase and add a Spring Boot @Configuration class that will override some defaults. According to the reference documentation (<https://tpd.io/spring-cors>), we can implement the interface WebMvcConfigurer and override the method addCorsMapping to add a generic CORS configuration. To keep classes organized, we create a new package named configuration for this class. See Listing 4-12.

Listing 4-12. Adding the CORS Configuration to the Back-End Application

```
package microservices.book.multiplication.configuration;

import org.springframework.context.annotation.Configuration;
import org.springframework.web.servlet.config.annotation.CorsRegistry;
import org.springframework.web.servlet.config.annotation.WebMvcConfigurer;

@Configuration
public class WebConfiguration implements WebMvcConfigurer {

    @Override
    public void addCorsMappings(final CorsRegistry registry) {
        registry.addMapping("/**").allowedOrigins("http://localhost:3000");
    }
}
```

This method works with an injected CorsRegistry instance that we can customize. We add a mapping allowing the front end's origin to access *any* path, represented by `/**`. We could also omit the `allowedOrigins` part in this line. Then, all origins would be allowed instead of only `http://localhost:3000`.

Remember that Spring Boot scans your packages looking for configuration classes. This is one of them, so this CORS configuration will be applied automatically the next time you start the application.

An important remark about CORS, in general, is that you probably need it only for development purposes. If you deploy your application's front end and back end to the same host, you won't experience any issue, and you shouldn't enable CORS to keep the security policies as strict as you can. When you deploy the back end and front end to different hosts, you should still be very selective in your CORS configuration and avoid adding complete access to all origins.

Playing with the Application

Now our front end and back end should work together. Restart the Spring Boot app if you haven't done it yet and refresh your browser (Figure 4-4).

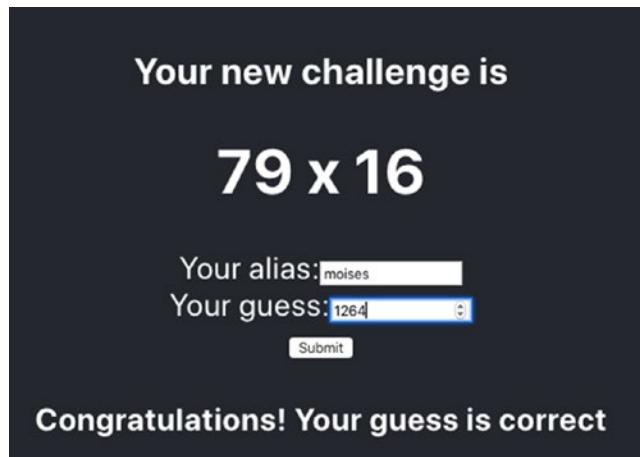


Figure 4-4. The first version of our application

Exciting times! Now you can enter your alias and make a few attempts. Remember to respect the rules and use only your brain to guess the result.

Deploying the React App

So far we've been using the development mode for our front end. We started the web server with `npm start`. This is not how it would work in a production environment, of course.

To prepare our React application for deployment, we need to build it first. See Listing 4-13.

Listing 4-13. Building the React App for a Production Deployment

```
$ npm run build
```

```
> challenges-frontend@0.1.0 build /Users/moises/dev/apress2/learn-microservices/
challenges-frontend
> react-scripts build
```

Creating an optimized production build...

Compiled successfully.

File sizes after gzip:

```
39.92 KB (+540 B)  build/static/js/2.548ff48a.chunk.js
1.32 KB (+701 B)   build/static/js/main.3411a94e.chunk.js
782 B              build/static/js/runtime-main.8b342bfc.js
547 B              build/static/css/main.5f361e03.chunk.css
```

The project was built assuming it is hosted at `/`.

You can control this with the `homepage` field in your `package.json`.

The build folder is ready to be deployed.

You may serve it with a static server:

```
npm install -g serve
serve -s build
```

Find out more about deployment here:

bit.ly/CRA-deploy

As you can see, this command generated all the scripts and files under the `build` folder. We find there as well a copy of the files we placed in the `public` folder. These logs also tell us how to install a static web server using `npm`. But actually, we already have a web server, Tomcat, embedded in the Spring Boot application. Couldn't we just use that one? Sure we can.

For our deployment example, we'll follow the easiest path and pack our entire app, back end and front end, within the same deployable unit: the *fat JAR* file generated by Spring Boot.

What we need to do is to copy all the files inside the front end's build folder to a folder named `static` inside the `src/main/resources` folder in the Multiplication codebase. See Figure 4-5. The default server configuration in Spring Boot adds some predefined locations for static web files, and this `static` folder in our classpath is one of them. These files will be mapped to the root context of the application, located at `/`.

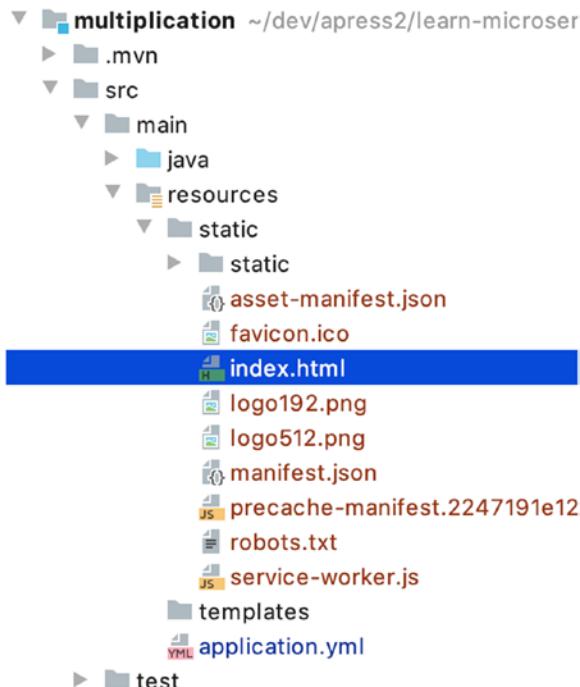


Figure 4-5. Static resources in the project structure

As usual, you could configure these resource locations and their mappings if you want. One of the places where you can fine-tune this is actually the same WebMvcConfigurer interface implementation that we used for the CORS registry configuration. Check the section Static Content in the Spring Boot reference documentation if you want to know more about configuring the web server to serve static pages (<https://tpd.io/mvc-static>).

Then, we restart the multiplication application. This time it is important you run it via the command line (not through your IDE), using `./mvnw spring-boot:run`. The reason is that IDEs might use the classpath differently while running the app, and you could get errors in that case (e.g., the page isn't found).

If we navigate to `http://localhost:8080`, the embedded Tomcat server in our Spring Boot application will try to find a default `index.html` page, which exists because we copied it from our React build. We have our React application now loaded from the same embedded server we use for the back-end side. See Figure 4-6.

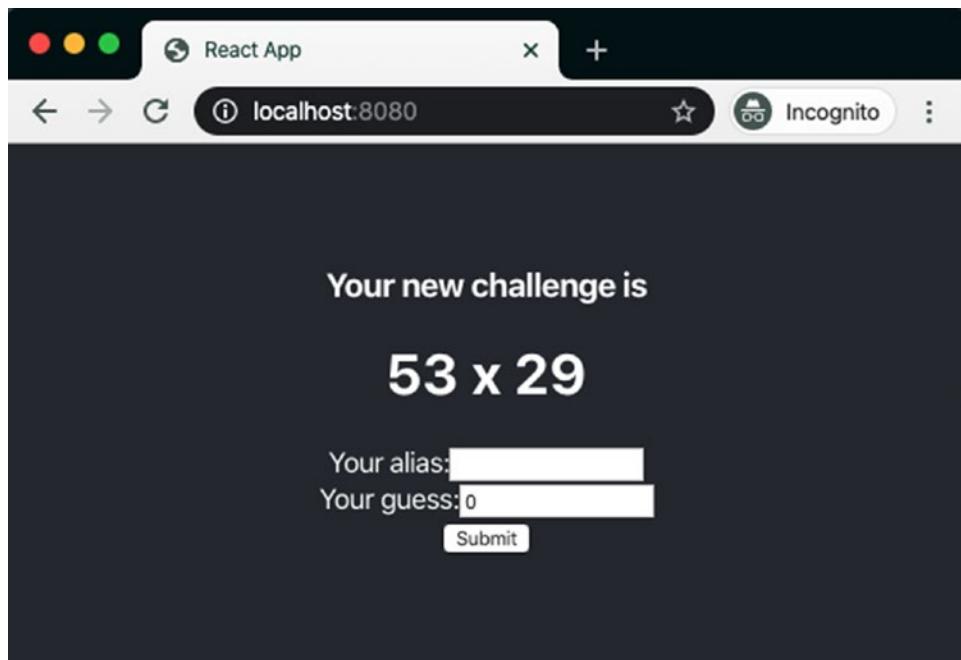


Figure 4-6. React app served from embedded Tomcat

You might be wondering what happens now with the CORS configuration we added in a previous section, given that now the front end and back end share the same origin. The CORS addition is no longer needed when we deploy the React application inside the same server. You could remove it since both the static front-end files and the back-end API are located in the origin `http://localhost:8080`. Anyway, let's keep that configuration there since we'll be using the development server while we evolve our React app. You can now remove again the contents inside the static folder in our Spring Boot application.

Summary and Achievements

It's time to look back at what we achieved within this chapter. When we started, we had a REST API we interacted with through command-line tools. Now, we added a user interface that interacts with the back end to retrieve challenges and send attempts. We have a real web application for our users. See Figure 4-7.

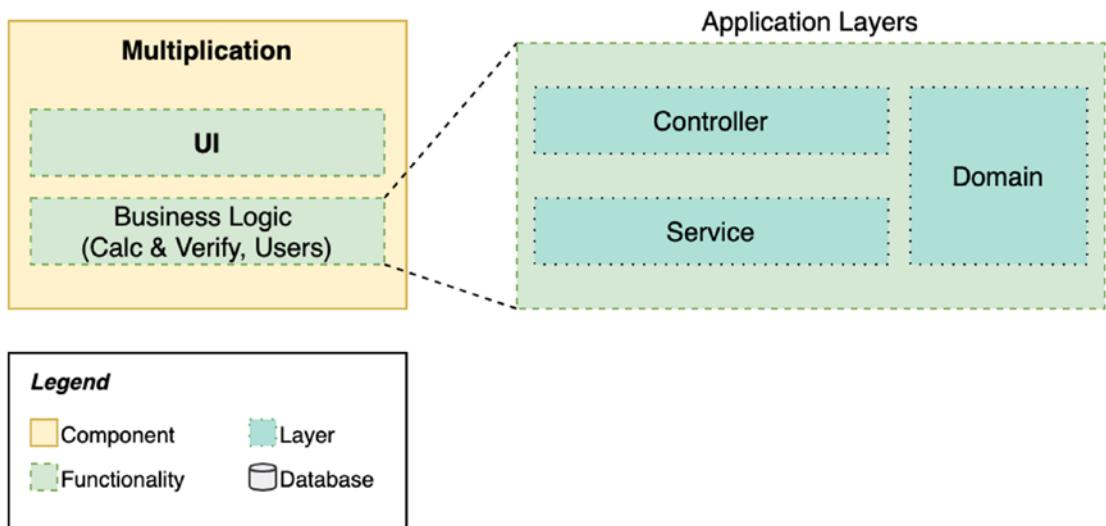


Figure 4-7. Logical view of our application at the end of Chapter 4

We created the foundations of the React application using the `create-react-app` tool, and we had a look at how it is structured. Then, we developed a service in JavaScript to connect with the APIs, as well as a React component that uses this service and renders a simple HTML code block.

To be able to interconnect the back end and front end living in different origins, we added CORS configuration to the back end.

Finally, we saw how to build our React application for production. We also took the resulting static files and moved them to the back-end project codebase to illustrate how to serve this static content from the embedded Tomcat server.

Ideally, this chapter helped you understand the basics of a front-end application and see a practical example of interaction with your APIs. Even being only the basics, that knowledge might be helpful in your career.

We'll use this front-end application in the next chapters to illustrate how a microservice architecture can impact the REST API clients.

Chapter's Achievements:

- You learned the basics of React, one of the most popular JavaScript frameworks in the market.
- You built the skeleton of a React application using the `create-react-app` tool.
- You developed a React component with a basic user interface for the user to send attempts.
- You understood what CORS is and how we can add exceptions in the back end to allow these requests.
- You had a quick look at how to debug the front end using the browser's developer tools.
- You learned how to package the HTML and JavaScript resulting from the React project's build, as well as how to distribute it within the same JAR file as the back-end application.
- You saw the application working for the first time, both the back end and the front end, in its minimal version.

CHAPTER 5

The Data Layer

It took us two chapters to complete our first user story. Now, we have a minimum viable product (MVP) that we can experiment with. In agile, slicing requirements in this way is really powerful. We could start collecting feedback from some test users and decide what is the next feature we should build. Also, it's early enough to change something if our product ideas were wrong.

Learning how to slice your product requirements vertically instead of horizontally may save you a lot of time while building software. That means you don't wait until you have a full layer completed to move to the next one. Instead, you develop pieces in multiple layers to be able to have something that works. This also helps you build a better product or service since you'll get feedback when you can easily react to it. If you want to know more about strategies for story splitting, check out <http://tpd.io/story-splitting>.

Let's imagine that our test users gave our application a try. Most of them came back to us saying that it would be great if they could access their statistics to know how they're performing over time. The team sits together and comes back with a new user story.

User Story 2

As a user of the application, I want to have access to my last attempts so I can see if I'm improving my brain skills over time.

When mapping this story to a technical solution, we quickly notice that we need to store the attempts somewhere. In this chapter, we'll introduce the missing layer in our three-layer application architecture: the data layer. That also means we'll be working with a different tier of the three-tier architecture: the database. See Figure 5-1.

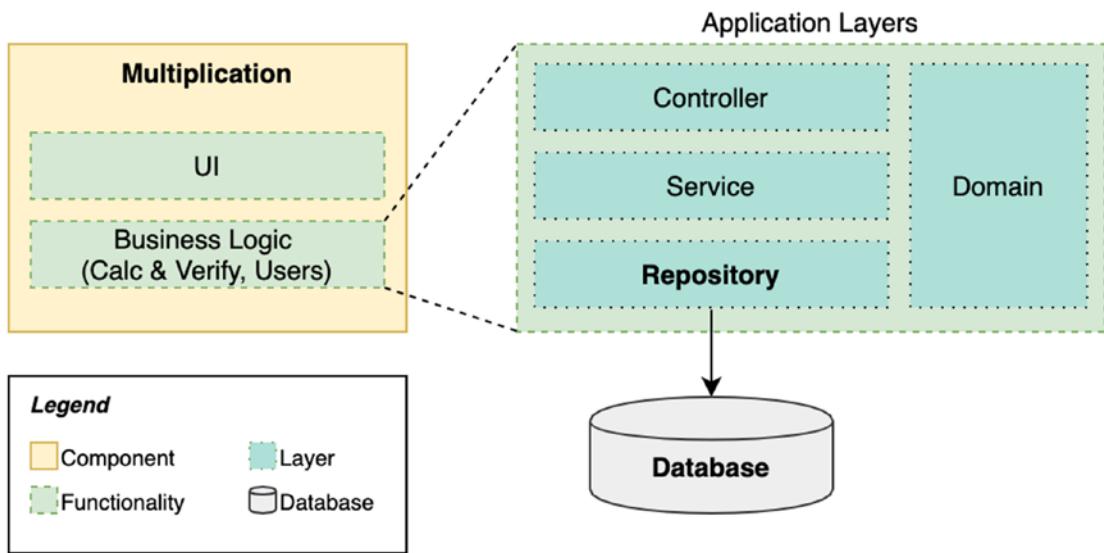


Figure 5-1. Our target application design

We'll need to integrate these new requirements into the rest of the layers too. To sum up, we could work with this list of tasks:

- Store all the user attempts and have a way to query them per user.
- Expose a new REST endpoint to get the latest attempts for a given user.
- Create a new service (business logic) to retrieve those attempts.
- Show the attempts' history to the users on the web page after they send a new one.

The Data Model

In the conceptual model we created in Chapter 3, there were three domain objects: Users, Challenges, and Attempts. Then, we made the decision to break the link between a challenge and an attempt. Instead, to keep our domain simple, we copied both factors inside attempts. That leaves us with only one relationship to model between our objects: attempts belong to a particular user.

Note that we could have gone one step further in our simplification and include the user's data (for now, the alias) in the attempts as well. In that case, the only objects we would have needed to store now would be the attempts. We could then use the user alias within the same table to query our data. But that comes with a price, higher than the one we assumed by copying the factors: we considered Users a different domain that may evolve over time and have interactions with other domains. Mixing up domains so tightly in the data layer is not a good idea.

There was also another design alternative. We could have created our domain classes by mapping exactly our conceptual domain with three separate objects and have a link between ChallengeAttempt and Challenge. See Figure 5-2.

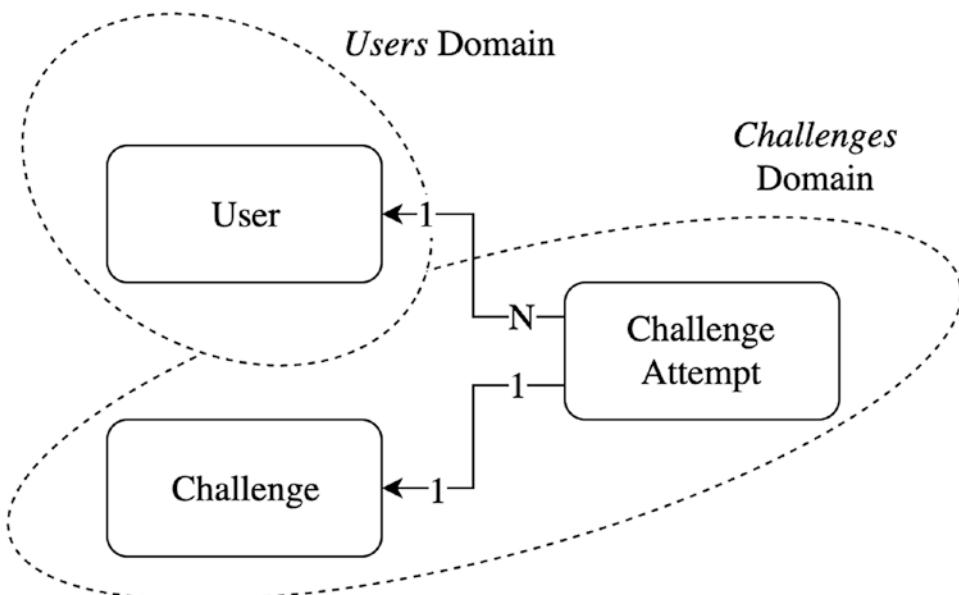


Figure 5-2. A reminder of the conceptual model

This could have been accomplished the same way we did with User. See Listing 5-1.

Listing 5-1. An Alternative Implementation of ChallengeAttempt

```

@Getter
@ToString
@EqualsAndHashCode
@AllArgsConstructor
public class ChallengeAttempt {
  
```

```

private Long id;
private User user;

// We decided to include factors
// private final int factorA;
// private final int factorB;

// This is an alternative
private Challenge challenge;

private int resultAttempt;
private boolean correct;
}

```

Then, we could have opted for a simplification only now, while designing the data model. In that approach, we would have the new version of the domain class `ChallengeAttempt`, as shown in the previous code snippet, and a different class in the data layer. We could name that class `ChallengeAttemptDataObject`, for example. That one would include the factors inside, so we would need to implement mappers between layers to combine and split challenges and attempts. As you may have identified already, this approach is similar to what we did with the DTO pattern. Back then, we created a new version of the `Attempt` object in the presentation layer, where we also added some validation annotations.

As in many other aspects of software design, there are multiple opinions in favor and against of having DTOs, domain classes, and data classes completely isolated. One of the main advantages, as we saw already in our hypothetical case, is that we get an even higher level of isolation. We could replace the implementation of the data layer without having to modify the code in the service layer. A big disadvantage, though, is the amount of code duplication and complexity we introduce in our application.

In this book, we follow a pragmatic approach and try to keep things simple while we keep applying proper design patterns. We chose a domain model in the previous chapter that we can now map directly to our data model. Therefore, we can reuse the same classes for domain and data representations. It's a good compromise solution since we still keep our domains isolated. See Figure 5-3 showing the objects and the relationship we have to persist in our database.

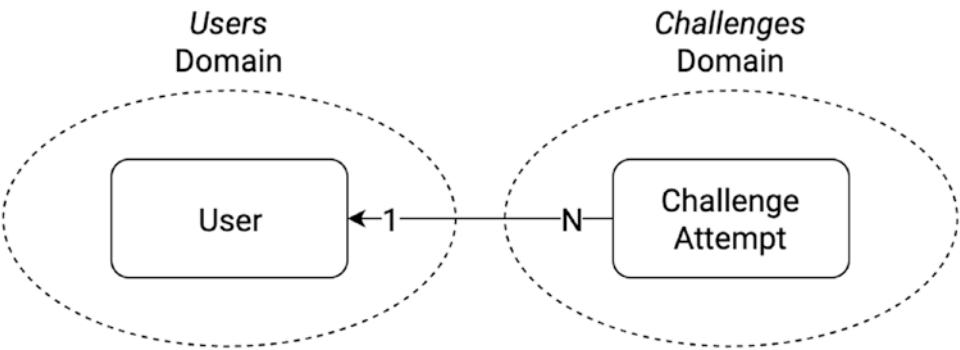


Figure 5-3. Data model of the Multiplication application

Choosing a Database

This section will discuss how to choose a database for our project considering the project's requirements, and the abstraction level that we'll use.

SQL vs. NoSQL

There are plenty of database engines available in the market. Each of them has its own particularities, but, most of the time, everybody groups them in two: SQL and NoSQL. SQL databases are relational, with a fixed schema, and they allow us to make complex queries. NoSQL databases are intended for unstructured data and can be oriented for example to key-value pairs, documents, graphs, or column-based data.

In short, we could also say that NoSQL databases are better for big volumes of records since these databases are distributed. We can deploy multiple nodes (or instances), so they allow for good performance at writing data, reading it, or both. The price we pay is that these databases follow the CAP theorem (https://en.wikipedia.org/wiki/CAP_theorem). When we store data in a distributed way, we have to choose only two of the availability, consistency, and partition tolerance guarantees. We normally want partition tolerance since network errors will simply happen, so we should be able to cope with them. Therefore, most of the time we have to choose between making the data available as much time as possible or making it consistent.

On the other hand, relational databases (SQL) follow the ACID guarantees: *atomicity* (transactions either succeed or fail as a whole unit), *consistency* (data always transitions between valid states), *isolation* (ensures that concurrency doesn't cause

side effects), and *durability* (after a transaction the state is persisted even in the event of a system failure). Those are great features, but to ensure them, these databases can't deal properly with horizontal scalability (multiple distributed nodes), meaning they don't scale that well.

It's important that you analyze carefully what your data requirements are. How are you planning to query the data? Do you need high availability? Are you writing millions of records? Do you need very fast readings? Also, keep in mind the nonfunctional requirements of your system. For example, in our particular case, we could accept that the system is not available for a few hours per year (or even days). However, we would be in a different situation if we're developing a web application for the healthcare sector where lives might be at risk. We'll come back to nonfunctional requirements in the coming chapters to analyze some of them in more detail.

Our model is *relational*. Besides, we don't plan to deal with millions of concurrent reads and writes. We'll choose a SQL database for our web application to benefit from the ACID guarantees.

In any case, one of the advantages of keeping our application (a future microservice) small enough is that we could change the database engine later in case we need it, without a big impact on the overall software architecture.

H2, Hibernate, and JPA

The next step is to decide what relational database we pick from all the possibilities: MySQL, MariaDB, PostgreSQL, H2, Oracle SQL, etc. In this book, we choose the H2 Database Engine since it's small and easy to install. It's so easy that it can be embedded within our application.

On top of the relational database, we'll go for an object/relational mapping (ORM) framework: Hibernate ORM. Instead of dealing with tabular data and plain queries, we'll use Hibernate to map our Java objects to SQL records. If you want to know more about ORM technologies, check out <http://tpd.io/what-is-orm>.

Instead of using the native API in Hibernate to map our objects to database records, we'll use an abstraction: the Java Persistence API (JPA).

This is how our technology choices relate to each other:

- From our Java code, we'll use the Spring Boot JPA annotations and integrations, so we keep our code decoupled from Hibernate specifics.

- On the implementation side, Hibernate takes care of all the logic to map our objects to database entities.
- Hibernate supports multiple SQL dialects for different databases, and the H2 dialect is one of them.
- Spring Boot autoconfiguration sets up H2 and Hibernate for us, but we can also customize behaviors.

This loose coupling between specifications and implementations gives us a big advantage: changing to a different database engine would be seamless since it's abstracted by Hibernate and Spring Boot configuration.

Spring Boot Data JPA

Let's analyze what the Spring Boot Data JPA module offers.

Dependencies and Autoconfiguration

The Spring Framework has multiple modules available to work with databases, grouped into the Spring Data family: JDBC, Cassandra, Hadoop, Elasticsearch, etc. One of them is Spring Data JPA, which abstracts access to databases using the Java Persistence API in a Spring-based programming style.

Spring Boot takes the extra step with a dedicated starter that uses autoconfiguration and some extra tooling to quickly bootstrap database access: the `spring-boot-starter-data-jpa` module. It can also autoconfigure embedded databases such as H2, our choice for the application.

We didn't add these dependencies when we created the application to respect the step-by-step approach. Now it's time to do that. In our `pom.xml` file, we add the Spring Boot starter and the H2 embedded database implementation. See Listing 5-2. We only need the H2 artifact in runtime since we'll be using the JPA and Hibernate abstractions in our code.

Listing 5-2. Adding the Data Layer Dependencies to Our Application

```
<dependencies>
[...]
<dependency>
    <groupId>org.springframework.boot</groupId>
    <artifactId>spring-boot-starter-data-jpa</artifactId>
</dependency>
<dependency>
    <groupId>com.h2database</groupId>
    <artifactId>h2</artifactId>
    <scope>runtime</scope>
</dependency>
[...]
</dependencies>
```

Source Code

You can find all the source code for this chapter on GitHub, in the `chapter05` repository.

See <https://github.com/Book-Microservices-v2/chapter05>.

Hibernate is the reference implementation for JPA in Spring Boot. That means that the starter brings the Hibernate dependencies inside. It includes also the core JPA artifacts and the dependency with its parent module, Spring Data JPA.

We already mentioned that H2 can behave as an embedded database. Therefore, we don't need to install, start, or shut down the database ourselves. Our Spring Boot application will control its lifecycle. Nevertheless, we'd also like to access the database from outside for our educational purposes, so let's add a property in the application.properties file to enable the H2 database console.

```
# Gives us access to the H2 database web console
spring.h2.console.enabled=true
```

The H2 console is a simple web interface that we can use to manage and query data. Let's verify that this new configuration works by starting our application again. We'll see some new log lines, coming from the Spring Boot Data JPA autoconfiguration logic. See Listing 5-3.

Listing 5-3. Application Logs Showing Database Autoconfiguration

```
INFO 33617 --- [main] o.s.web.context.ContextLoader : Root
WebApplicationContext: initialization completed in 1139 ms
INFO 33617 --- [main] com.zaxxer.hikari.HikariDataSource : HikariPool-1 -
Starting...
INFO 33617 --- [main] com.zaxxer.hikari.HikariDataSource : HikariPool-1 -
Start completed.
INFO 33617 --- [main] o.s.b.a.h2.H2ConsoleAutoConfiguration : H2 console
available at '/h2-console'. Database available at 'jdbc:h2:mem:testdb'
INFO 33617 --- [main] o.hibernate.jpa.internal.util.LogHelper : HHH000204:
Processing PersistenceUnitInfo [name: default]
INFO 33617 --- [main] org.hibernate.Version : HHH000412:
Hibernate ORM core version 5.4.12.Final
INFO 33617 --- [main] o.hibernate.annotations.common.Version : HCANN000001:
Hibernate Commons Annotations {5.1.0.Final}
INFO 33617 --- [main] org.hibernate.dialect.Dialect : HHH000400: Using
dialect: org.hibernate.dialect.H2Dialect
INFO 33617 --- [main] o.h.e.t.j.p.i.JtaPlatformInitiator : HHH000490: Using
JtaPlatform implementation: [org.hibernate.engine.transaction.jta.platform.
internal.NoJtaPlatform]
INFO 33617 --- [main] j.LocalContainerEntityManagerFactoryBean : Initialized JPA
EntityManagerFactory for persistence unit 'default'
```

Spring Boot detects Hibernate in the classpath and configures a data source. Since H2 is also available, Hibernate connects to H2 and selects the H2Dialect. It also initialized an EntityManagerFactory for us; we'll see soon what that means. There is also a log line claiming that the H2 console is available at /h2-console and that there is a database, available at jdbc:h2:mem:testdb. If there is no other configuration specified, Spring Boot autoconfiguration creates a ready-to-use, in-memory database named testdb.

Let's navigate to <http://localhost:8080/h2-console> to take a look at the console UI. See Figure 5-4.

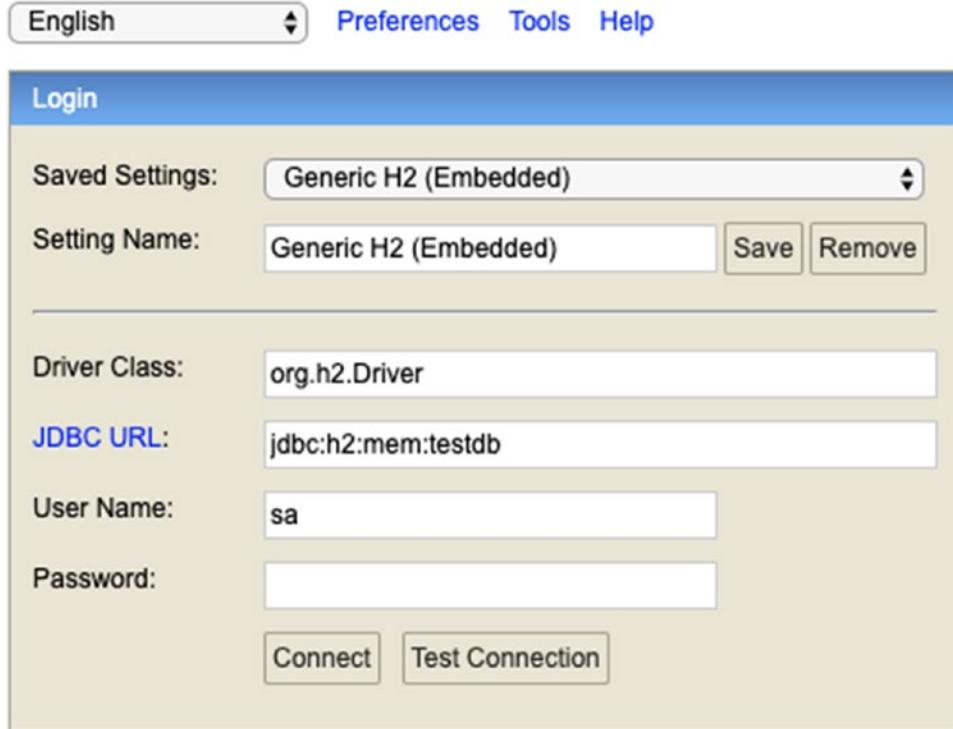


Figure 5-4. H2 Console, login

We can copy and paste `jdbc:h2:mem:testdb` as the JDBC URL and leave other values as they are. Then, we click Connect, and we have access to the main console view. See Figure 5-5.



Figure 5-5. H2 console, connected

It seems that we have indeed an in-memory database named testdb and that we're able to connect to it using the H2 default administrator credentials. Where is this database coming from? That's something we'll analyze soon.

We'll use the H2 console interface later in this chapter to query our data. For now, let's continue our learning path by exploring the technology stack that comes with Spring Boot and the Data JPA starter.

Spring Boot Data JPA Technology Stack

Let's start from the lowest level, using Figure 5-6 as visual support. There are some core Java APIs to handle SQL databases in the packages `java.sql` and `javax.sql`. There, we can find the interfaces `DataSource`, `Connection`, and some others for pooled resources such as `PooledConnection` or `ConnectionPoolDataSource`. We can find multiple implementations of these APIs by different vendors. Spring Boot comes with HikariCP (<http://tpd.io/hikari>), which is one of the most popular implementations of `DataSource` connection pools because it's lightweight and has a good performance.

Hibernate uses these APIs (and therefore the HikariCP implementation in our application) to connect to the H2 database. The JPA flavor in Hibernate for managing the database is the `SessionImpl` class (<http://tpd.io/h-session>), which includes *a lot* of code to perform statements, execute queries, handle the session's connections, etc. This class, via its hierarchy tree, implements the JPA interface `EntityManager` (<http://tpd.io/jpa-em>). This interface is part of the JPA specification. Its implementation, in Hibernate, is what does the complete ORM.

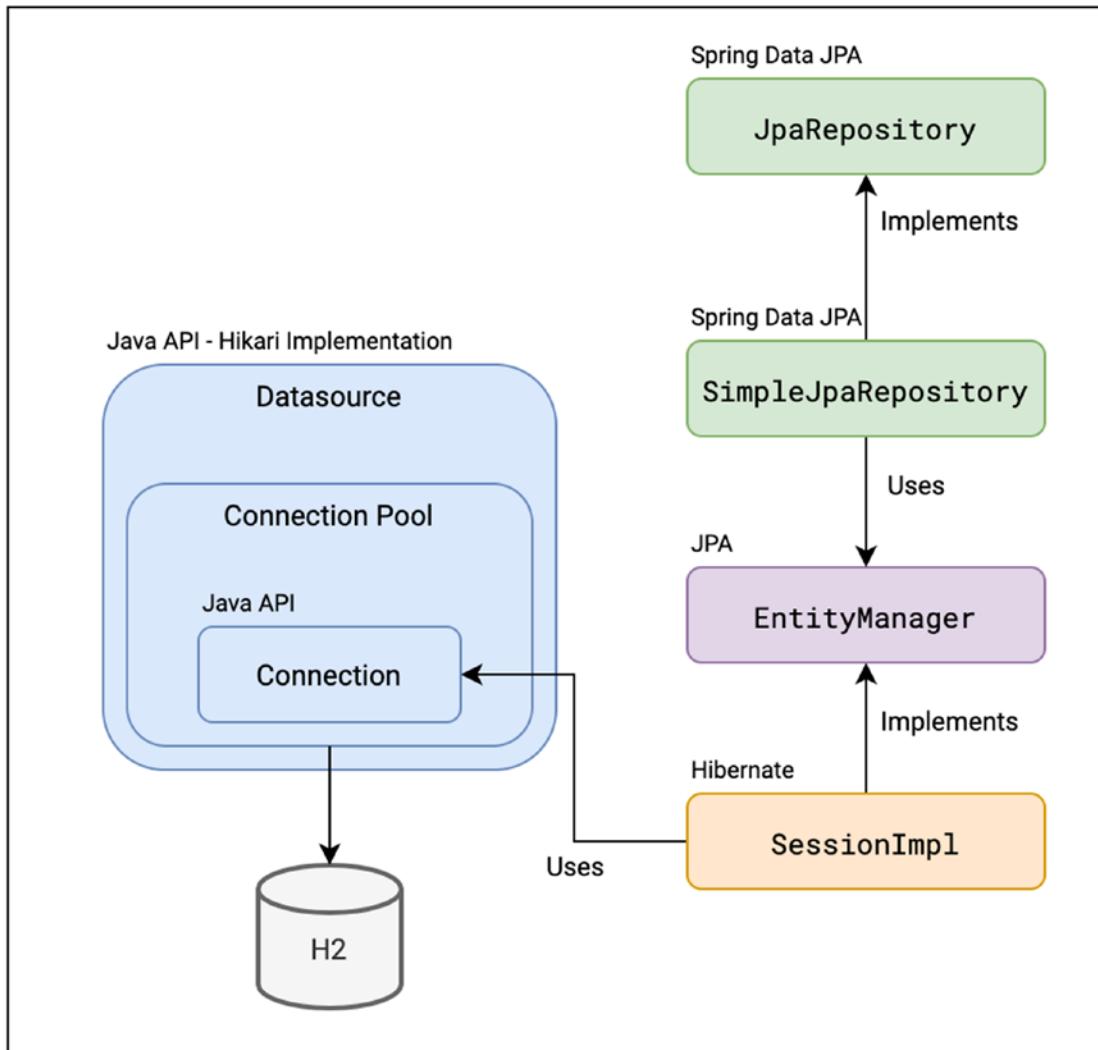


Figure 5-6. Spring Data JPA technology stack

On top of JPA's **EntityManager**, Spring Data JPA defines a **JpaRepository** interface (<http://tpd.io/jpa-repo>) with the most common methods we need to use normally: `find`, `get`, `delete`, `update`, etc. The **SimpleJpaRepository** class (tpd.io/simple-jpa-repo) is the default implementation in Spring and uses the **EntityManager** under the hood. This means we don't need to use the pure JPA standard nor Hibernate to perform database operations in our code since we can use these Spring abstractions.

We'll explore later in this chapter some of the cool features that Spring offers with the JPA Repository classes.

Data Source (Auto)configuration

There is something that might have amazed you when we ran again our application with the new dependencies. We didn't configure the data source yet, so why were we able to open a connection successfully with H2? The answer is always *autoconfiguration*, but this time it comes with a bit of extra magic.

Normally, we configure the data source using some values in our application properties. These properties are defined by the `DataSourceProperties` class (<http://tpd.io/dsprops>) within the Spring Boot autoconfiguration dependency, which contains the database's URL, username, and password, for example. As usual, there is also a `DataSourceAutoConfiguration` class (<http://tpd.io/ds-autoconfig>) that uses these properties to create the necessary beans in the context. In this case, it creates the `DataSource` bean to connect to the database.

The `sa` username comes actually from a piece of code within Spring's `DataSourceProperties` class. See Listing 5-4.

Listing 5-4. A Fragment of Spring Boot's `DataSourceProperties` Class

```
/**  
 * Determine the username to use based on this configuration and the environment.  
 * @return the username to use  
 * @since 1.4.0  
 */  
public String determineUsername() {  
    if (StringUtils.hasText(this.username)) {  
        return this.username;  
    }  
}
```

```

if (EmbeddedDatabaseConnection.isEmbedded(determineDriverClassName())) {
    return "sa";
}
return null;
}

```

Since the Spring Boot developers know these conventions, they can prepare Spring Boot so we can work with a database out of the box. There is no need to pass any configuration because they hard-coded the username, and the password is an empty String by default. There are other conventions such as the database name; that's how we also got the testdb database.

We won't use the default database created by Spring Boot. Instead, we set the name after the application's name, and we change the URL to create a database stored in a file. If we would go ahead with the in-memory database, all the attempts would be lost when we shut down the application. Besides, we have to add the parameter DB_CLOSE_ON_EXIT=false as described in the reference documentation (see this <http://tpd.io/sb-embed-db>), so we disable automatic shutdown and let Spring Boot decide when to close the database. See Listing 5-5 for the resulting URL, and the rest of the changes we're including in our application.properties file. There is some extra explanation afterward.

Listing 5-5. application.properties File with New Parameters for Database Configuration

```

# Gives us access to the H2 database web console
spring.h2.console.enabled=true
# Creates the database in a file
spring.datasource.url=jdbc:h2:file:~/multiplication;DB_CLOSE_ON_EXIT=FALSE
# Creates or updates the schema if needed
spring.jpa.hibernate.ddl-auto=update
# For educational purposes we will show the SQL in console
spring.jpa.show-sql=true

```

- As described earlier, we change the data source to use a file named `multiplication` in the user's home directory, `~`. We do that by specifying `:file:` within the URL. To learn all about the configuration possibilities you have in H2's URLs, check <http://tpd.io/h2url>.
- For simplicity, we're going to let Hibernate create our database schema for us. That feature is called an automatic data definition language (DDL). We're setting it to update because we want the schema to be both created and updated when we create or modify the entities (as we'll do in the next section).
- Last, we're enabling the property `spring.jpa.show-sql` so we see the queries in the logs. This is useful for learning purposes.

Entities

From a data perspective, JPA calls entities to the Java objects. Therefore, given that we intend to store users and attempts, we have to make the `User` and `ChallengeAttempt` classes become entities. As discussed, we could create new classes for the data layer and use mappers, but we want to keep our codebase simple, so we reuse the domain definitions.

First, let's add some JPA annotations to `User`. See Listing 5-6.

Listing 5-6. The User Class After Adding JPA Annotations

```
package microservices.book.multiplication.user;

import lombok.*;
import javax.persistence.*;

/**
 * Stores information to identify the user.
 */
@Entity
@Data
@AllArgsConstructor
```

```
@NoArgsConstructor
public class User {

    @Id
    @GeneratedValue
    private Long id;
    private String alias;

    public User(final String userAlias) {
        this(null, userAlias);
    }
}
```

Let's go through the characteristics of this updated User class, one by one:

- We added the `@Entity` annotation to mark this class as an object to be mapped to a database record. We could add a value to the annotation if we want to name our table differently from the default, `user`. Also by default, all fields exposed via getters in the class will be persisted in the mapped table with default column names. We could exclude fields by tagging them with the JPA's `@Transient` annotation.
- Hibernate's User Guide (<http://tpd.io/hib-pojos>) states that we should provide setters or make our fields modifiable by Hibernate. Luckily, Lombok has a shortcut annotation, `@Data`, which is perfect for classes that are used as data entities. This annotation groups `equals` and `hashCode` methods, `toString`, getters, and also setters. Another section in Hibernate's User Guide instructs us not to use `final` classes. This way we allow Hibernate to create runtime proxies, which improve performance. We'll see an example of how a runtime proxy works later in this chapter.
- JPA and Hibernate also require our entities to have a default, empty constructor (see <http://tpd.io/hib-constructor>). We can quickly add it with Lombok's `@NoArgsConstructor` annotation.

- Our id field is annotated with @Id and @GeneratedValue. This will be the column that uniquely identifies each row. We use a generated value so Hibernate will fill in that field for us, getting the next value of the sequence from the database.

For the ChallengeAttempt class, we're using some additional features. See Listing 5-7.

Listing 5-7. The ChallengeAttempt Class with JPA Annotations

```
package microservices.book.multiplication.challenge;

import lombok.*;
import microservices.book.multiplication.user.User;

import javax.persistence.*;

@Entity
@Data
@AllArgsConstructor
@NoArgsConstructor
public class ChallengeAttempt {
    @Id
    @GeneratedValue
    private Long id;
    @ManyToOne(fetch = FetchType.LAZY)
    @JoinColumn(name = "USER_ID")
    private User user;
    private int factorA;
    private int factorB;
    private int resultAttempt;
    private boolean correct;
}
```

Differently from the previous class, our challenge attempt model has not only basic types but also an embedded entity type, User. Hibernate knows how to map it because we added the JPA annotations, but it doesn't know the relationship between these two entities. In databases, we can model these relationships as *one-to-one*, *one-to-many*, *many-to-one*, and *many-to-many*.

We define here a many-to-one relationship since we already had a preference to avoid coupling users to attempts but to link attempts to users instead. To make these decisions in our data layer, we should also consider how we plan to query our data. In our case, we don't need the link from users to attempts. If you want to know more about entity relationships in Hibernate, check the Associations section in Hibernate User's Guide (<http://tpd.io/hib-associations>).

As you can see in the code, we're passing a parameter to the @ManyToOne annotation: the fetch type. When collecting our attempts from the data store, we have to tell Hibernate *when* to collect the values for the nested user too, which are stored in a different table. If we would set it to EAGER, the user data gets collected with the attempt. With LAZY, the query to retrieve those fields will be executed only when we try to access them. This works because Hibernate configures proxy classes for our entity classes. See Figure 5-7. These proxy classes extend ours; that's why we shouldn't declare them final if we want this mechanism to work. For our case, Hibernate will pass a proxy object that triggers the query to fetch the user only when the accessor (getter) is used for the first time. That's where the *laziness* term comes from—it doesn't do that until the very last moment.

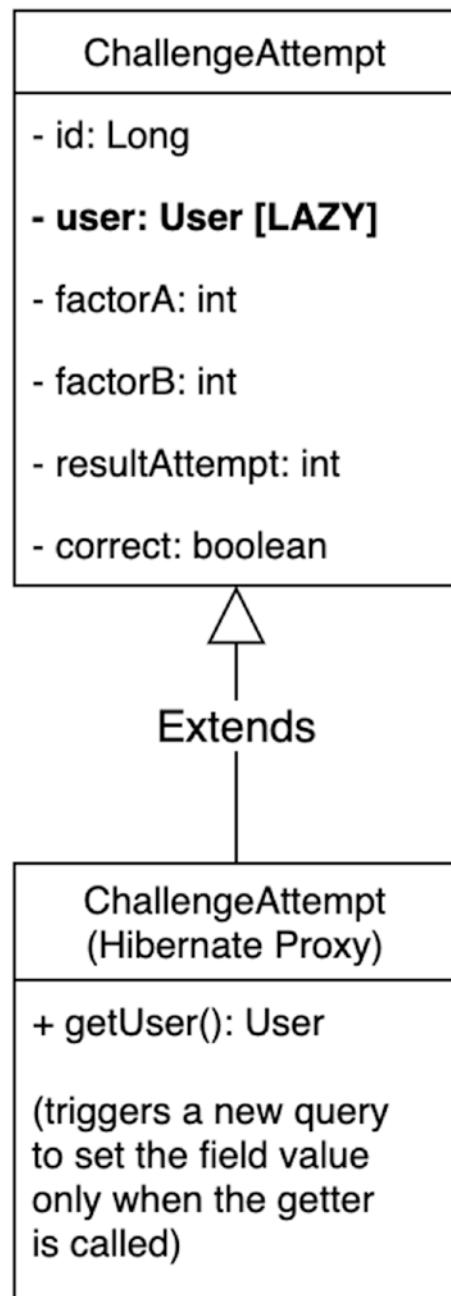


Figure 5-7. Hibernate, intercepting classes

In general, we should prefer lazy associations to avoid triggering extra queries for data you may not need. In our example, we don't need the user's data when we collect the attempts.

The `@JoinColumn` annotation makes Hibernate link both tables with a join column. For consistency, we're passing it the same name of the column that represents the index for users. This will translate to a new column added to the `CHALLENGE_ATTEMPT` table, `USER_ID`, which will store the reference to the ID record of the corresponding user in the `USER` table.

This is a basic yet representative example of ORM with JPA and Hibernate. If you want to extend your knowledge about all the possibilities you have with JPA and Hibernate, the User Guide (<http://tpd.io/hib-user-guide>) is a good place to start.

Consequences of Reusing Domain Objects as Entities

Because of JPA and Hibernate's requirements, we needed to add setters to our classes and an ugly empty constructor (Lombok hid it, but it's still there). This is inconvenient since it prevents us from creating classes following good practices such as immutability. We could argue that our domain classes became corrupted by the data requirements.

This is not a big problem when you are building small applications and you know the reasons behind these decisions. You simply avoid using the setters or the empty constructor in your code. However, when working with a big team or in a mid- or big-size project, this may become an issue since a new developer might be tempted to break good practices just because the class allows them to do so. In that scenario, you may consider splitting the domain and the entity classes as mentioned earlier. That'll bring some code duplication, but you can then better enforce good practices.

Repositories

When we described the three-layer architecture, we briefly explained that the data layer may contain data access objects (DAOs) and repositories. DAOs are typically classes that are coupled to the database structure, whereas repositories, on the other hand, are domain-centric, so these classes can work with aggregates.

Given that we're following domain-driven design, we'll use repositories to connect to the database. More specifically, we'll use JPA repositories and the features included in Spring Data JPA.

In the previous section about the technology stack, we introduced the Spring's `SimpleJpaRepository` class (see <https://tpd.io/sjparepo-doc>), which uses JPA's `EntityManager` (see <https://tpd.io/em-javadoc>) to manage our database objects. The Spring abstraction adds some features such as pagination and sorting, and some methods that make it more convenient to use than the plain JPA interface (e.g., `saveAll`, `existsById`, `count`, etc.).

Spring Data JPA also comes with a superpower not offered by plain JPA: the query methods (see <http://tpd.io/jpa-query-methods>).

Let's use our codebase to demonstrate this functionality. We need a query to get the last attempts for a given user so we can display the stats on the web page. Apart from that, we need some basic entity management to create, read, and delete attempts. The interface shown in Listing 5-8 provides that functionality.

Listing 5-8. The ChallengeAttemptRepository Interface

```
package microservices.book.multiplication.challenge;

import org.springframework.data.repository.CrudRepository;
import java.util.List;

public interface ChallengeAttemptRepository extends CrudRepository<ChallengeAttempt, Long> {

    /**
     * @return the last 10 attempts for a given user, identified by their alias.
     */
    List<ChallengeAttempt> findTop10ByUserAliasOrderByIdDesc(String userAlias);
}
```

We create the interface extending the `CrudRepository` interface (<http://tpd.io/crud-repo>) in Spring Data Commons. `CrudRepository` defines a list of basic methods to create, read, update, and delete (CRUD) objects. The `SimpleJpaRepository` class in Spring Data JPA implements this interface too (<http://tpd.io/simple-jpa-repo>). Apart from `CrudRepository`, there are two other alternatives we could use.

- If we choose to extend the plain `Repository`, we don't get CRUD functionality. However, that interface works as a marker when we want to fine-tune the methods we want to expose from `CrudRepository`, instead of getting them all by default. See <http://tpd.io/repo-tuning> to learn more about this technique.
- If we need also pagination and sorting, we could extend `PagingAndSortingRepository`. This is helpful if we have to deal with big collections that are better queried in chunks, or *pages*.

When we extend any of these three interfaces, we have to use Java generics, as we did in this line:

```
... extends CrudRepository<ChallengeAttempt, Long> {
```

The first type specifies what is the class of the returned entity, `ChallengeAttempt` in our case. The second class must match the type of the index, which is a `Long` in our repository (the `id` field).

The most striking part of our code is the method name we added to the interface. In Spring Data, we can create methods that define queries by using naming conventions in the method name. In this particular case, we want to query attempts by user alias, order them by `id` descending (the newest first), and pick the top 10 of the list. Following the method structure, we could describe the query as follows: find Top 10 (any matching `ChallengeAttempt`) by (field `userAlias` equals to passed argument) order by (field `id`) descending.

Spring Data will process the methods you define in your interface looking for those that don't have an explicit query defined and match the naming convention for creating query methods. That's exactly our case. Then, it parses the method name, decomposes it in chunks, and builds a JPA query that corresponds with that definition (keep reading for an example query).

We can build many other queries using the JPA query method definition; see <http://tpd.io/jpa-qm-create> for details.

Sometimes we may want to perform some queries that can't be achieved with a query method. Or maybe we just don't feel comfortable using this feature because the method names start getting a bit weird. No worries, it's also possible to define our own queries. In this case, we can still keep our implementation abstracted from the database engine by writing the queries in Java Persistence Query Language (JPQL), a SQL language that is also part of the JPA standard. See Listing 5-9.

Listing 5-9. A Defined Query as an Alternative to a Query Method

```
/**  
 * @return the last attempts for a given user, identified by their alias.  
 */  
@Query("SELECT a FROM ChallengeAttempt a WHERE a.user.alias = ?1 ORDER BY a.id DESC")  
List<ChallengeAttempt> lastAttempts(String userAlias);
```

As you can see, it looks like standard SQL. The following are the differences:

- We don't use a table name, but the class name instead (`ChallengeAttempt`).
- We refer to fields not as columns but as object fields, using dots to traverse the object structure (`a.user.alias`).
- We can use argument placeholders, like `?1` in our example to refer to the first (and only) passed argument.

We'll stick to the query method since it's shorter and descriptive enough, but we'll need to write JPQL queries soon for other requirements we have.

That was all we need to manage attempt entities in the database. Now, we're missing the repository to manage the User entities. This one is straightforward to implement, as shown in Listing 5-10.

Listing 5-10. The UserRepository Interface

```
package microservices.book.multiplication.user;

import org.springframework.data.repository.CrudRepository;
import java.util.Optional;

public interface UserRepository extends CrudRepository<User, Long> {

    Optional<User> findByAlias(final String alias);

}
```

The `findByAlias` query method will return a user wrapped in a Java `Optional` if there was a match, or an empty `Optional` object if no user is matching the passed alias. This is another feature provided by Spring Data's JPA query methods.

With these two repositories, we have everything we need to manage our database entities. We don't need to implement these interfaces. We don't even need to add the Spring's `@Repository` annotation. Spring, using the Data module, will find interfaces extending the base ones and will inject beans that implement the desired behavior. That also involves processing the method names and creating the corresponding JPA queries.

Storing Users and Attempts

After finishing the data layer, we can start using the repositories from our service layer.

First, let's extend our test cases with the new expected logic:

- The attempt should be stored, no matter if it was correct or not.
- If it's the first attempt for a given user, identified by their alias, we should create the user. If the alias exists, the attempt should be linked to that existing user.

We have to make some updates to our `ChallengeServiceTest` class. To start with, we need to add two mocks for both repositories. This way, we keep the unit test focused on the service layer, without including any real behavior from other layers. As introduced in Chapter 2, this is one of the advantages of Mockito.

To use mocks with Mockito, we can annotate the fields with the `@Mock` annotation and add the `MockitoExtension` to the test class to have them initialized automatically. With this extension, we also get other Mockito features like the detection of unused stubs, which makes our tests fail if we specify a mock behavior that we don't use it during the test case. See Listing 5-11.

Listing 5-11. Using Mockito in the `ChallengeServiceTest` Class

```
@ExtendWith(MockitoExtension.class)
public class ChallengeServiceTest {

    private ChallengeService challengeService;

    @Mock
    private UserRepository userRepository;
    @Mock
    private ChallengeAttemptRepository attemptRepository;

    @BeforeEach
    public void setUp() {
        challengeService = new ChallengeServiceImpl(
            userRepository,
            attemptRepository
        );
        given(attemptRepository.save(any()))
            .will(returnsFirstArg());
    }

    //...
}
```

Besides, we can use the method annotated with JUnit's `@BeforeEach` to add some common behavior to all our tests. In this case, we use the service's constructor to include the repositories (note that this constructor doesn't exist yet). We added this line too:

```
given(attemptRepository.save(any()))
    .will(returnsFirstArg());
```

This instruction uses BDDMockito's `given` method to define what the mock class should do when we call specific methods during the test. Remember that we don't want to use the real class's functionality, so we have to define, for example, what to return when invoking

functions on this fake object (or *stub*). The method we want to override is passed as an argument: `attemptRepository.save(any())`. We could match a specific argument passed to `save()`, but we can also define this predefined behavior for any argument by using `any()` from Mockito's *argument matchers* (check <https://tpd.io/mock-am> for the complete list of matchers). The second part of the instruction, using `will()`, specifies what Mockito should do when the previously defined condition matches. The `returnsFirstArg()` utility method is defined in Mockito's `AdditionalAnswers` class, which includes some convenient predefined answers we can use (see <http://tpd.io/mockito-answers>). You could also declare your own functions to provide custom answers if you need to implement more complex scenarios. In our case, we want the `save` method to do nothing but return the first (and only) argument passed. That's good enough for us to test this layer without calling the real repository.

Now we add the extra verifications to our existing test cases. See Listing 5-12, which includes the correct attempt's test case as an example.

Listing 5-12. Verifying Stub Calls in ChallengeServiceTest

```
@Test
public void checkCorrectAttemptTest() {
    // given
    ChallengeAttemptDTO attemptDTO =
        new ChallengeAttemptDTO(50, 60, "john_doe", 3000);

    // when
    ChallengeAttempt resultAttempt =
        challengeService.verifyAttempt(attemptDTO);

    // then
    then(resultAttempt.isCorrect()).isTrue();
    // newly added lines
    verify(userRepository).save(new User("john_doe"));
    verify(attemptRepository).save(resultAttempt);
}
```

We use Mockito's `verify` to check that we store a new user with a null ID and the expected alias. The identifier will be set at the database level. We also verify that the attempt should be saved. The test case that verifies a wrong attempt should contain those two new lines as well.

To make our tests more complete, we add a new case that verifies that extra attempts from the same user won't create new user entities but reuse the existing one. See Listing 5-13.

Listing 5-13. Verifying That Only the First Attempt Creates the User Entity

```
@Test
public void checkExistingUserTest() {
    // given
    User existingUser = new User(1L, "john_doe");
    given(userRepository.findByAlias("john_doe"))
        .willReturn(Optional.of(existingUser));
    ChallengeAttemptDTO attemptDTO =
        new ChallengeAttemptDTO(50, 60, "john_doe", 5000);

    // when
    ChallengeAttempt resultAttempt =
        challengeService.verifyAttempt(attemptDTO);

    // then
    then(resultAttempt.isCorrect()).isFalse();
    then(resultAttempt.getUser()).isEqualTo(existingUser);
    verify(userRepository, never()).save(any());
    verify(attemptRepository).save(resultAttempt);
}
```

In this case, we define the behavior of the `userRepository` mock to return an existing user. Since the challenge DTO contains the same alias, the logic should find our predefined user, and the returned attempt must include it, with the same alias and ID. To make the test more exhaustive, we check that the method `save()` in `UserRepository` is never called.

At this point, we have a test that doesn't compile. Our service should provide a constructor with both repositories. When we start the application, Spring will use dependency injection via the constructor to initialize the repositories. This is how Spring helps us keep our layers loosely coupled.

Then, we also need the main logic to store both the attempt and the user (if it doesn't exist yet). See Listing 5-14 for the new implementation of `ChallengeServiceImpl`.

Listing 5-14. The Updated ChallengeServiceImpl Class Using the Repository Layer

```

package microservices.book.multiplication.challenge;

import lombok.RequiredArgsConstructor;
import lombok.extern.slf4j.Slf4j;
import microservices.book.multiplication.user.User;
import microservices.book.multiplication.user.UserRepository;
import org.springframework.stereotype.Service;

@Slf4j
@RequiredArgsConstructor
@Service
public class ChallengeServiceImpl implements ChallengeService {

    private final UserRepository userRepository;
    private final ChallengeAttemptRepository attemptRepository;

    @Override
    public ChallengeAttempt verifyAttempt(ChallengeAttemptDTO attemptDTO) {
        // Check if the user already exists for that alias, otherwise create it
        User user = userRepository.findByAlias(attemptDTO.getUserAlias())
            .orElseGet(() -> {
                log.info("Creating new user with alias {}", attemptDTO.getUserAlias());
                return userRepository.save(
                    new User(attemptDTO.getUserAlias())
                );
            });

        // Check if the attempt is correct
        boolean isCorrect = attemptDTO.getGuess() ==
            attemptDTO.getFactorA() * attemptDTO.getFactorB();

        // Builds the domain object. Null id since it'll be generated by the DB.
        ChallengeAttempt checkedAttempt = new ChallengeAttempt(null,
            user,
            attemptDTO.getFactorA(),
            attemptDTO.getFactorB(),
    }
}

```

```

        attemptDTO.getGuess(),
        isCorrect
    );

    // Stores the attempt
    ChallengeAttempt storedAttempt = attemptRepository.save(checkedAttempt);

    return storedAttempt;
}
}
}

```

The first block inside `verifyAttempt` uses the `Optional` returned by the repository to decide whether the user should be created. The method `orElseGet` in an `Optional` invokes the passed function only if it's empty. Therefore, we create a new user only if it's not there yet.

We pass the returned `User` object from the repository when we construct an attempt. Hibernate will take care of linking them properly in the database when we call `save()` to store the attempt entity. We return the result, so it includes all the identifiers from the database.

All the test cases should pass now. Again, we used TDD to create our logic based on our expectations. It's clear now how a unit test helps us verify the specific layer's behavior without depending on the other layers. For our service class, we replaced both repositories by stubs for which we defined preset values.

There is an alternative implementation of these tests. We could use the `@SpringBootTest` flavor and `@MockBean` for the repository classes. However, that doesn't bring any added value and requires the Spring context, so the tests take more time to finish. As we said in a previous chapter, we prefer to keep our unit tests as simple as possible.

Repository Tests

We're not creating tests for the application's data layer. These tests don't make much sense since we're not writing any implementation anyway. We would end up verifying the Spring Data implementation itself.

Displaying Last Attempts

We modified our existing service logic to store users and attempts, but we're still missing the other half of the functionality: retrieving the last attempts and displaying them on the page.

The service layer can simply use the query method from the repository. On the controller layer, we'll expose a new REST endpoint to get the list of attempts by user alias.

Exercise

Keep following TDD and complete some tasks before moving forward with the implementation. You'll find the solution in this chapter's code repository (<https://github.com/Book-Microservices-v2/chapter05>).

- Extend the `ChallengeServiceTest` and create a test case to verify we can retrieve the last attempts. The logic behind the test is a one-liner, but it's good to have the test in case the service layer grows. Note that you may get complaints from Mockito about the unnecessary stub of the `save` method in this test case. That's one of the features of the `MockitoExtension`. You can then move that stub inside the test cases that use it.
 - Update the `ChallengeAttemptController` class to include a test for the new endpoint: GET `/attempts?alias=john_doe`.
-

Service Layer

Let's add a method to the `ChallengeService` interface called `getStatsForUser`. See Listing 5-15.

Listing 5-15. Adding the getStatsForUser Method to the ChallengeService Interface

```
package microservices.book.multiplication.challenge;

import java.util.List;

public interface ChallengeService {

    /**
     * Verifies if an attempt coming from the presentation layer is correct or
     * not.
     *
     * @return the resulting ChallengeAttempt object
     */
    ChallengeAttempt verifyAttempt(ChallengeAttemptDTO attemptDTO);

    /**
     * Gets the statistics for a given user.
     *
     * @param userAlias the user's alias
     * @return a list of the last 10 {@link ChallengeAttempt}
     * objects created by the user.
     */
    List<ChallengeAttempt> getStatsForUser(String userAlias);
}
```

The code block in Listing 5-16 shows the implementation. As predicted, it's just a single line of code.

Listing 5-16. Implementing the getStatsForUser Method

```
@Slf4j
@RequiredArgsConstructor
@Service
public class ChallengeServiceImpl implements ChallengeService {
    // ...
}
```

```

@Override
public List<ChallengeAttempt> getStatsForUser(final String userAlias) {
    return attemptRepository.findTop10ByUserAliasOrderByIdDesc(userAlias);
}
}

```

Controller Layer

Let's move one layer up and see how we connect the service layer from the controller. This time, we make use of a query parameter, but that doesn't add much complexity to our API definitions. Similarly, as we got the request body injected as a parameter in our first method, we can now use `@RequestParam` to tell Spring to pass us a URL parameter. Check the reference docs (<http://tpd.io/mvc-ann>) for other method arguments you can define (e.g., session attributes or cookie values). See Listing 5-17.

Listing 5-17. Adding New Endpoint in the Controller to Retrieve Statistics

```

@Slf4j
@RequiredArgsConstructor
@RestController
@RequestMapping("/attempts")
class ChallengeAttemptController {

    private final ChallengeService challengeService;

    @PostMapping
    ResponseEntity<ChallengeAttempt> postResult(
        @RequestBody @Valid ChallengeAttemptDTO challengeAttemptDTO) {
        return ResponseEntity.ok(challengeService.verifyAttempt(challenge
            AttemptDTO));
    }

    @GetMapping
    ResponseEntity<List<ChallengeAttempt>> getStatistics(@RequestParam("alias")
        String alias) {
        return ResponseEntity.ok(

```

```

        challengeService.getStatsForUser(alias)
    );
}
}
}
```

If you implemented the tests, they should pass now. However, if we run a quick test using HTTPie, we find an unexpected result. See Listing 5-18. Sending one attempt and then trying to retrieve the list gives us an error.

Listing 5-18. Error During Serialization of the Attempt List

```

$ http POST :8080/attempts factorA=58 factorB=92 userAlias=moises guess=5303
HTTP/1.1 200
...
$ http ":8080/attempts?alias=moises"
HTTP/1.1 500
...
{
    "error": "Internal Server Error",
    "message": "Type definition error: [simple type, class org.hibernate.proxy.pojo.bytebuddy.ByteBuddyInterceptor]; nested exception is com.fasterxml.jackson.databind.exc.InvalidDefinitionException: No serializer found for class org.hibernate.proxy.pojo.bytebuddy.ByteBuddyInterceptor and no properties discovered to create BeanSerializer (to avoid exception, disable SerializationFeature.FAIL_ON_EMPTY_BEANS) (through reference chain: java.util.ArrayList[0]->microservices.book.multiplication.challenge.ChallengeAttempt[\"user\"]->microservices.book.multiplication.user.User$HibernateProxy$mk4Fwvp[\"hibernateLazyInitializer\"])",
    "path": "/attempts",
    "status": 500,
    "timestamp": "2020-04-15T05:41:53.993+0000"
}
```

That's an ugly server error. We can also find the counterpart exception in our backend logs. What is a ByteBuddyInterceptor, and why is our ObjectMapper trying to serialize it? There should be only ChallengeAttempt objects in the result, with nested User instances, right? Well, not really.

We configured our nested User entities to be fetched in LAZY mode, so they're not being queried from the database. We also said that Hibernate creates proxies for our classes in runtime. That's the reason behind the ByteBuddyInterceptor class. You can try switching the fetch mode to EAGER, and you will no longer get this error. But that's not the proper solution to this problem since then we'll be triggering many queries for data we don't need.

Let's keep the lazy fetch mode and fix this accordingly. The first option we have is to customize our JSON serialization so it can handle Hibernate objects. Luckily, FasterXML, the provider of Jackson libraries, has a specific module for Hibernate that we can use in our ObjectMapper objects: jackson-datatype-hibernate (<http://tpd.io/json-hib>). To use it, we have to add this dependency to our project since it's not included by Spring Boot starters. See Listing 5-19.

Listing 5-19. Adding the Jackson Module for Hibernate to Our Dependencies

```
<dependencies>
<!-- ... -->
<dependency>
  <groupId>com.fasterxml.jackson.datatype</groupId>
  <artifactId>jackson-datatype-hibernate5</artifactId>
</dependency>
<!-- ... -->
</dependencies>
```

Then we follow the documented way in Spring Boot (see <http://tpd.io/om-custom>) to customize ObjectMappers:

“Any beans of type com.fasterxml.jackson.databind.Module are automatically registered with the auto-configured Jackson2ObjectMapperBuilder and are applied to any ObjectMapper instances that it creates. This provides a global mechanism for contributing custom modules when you add new features to your application.”

We create a bean for our new Hibernate module for Jackson. Spring Boot's `JacksonObjectMapperBuilder` will use it via autoconfiguration, and all our `ObjectMapper` instances will use the Spring Boot defaults plus our own customization. See Listing 5-20 showing this new `JsonConfiguration` class.

Listing 5-20. Loading the Jackson's Hibernate Module to Be Picked Up by Auto-Configuration

```
package microservices.book.multiplication.configuration;

import com.fasterxml.jackson.databind.Module;
import com.fasterxml.jackson.datatype.hibernate5.Hibernate5Module;
import org.springframework.context.annotation.Bean;
import org.springframework.context.annotation.Configuration;

@Configuration
public class JsonConfiguration {

    @Bean
    public Module hibernateModule() {
        return new Hibernate5Module();
    }
}
```

Now we start our application, and we verify that we can retrieve the attempts successfully. The nested `user` object is `null`, which is perfect since we don't need it for the list of attempts. See Listing 5-21. We avoided the extra queries.

Listing 5-21. Correct Serialization of the Attempts After Adding the Hibernate Module

```
$ http ":8080/attempts?alias=moises"
HTTP/1.1 200
...
[
  {
    "correct": false,
    "factorA": 58,
    "factorB": 92,
    "id": 11,
```

```

    "resultAttempt": 5303,
    "user": null
},
...
]

```

An alternative to adding this new dependency and the new configuration is to follow the recommendation that was printed in the message of the exception that we got:

```
...(to avoid exception, disable SerializationFeature.FAIL_ON_EMPTY_BEANS)[...]
```

Let's try it. We can add features to Jackson serializers directly in the application properties file (see <http://tpd.io/om-custom>). This is achieved with some naming convention, prefixing the Jackson properties with `spring.jackson.serialization`. See Listing 5-22.

Listing 5-22. Adding a Property to Avoid Serialization Errors on Empty Beans

```
[...]
spring.jpa.show-sql=true
spring.jackson.serialization.fail_on_empty_beans=false
```

If you try this (after removing the code from the previous solution) and then collect the attempts, you'll find a funny result. See Listing 5-23.

Listing 5-23. Retrieving Attempts with `fail_on_empty_beans=false`

```
$ http ":8080/attempts?alias=moises"
```

```
HTTP/1.1 200
```

```
...
```

```
[
```

```
{
    "correct": false,
    "factorA": 58,
    "factorB": 92,
    "id": 11,
    "resultAttempt": 5303,
    "user": {
        "alias": "moises",
```

```

    "hibernateLazyInitializer": {},
    "id": 1
  },
},
...
]

```

There are two unexpected outcomes. First, the property `hibernateLazyInitializer` from the proxy object is being serialized to JSON, and it's empty. That's the empty bean, and it's actually the source of the error we got earlier. We could avoid that with some Jackson configuration to ignore that field. But the real issue is that the user's data is there too. The serializer traversed the proxy to get the user's data, and that triggered the extra query from Hibernate to fetch it, which makes our lazy parameter configuration useless. We can also verify that in the logs, where we got an extra query compared to the previous solution. See Listing 5-24.

Listing 5-24. Unwanted Query When Fetching Attempts with Suboptimal Configuration

```

Hibernate: select challengea0_.id as id1_0_, challengea0_.correct as correct2_0_,
challengea0_.factora as factora3_0_, challengea0_.factorb as factorb4_0_,
challengea0_.result_attempt as result_a5_0_, challengea0_.user_id as user_id6_0_
from challenge_attempt challengea0_ left outer join user user1_ on challengea0_.
user_id=user1_.id where user1_.alias=? order by challengea0_.id desc limit ?
Hibernate: select user0_.id as id1_1_0_, user0_.alias as alias2_1_0_ from user
user0_ where user0_.id=?

```

We'll hold to the first option with the Hibernate module for Jackson since it's the proper way to handle lazy fetching with JSON serialization.

The takeaway from the analysis we did of both alternatives is that, with so much behavior hidden behind the scenes in Spring Boot, you should avoid going for quick solutions without really understanding what the implications are. Get to know the tools and read the reference documentation.

User Interface

The last part in our stack where we need to integrate the new functionality to display the last attempts is in our React front end. Like in the previous chapter, you can skip this section if you don't want to dive into details about the UI.

Let's stick to a basic user interface for now and add a table to our page that will show the last tries of the user. We can make this request after a new attempt is sent since we'll get the user's alias.

But, before that, let's replace the predefined CSS to make sure all contents fit on the page.

First, we move the `ChallengeComponent` to be rendered directly, without any wrapper. See the resulting `App.js` file in Listing 5-25.

Listing 5-25. App.js File After Moving the Component Up

```
import React from 'react';
import './App.css';
import ChallengeComponent from './components/ChallengeComponent';

function App() {
  return <ChallengeComponent/>;
}

export default App;
```

Then, we remove all the predefined CSS and adapt it to our needs. We can add these basic styles to the `index.css` and `App.css` files, respectively. See Listings 5-26 and 5-27.

Listing 5-26. The Modified index.css File

```
body {
  font-family: 'Segoe UI', Roboto, Arial, sans-serif;
}
```

Listing 5-27. The Modified app.css File

```
.display-column {
  display: flex;
  flex-direction: column;
  align-items: center;
}
```

```
.challenge {
  font-size: 4em;
}

th {
  padding-right: 0.5em;
  border-bottom: solid 1px;
}
```

We'll apply the `display-column` to the main HTML container to stack our components vertically and align them to the center. The `challenge` style is for the multiplication, and we also customize the table header style to have some padding and use a bottom line.

Once we've made some room for the new table, we have to extend our `ApiClient` in JavaScript to retrieve the attempts. Like before, we use `fetch` with its default GET verb and build the URL to include the user's alias as a query parameter. See Listing 5-28.

Listing 5-28. `ApiClient` Class Update with the Method to Fetch Attempts

```
class ApiClient {

  static SERVER_URL = 'http://localhost:8080';
  static GET_CHALLENGE = '/challenges/random';
  static POST_RESULT = '/attempts';
  static GET_ATTEMPTS_BY_ALIAS = '/attempts?alias=';

  static challenge(): Promise<Response> {
    return fetch(ApiClient.SERVER_URL + ApiClient.GET_CHALLENGE);
  }

  static sendGuess(user: string,
                  a: number,
                  b: number,
                  guess: number): Promise<Response> {
    // ...
  }

  static getAttempts(userAlias: string): Promise<Response> {
    return fetch(ApiClient.SERVER_URL +
      ApiClient.GET_ATTEMPTS_BY_ALIAS + userAlias);
  }
}

export default ApiClient;
```

Our next task is to create a new React component for this list of attempts. This way, we keep our front end modularized. This new component doesn't need to have a state since we will use the parent's one to hold the last attempts.

We use a simple HTML table to render the objects passed through the props object. As a nice addition on the UI level, we'll show the correct result of the challenge if it was incorrect. Also, we'll have a conditional style attribute that will make the text color green or red depending on whether the attempt was correct or not. See Listing 5-29.

Listing 5-29. The New LastAttemptsComponent in React

```
import * as React from 'react';

class LastAttemptsComponent extends React.Component {

  render() {
    return (
      <table>
        <thead>
          <tr>
            <th>Challenge</th>
            <th>Your guess</th>
            <th>Correct</th>
          </tr>
        </thead>
        <tbody>
          {this.props.lastAttempts.map(a =>
            <tr key={a.id}>
              style={{ color: a.correct ? 'green' : 'red' }}>
              <td>{a.factorA} x {a.factorB}</td>
              <td>{a.resultAttempt}</td>
              <td>{a.correct ? "Correct" :
                ("Incorrect (" + a.factorA * a.factorB + ")")}</td>
            </tr>
          )}
        </tbody>
      </table>
    );
  }
}

export default LastAttemptsComponent;
```

As shown in the code, we can use `map` when rendering React components to easily iterate over arrays. Each element of the array should use a `key` attribute to help the framework identify changing elements. See <http://tpd.io/react-keys> for more details about rendering lists with unique keys.

Now we need to put everything to work together in our existing `ChallengeComponent` class. See Listing 5-30 for the code after adding some modifications.

- A new function that uses `ApiClient` to retrieve the last attempts, check if the HTTP response was OK, and store the array in the state.
- A call to this new function right after we get a response to the request that sends a new attempt.
- The component's HTML tag inside the `render()` function of this parent component.
- As an improvement, we also extract the logic to refresh the challenge (included before in `componentDidMount`) to a new function, `refreshChallenge`. We'll create a new challenge for the users after they send an attempt.

Listing 5-30. The Updated `ChallengeComponent` to Include the `LastAttemptsComponent`

```
import * as React from "react";
import ApiClient from "../services/ApiClient";
import LastAttemptsComponent from './LastAttemptsComponent';

class ChallengeComponent extends React.Component {
  constructor(props) {
    super(props);
    this.state = {
      a: '',
      b: '',
      user: '',
      message: '',
      guess: 0,
      lastAttempts: []
    };
  }
}
```

```

    this.handleSubmitResult = this.handleSubmitResult.bind(this);
    this.handleChange = this.handleChange.bind(this);
}

// ...

handleSubmitResult(event) {
    event.preventDefault();
    ApiClient.sendGuess(this.state.user,
        this.state.a, this.state.b,
        this.state.guess)
    .then(res => {
        if (res.ok) {
            res.json().then(json => {
                if (json.correct) {
                    this.updateMessage("Congratulations! Your guess is
                        correct");
                } else {
                    this.updateMessage("Oops! Your guess " + json.
                        resultAttempt +
                        " is wrong, but keep playing!");
                }
                this.updateLastAttempts(this.state.user); // NEW!
                this.refreshChallenge(); // NEW!
            });
        } else {
            this.updateMessage("Error: server error or not available");
        }
    });
}

// ...

updateLastAttempts(userAlias: string) {
    ApiClient.getAttempts(userAlias).then(res => {
        if (res.ok) {
            let attempts: Attempt[] = [];
            res.json().then(data => {
                data.forEach(item => {

```

```

        attempts.push(item);
    });
    this.setState({
        lastAttempts: attempts
    });
}
}

render() {
    return (
        <div className="display-column">
            <div>
                <h3>Your new challenge is</h3>
                <div className="challenge">
                    {this.state.a} x {this.state.b}
                </div>
            </div>
            <form onSubmit={this.handleSubmitResult}>
                {/* ... */}
            </form>
            <h4>{this.state.message}</h4>
            {this.state.lastAttempts.length > 0 &&
                <LastAttemptsComponent lastAttempts={this.state.
                    lastAttempts}/>
            }
        </div>
    );
}

export default ChallengeComponent;

```

In React, we can update parts of the state if we pass only that property to the `setState` method, which will then merge the contents. We add the new property `lastAttempts` to the state and update it with the contents of the array returned by the back end. Given that the items of the array are JSON objects, we can access its attributes in plain JavaScript using the property names.

We also used something new in this code: a conditional rendering with the `&&` operator. Only if the condition on the left is true, the content on the right will be rendered by React. See <http://tpd.io/react-inline-if> for different ways of doing this. The `lastAttempts` HTML property added to the component tag is passed to the child component's code via the `props` object.

Note that we also used the new styles `display-column` and `challenge`. In React, we use the `className` attribute, which will be mapped to the standard HTML `class`.

Playing with the New Feature

After adding the new data layer and the logic in all other layers up to the UI, we're ready to play with the full application. Either using the IDE or using two different terminal windows, we run both the back end and the front end with the commands shown in Listing 5-31.

Listing 5-31. Commands to Start the Back End and the Front End, Respectively

```
/multiplication $ mvnw spring-boot:run  
...  
/challenges-frontend $ npm start  
...
```

Then, we navigate to `http://localhost:3000` to access the React front end running in development mode. Since the rendering of our new component is conditional, we don't see yet the new table, so let's play a few challenges and see how the table gets populated with our attempts. You should see something similar to Figure 5-8.

Your new challenge is

$$43 \times 70$$

Your alias:

Your guess:

Oops! Your guess 1564 is wrong, but keep playing!

Challenge	Your guess	Correct
36 x 44	1564	Incorrect (1584)
15 x 85	1275	Correct
27 x 18	486	Correct
78 x 71	1488	Incorrect (5538)

Figure 5-8. Our app after adding the Last Attempts feature

Great, we made it! It's not the most beautiful front end, but it's nice to see our new functionality up and running. The Challenge component performs the requests to the back end and renders the child component, the Last Attempts table.

If you're curious about how the data looks, we can also navigate to the back end's H2 console to get access to the data in our tables. Remember that the console is located at <http://localhost:8080/h2-console>. You should see both tables for users and attempts and some contents in them. This basic console allows you to perform queries and even edit the data. For example, you can click the name of the table CHALLENGE_ATTEMPT, and a SQL query will be generated for you in the panel to the right. Then, you can click the button Run to query the data. See Figure 5-9.

The screenshot shows the H2 console interface. On the left, there's a sidebar with database connections (H2), schema navigation (CHALLENGE_ATTEMPT, USER, INFORMATION_SCHEMA, Sequences, Users), and a note about version H2 1.4.200 (2019-10-14). The main area has tabs for Run, Run Selected, Auto complete, Clear, and SQL statement: "SELECT * FROM CHALLENGE_ATTEMPT". Below this is a table titled "SELECT * FROM CHALLENGE_ATTEMPT;" with columns ID, CORRECT, FACTORA, FACTORB, RESULT_ATTEMPT, and USER_ID. The table contains 13 rows of data.

ID	CORRECT	FACTORA	FACTORB	RESULT_ATTEMPT	USER_ID
21	FALSE	16	35	2323	1
22	TRUE	16	35	560	1
23	FALSE	72	49	230	1
24	FALSE	72	21	2323	1
25	FALSE	35	76	2323	12
26	FALSE	98	29	3239	12
27	FALSE	47	88	2323	12
28	FALSE	84	84	2390	12
29	FALSE	89	77	2837	12
31	FALSE	92	12	3290	30
33	TRUE	70	21	1470	32
34	TRUE	52	60	3120	32
35	FALSE	76	92	1682	32

(13 rows, 2 ms)

Figure 5-9. Attempts data in H2 console

Summary and Achievements

In this chapter, we saw how to model our data for persistence, and we used object-relational mapping to convert our domain object to database records. During the journey, we covered some Hibernate and JPA fundamentals. See Figure 5-10 with the current status of our application.

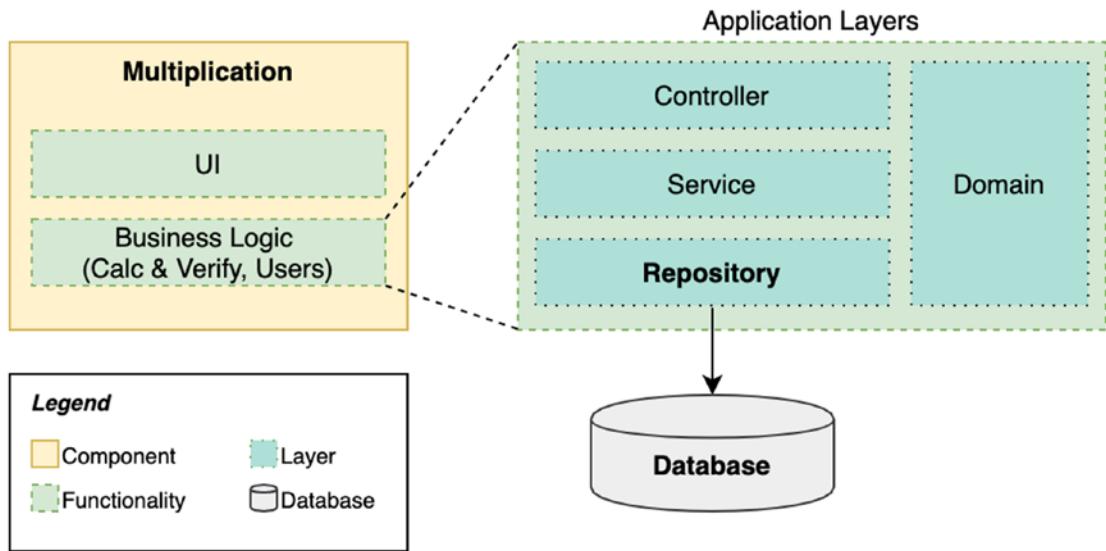


Figure 5-10. Application after Chapter 5

We learned how to use JPA annotations to map our Java classes and how simple associations work, based on our many-to-one use case between challenges and attempts. Also, we used Spring Data repositories to get a lot of out-of-the-box functionality. Among all the offered features, we saw how query methods are a powerful way to write simple queries with some method's naming conventions.

The second user story is done. We went through the service and the controller layers and added our new functionality there. We also included a new component in the UI to visualize the last attempts on our web application.

We finished the first part of the book. Until now, we have seen in detail how a small web application works. We dedicated time to understand Spring Boot's core concepts and some specific modules for the different layers, like Spring Data and Spring Web. We even built a small React front end.

Now it's time to start the microservices adventure.

Chapter's Achievements:

- You got the complete picture of how a three-layer, three-tier architecture works, with the introduction of the repository classes and the database.
- You learned how to model your data, taking into account your requirements for querying it and proper domain isolation.

CHAPTER 5 THE DATA LAYER

- You went through the main differences between SQL and NoSQL and the criteria you can use to make future choices.
- You learned about JPA and its integration with Hibernate, Spring Data, and Spring Boot.
- You developed a real persistence layer for an application using JPA annotations and Spring Data repositories, using query methods and defined queries.
- You integrated the new attempts history feature all the way up to the front end, improving our practical case study.

CHAPTER 6

Starting with Microservices

This chapter begins with a pause in our practical journey. It's important that we take the time to analyze how we built our application up until now and the impact of our future move to microservices.

First, we'll see what the advantages of starting with a single codebase are. Then, we'll describe our new requirements, including a short summary of the basic gamification techniques. After that, we'll see what factors we need to take into account when moving to microservices, as well as the pros and cons we should consider to make that decision.

The second part of this chapter is all practical again. We'll build our new microservice using the design patterns we learned in previous chapters, and we'll connect it to our existing application, the first microservice. Then, we'll analyze some new challenges we'll face because of our new *distributed system*.

The Small Monolith Approach

The previous chapter ended with a single deployment unit that contained all the required functionalities (even the front end if we'd like to include it). This single-application strategy has been our choice despite having identified already two domains: challenges and users. Their domain objects have relationships, but they are loosely coupled. Yet we decided not to split them into multiple applications or microservices from the beginning.

We could consider our Multiplication application a small monolith.

Why a Small Monolith?

Compared to microservices, starting with a single codebase simplifies the development process, so we reduce the time we need to deploy the first version of our product. Besides, it makes changing our architecture and software designs much easier at the beginning of the project's lifecycle, when it's critical to make adaptations after we validated our ideas.

If you haven't worked yet in an organization using microservices, you might underestimate the dimension of the technical complexity microservices introduce in a software project. Ideally, you'll have a clearer idea when you finish this book. The next chapters focus on the different patterns and technical requirements you should adopt when you move to microservices, and as you'll see, they're not easy to implement.

The Problems with Microservices from Day Zero

As an alternative to the approach we followed, we could have chosen a microservices architecture from the beginning and split `Users` and `Challenges` into two separate Spring Boot applications.

A reason to start directly with a split is that we may have multiple teams in our organization who could work in parallel without interfering with each other. We could take advantage of mapping microservices to teams from the start. In our example, there are only two domains, but imagine we would have identified ten different bounded contexts. Theoretically, we could leverage the power of a big organization so we finish our project much earlier.

We could also use the split into microservices to accomplish our target architecture. This plan is usually referred to as the *reverse Conway*. Conway's law (see <https://tpd.io/conway>) states that a system design tends to resemble the structure of the organization that is building it. Therefore, it makes sense that we try to use this prediction and adapt our organization to be similar to the software architecture we want to materialize. We draft our complete software architecture, identify the domains, and divide them between the teams.

Being able to work in parallel and achieving the target architecture seem to be great advantages. However, there are two problems behind a too-early split into microservices. The first is that, when we develop software the agile way, we normally can't spend weeks designing the complete system in advance. We will make mistakes when trying to identify loosely coupled domains. And, when we realize that there were flaws, it'll be

too late. It's hard to fight against the inertia of multiple teams working at the same time based on a wrong split of domains, especially if the organization is not flexible enough to cope with those changes. Under these circumstances, the reverse Conway and an early split will play against our goals. We'll create a software system that reflects our initial architecture, but that might be no longer what we want. Check out <https://tpd.io/reverse-conway> where I give more insights about this topic.

The second big issue of starting directly with microservices is that it usually means that we're not splitting the system into vertical slices. We started the previous chapter describing why it's a good idea to deliver our software as soon as possible so we can get feedback. Then, we explained how we can build small portions of the application layers to deliver value, instead of designing and implementing the complete layers one by one. If we start from scratch with multiple microservices, we're going horizontal. A microservice architecture always introduces technical complexity. They're harder to set up, deploy, orchestrate, and test. It will simply cost us more time to have a minimum viable product, and that may have a technical impact too. In the worst case, we could have made software designs based on wrong assumptions, so they become obsolete after we get feedback from users.

Small Monoliths Are for Small Teams

A small monolith is a good plan if we can keep the team small at the beginning. We can focus on domain definition and experimentation. When we have our product ideas validated and a clearer software architecture, more people can gradually join the team, and we can think of splitting the codebase and conforming new teams. We could then move to microservices or choose another approach such as a modular system, depending on our requirements (we'll elaborate more on these two options at the end of this chapter).

However, sometimes we can't avoid starting a project with a big team. It's just a given in our organization. We can't convince the right people that this is not a good idea. If that's the case, the small monolith could become a big monolith quickly, with a spaghetti codebase that might be difficult to modularize later. Besides, it's hard to focus on one vertical slice at a time since there would be then a lot of people doing nothing. The small monolith for a small team idea doesn't work well under this organizational constraint. We need to do some splitting. In that scenario, we must put extra effort into defining not only the bounded contexts but also the communication interfaces between those future

modules. Whenever we design features that span multiple modules or microservices, we make sure we involve the corresponding teams to define what kind of inputs/outputs these modules will produce and consume. The better we define these contracts, the more independent the teams will be. In an Agile environment, this means that the delivery of features can go slower than expected at the beginning since the teams need to define not only these contracts but a lot of common technical foundations.

Embracing Refactoring

Another situation where a small monolith seems problematic is when our organization doesn't embrace code changes. As described before, we start with a small monolith to validate product ideas and get feedback. Then, there will be a point in time where we see the need to start splitting the monolith into microservices. That split comes with both organizational and technical advantages, as we'll detail later. Both technical people and project managers should have a conversation at the beginning of the project to decide when is a good moment to do this split, based on both functional and technical requirements.

Nevertheless, sometimes we as developers think that this moment will never arrive: if we start with a monolith, we'll be chained to it forever. We fear there'll be never a pause in the project's road map to plan and accomplish the needed refactor into microservices. With that in mind, technical people may try to push the organization and force microservices from the beginning. This is a bad idea because it normally frustrates people who may think such technical complexity is delaying the project unnecessarily. Instead of selling a microservices architecture as the only option, it's better to improve communication with business stakeholders and project managers to shape a good plan.

Planning the Small Monolith for a Future Split

When you choose to go for a small monolith, you can follow some good practices for it to be split later with little effort.

- *Compartmentalize the code in root packages defining the domain contexts:* This is what we did in our application with the challenge and user root packages. Then, you could create subpackages for the layering (e.g., controller, repository, domain, and service) if you start dealing with many classes, to ensure layer isolation. Make

sure you follow good practices for class visibility (e.g., interfaces are public, but their implementations are package-private). The main advantages you get with this structure are that you keep business logic inaccessible across domain contexts and that later you should be able to extract one complete root package as a microservice if you need it, with less refactoring.

- *Take advantage of dependency injection:* Base your code on interfaces, and let Spring do its job injecting the implementations. Refactoring using this pattern is much easier. For example, you may change an implementation to call to a different microservice later instead of using local classes, without the rest of the logic being impacted.
- *Once you have identified the contexts (e.g., Challenges and Users), give them a consistent name across your application:* Naming concepts properly is critical at the beginning of the design phase to make sure everybody understands the different domain boundaries.
- *Don't be afraid of moving classes around (easier with a small monolith) during the design phase until boundaries are clear:* After that, respect the boundaries. Never take shortcuts tangling business logic across contexts just because you can. Always keep in mind that the monolith should be prepared to evolve.
- *Find common patterns and identify what can be later extracted as common libraries, for example:* Move them to a different root package.
- *Use peer reviews to make sure the architecture designs are sound and to facilitate knowledge transfer:* It's better to do this as a small group instead of following a top-bottom approach where all designs come from a single person.
- *Clearly communicate to the project manager and/or business representatives to plan time later to split the monolith:* Explain the strategy and create the culture. Refactoring is going to be necessary, and there is nothing wrong with it.

Try to keep a small monolith at least until your first release. Don't be afraid of it; a small monolith will bring you some advantages.

- Faster development in early phases is better to get quick feedback on your product.
- You can easily change the domain boundaries.
- People get used to the same technical guidelines. That helps achieve future consistency.
- Common cross-domain functionality can be identified and shared as libraries (or guidelines).
- The team will get a complete view of the system instead of only parts of it. Then, these people can move to other teams and bring that useful knowledge with them.

New Requirements and Gamification

Imagine that we release our application and connect it to an analytics engine. We get new users every day and also recurrent users who come back regularly thanks to our last feature showing the attempts' history. However, we see in our metrics that, after a week, users tend to abandon the routine of training their brains with new challenges.

Therefore, we make a decision based on our data to try to improve these figures. This simple process is called *data-driven decision-making* (DDDM), and it's important for all kinds of projects. We use data to choose our next move, instead of basing it on intuition or only observation. If you're interested in DDDM, there are multiple articles and courses available on the Internet. The article at <https://tpd.io/ddd> is a good one to start with.

In our case, we plan to introduce some *gamification* to improve engagement in our application. Bear in mind that we'll reduce gamification to points, badges, and leaderboards to keep the book focused on the technical topics. If you're interested in this field, the books *Reality Is Broken* and *For the Win* are good places to start. Before introducing gamification and how it applies to our application, let's present our new user story.

User Story 3

As a user of the application, I want to be motivated to keep solving challenges every day and not abandon it after a while. This way, I keep exercising my brain and improving over time.

Gamification: Points, Badges, and Leaderboards

Gamification is the design process in which you apply techniques that are used in games to another field that is not a game. You do that because you want to get some well-known benefits from games, such as getting players motivated and interacting with your process, application, or whatever you're *gamifying*.

One basic idea about making a game out of something else is introducing *points*: every time you perform an action, and you do well, you get some points. You can even get points if you didn't perform so well, but it should be a fair mechanism: you get more if you do better. Winning points make the player feel as if they're progressing and gives them feedback.

Leaderboards make the points visible to everybody, so they motivate players by activating feelings of competition. We want to get more points than the person above us and rank higher. This is even more fun if you play with friends.

Last but not least, *badges* are virtual symbols of achieving status. We like badges because they say more than points. Also, they can represent different things: you can have the same points as another player (e.g., five correct answers), but you could have won them in a different way (e.g., five in a minute!).

Some software applications that are not games use these elements very well. Take StackOverflow, for example. It's full of game elements to encourage people to keep participating.

What we'll do is assign points to every correct answer that users submit. To keep it simple, we'll give points only if they send a correct attempt. We'll make it 10 points each time.

A leaderboard with the top scores will be shown on the page, so players can find themselves in the ranking and compete with others.

We'll create also some basic badges: Bronze (10 correct attempts), Silver (25 correct attempts), and Gold (50 correct attempts). Because the first correct attempt deserves a nice feedback message, we'll also introduce a First Correct! badge. Also, to introduce a surprise element, we'll have a badge that users can win only if they solve a multiplication where the number 42 is one of the factors.

With these basics, we believe we'll motivate our users to come back and keep playing, competing with their peers.

Moving to Microservices

There is nothing in our new requirements that we couldn't achieve with our small monolith. Actually, if this were a project with one developer and the objective weren't educational, the best option would be to just create a new root package called gamification and start coding our classes within the same deployable unit.

Let's situate ourselves in a different scenario. Imagine that we identify other new features that could help us achieve our business goals. Some of these improvements could be the following:

- Adapt the complexity of the challenge based on the user's statistics.
- Add hints.
- Allow the user to log in instead of using an alias.
- Ask for some user's personal information to collect better metrics.

Those improvements would impact the existing Challenges and Users domains. Besides, since our first release went really well, let's imagine that we also got some capital investment. Our team can grow. The development of our application doesn't need to be sequential anymore. We could work on the Gamification domain at the same time we're improving the existing domains with extra features.

Let's say also that the investors brought some conditions, and now we want to scale up to 100,000 monthly active users. We need to design our architecture to cope with that. And we could soon realize that the new gamification component we're planning to build

is not as critical as the main feature: solving a challenge. If the gamification features are not available for a short period of time, as long as the users can still solve challenges, we're fine.

From our analysis, we could conclude the following:

1. The Users and Challenges domains are critical in our application.
We should aim to keep them highly available. Horizontal scalability would fit very well in this case: we deploy multiple instances of our first application, and we use load balancing and divert the traffic if one of the instances goes down. Besides, replicating the service would also give us more capacity to cope with many concurrent users.
2. The new Gamification domain has different requirements in terms of availability. We don't need to scale up this logic at the same pace. We can accept that it performs slower than other domains, and we can also allow it to stop working for some time.
3. We could benefit from having independently deployable units since our team is growing. If we keep the Gamification module loosely coupled and release it independently, we can work in our organization with multiple teams and minimal interference.

Taking those nonfunctional requirements (e.g., scalability, availability, and extensibility) into account, it seems a good idea to move to microservices. Let's cover these advantages in more detail.

Independent Workflows

We already saw in previous chapters how to accomplish a modular architecture following DDD principles. We could split the resulting bounded contexts into different code repositories so multiple teams can work on them more independently.

However, if these modules are part of the same deployable unit, we still have some dependencies between teams. We need to integrate all these modules together, make sure they work with each other, and deploy the whole thing to our production environment. If other infrastructure elements are shared across modules, like databases, these dependencies become even bigger.

Microservices take modularity to the next level because we can deploy them independently. Teams can have not only different repositories but also different workflows.

In our system, we can develop a few Spring Boot applications in separate repositories. Each of them has its own embedded web server, so we can deploy them separately. This removes all the friction generated during the release of a big monolith: tests, packaging, interdependent database updates, etc.

If we also look at the maintenance and support aspects, microservices help build a DevOps culture because each application may own its corresponding infrastructure elements: web server, database, metrics, logs, etc. When we use a framework like Spring Boot, the system can be seen as a group of mini-applications interacting with each other. If there is a problem in one of these pieces, the team that owns that microservice can be the one fixing it. With a monolith, it's usually harder to draw these lines.

Horizontal Scalability

When we want to scale up a monolithic application, we have the option of doing it *vertically*, with a bigger server/container, or *horizontally*, with more instances and a load balancer. Horizontal scalability is normally the preferred choice since multiple small-size machines are cheaper than a big powerful one. Besides, we can better react to different workload patterns by switching instances on and off.

With microservices, you can choose more flexible strategies for scalability. In our practical example, we considered the multiplication application a critical part of our system, having to cope with a big number of concurrent requests. Therefore, we could decide to deploy two instances of the multiplication microservice but only one instance of the (not yet developed) gamification microservice. If we would keep all the logic in one place, we would replicate also the gamification logic, even though we probably don't need those resources. Figure 6-1 shows an example of this.

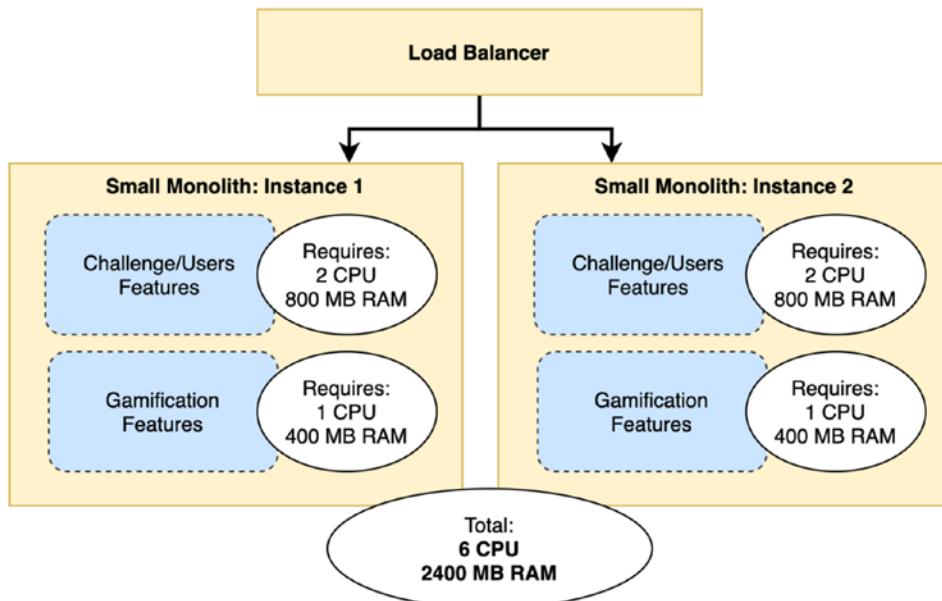
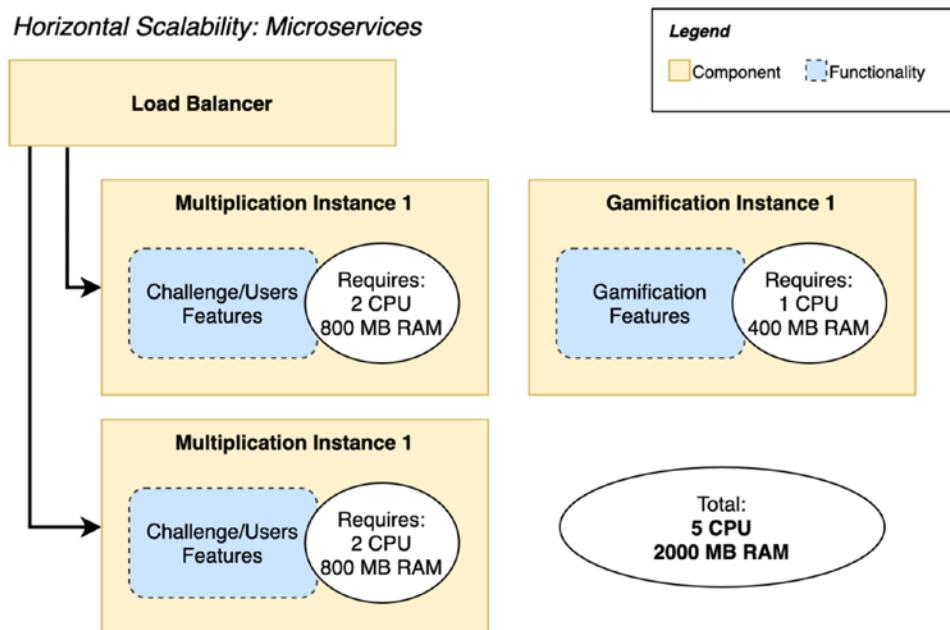
Horizontal Scalability: Monolith*Horizontal Scalability: Microservices*

Figure 6-1. Horizontal scalability in a monolith versus microservices

Fine-Grained Nonfunctional Requirements

We could extrapolate the horizontal scalability advantages to other nonfunctional requirements. For instance, we said it's not that bad if the new gamification microservice is not available for a short time. If we are running a monolithic application, the whole system could potentially go down due to an unexpected condition on the gamification module. With microservices, we have the choice to allow for complete parts in our system to be down for a while. We'll see later in this book how we can implement resilience patterns to build a more fault-tolerant back end.

The same applies to security, for example. We might need more restrictive access to a microservice that manages a user's personal data, but we don't need to deal with that security overhead in the gamification domain. Being independent applications, microservices bring more flexibility.

Other Advantages

There are some other reasons why we might want to choose a microservices architecture. However, I grouped them separately since I don't think you should treat them as decisive factors to make the move.

- *Multiple technologies:* We may want to build some microservices in Java and some others in Golang, for example. However, that comes with a cost since it won't be that easy to use common artifacts or frameworks or share knowledge across teams. We may also want to use different database engines, but that is also possible in a modular monolith.
- *Consistency with the organizational structure:* As we described earlier in this chapter, you might be tempted to use Conway's law and try to design microservices following your organization structure, or vice versa. But we already covered how that has both pros and cons, depending also on whether you do the split directly at the beginning or later in the project's lifecycle.
- *Ability to replace parts of your system:* If microservices bring more isolation to your software architecture, it's logical to think that it should be easier to replace them without causing much impact on other services. But, in real life, microservices could also become

strongly coupled to each other when some basic rules are not respected. And, on the other hand, you can achieve replaceability with a good modular system. Therefore, I don't see this as a decisive driver for change either.

Disadvantages

As we covered already in the previous sections, microservice architectures have many disadvantages too, so they're not the panacea to solve all the issues that may arise with a monolithic architecture. We covered some of these disadvantages while analyzing why it's a good idea to start with a small monolith.

- *You need more time to deliver a first working version:* Due to its complexity, a microservice architecture requires much more time to set it up correctly when compared to a single service.
- *Moving functionality across domains becomes harder:* Once you've made the first split, it requires extra work to merge code or move functionalities across microservices, when compared to a single codebase or deployment unit.
- *There is an implicit introduction of new paradigms:* Given that a microservice architecture makes your system distributed, you'll face new challenges like asynchronous processing, distributed transactions, and eventual consistency. We'll analyze these new paradigms in detail in this chapter and the next one.
- *It requires learning new patterns:* When you have a distributed system, you better know how to implement routing, service discovery, distributed tracing and logging, etc. These patterns are not easy to implement and maintain. We'll cover them in Chapter 8.
- *You may require adopting new tools:* There are some frameworks and tools that can help you implement a microservice architecture: Spring Cloud, Docker, Message Brokers, Kubernetes, etc. You may not need them in a monolithic architecture, and that means extra maintenance, setup, potential costs, and time to learn all these new concepts. Again, this book will help you understand these tools over the next chapters.

- *More resources are needed to run your system:* At the beginning of a project, when the system traffic is not high yet, maintaining a microservice-based system can be much more expensive than a monolithic one. Having multiple idle services is less efficient than having a single one. Besides, the surrounding tools and patterns (that we'll cover in this book) introduce an extra overload. The impact only starts being positive when you benefit from scalability, fault tolerance, and other characteristics we'll describe later in this book.
- *There might be a diversion from standards and common practices:* One of the reasons to move to microservices could be achieving more independence across teams. However, that may also have a negative impact if everybody starts creating their own solutions to solve the same problems, instead of reusing common patterns. That could cause a waste of time and makes it harder for people to understand other parts of the system.
- *The architecture is much more complex:* Explaining how your system works might become harder with a microservice architecture. That means new joiners will require extra time to understand how the complete system works. One might argue that this is not needed as long as people understand the domain they work on, but it's always better to get to know how all the pieces interact together.
- *You could be distracted with new techniques you don't need:* Once you board the train of microservices with their fancy tools around them, some people could be attracted by new products and patterns that are *cool* to implement. However, you might not need them, so they become just a distraction. Even though this can happen with any type of architecture, it happens more often when you work with microservices.

Some of these points might be still unclear for you. Don't worry, you'll be able to understand exactly what they mean at the end of our journey. This book follows a pragmatic and realistic approach to these subjects to help you understand both the advantages and the disadvantages of a microservice architecture so you can make the best decision in the future.

Architecture Overview

After comparing the alternatives we have and analyzing the pros and cons of microservices, we decided to make the move and create a new Spring Boot application for our gamification requirements. The scalability of both the system and the organization plays a major role in this decision, given our hypothetical scenario.

Now we can refer to these two applications as the Multiplication microservice and the Gamification microservice. It didn't make sense until now to call our first application a microservice since it wasn't part of a microservices architecture yet.

Figure 6-2 represents the different components in our system and how they'll be connected by the end of this chapter.

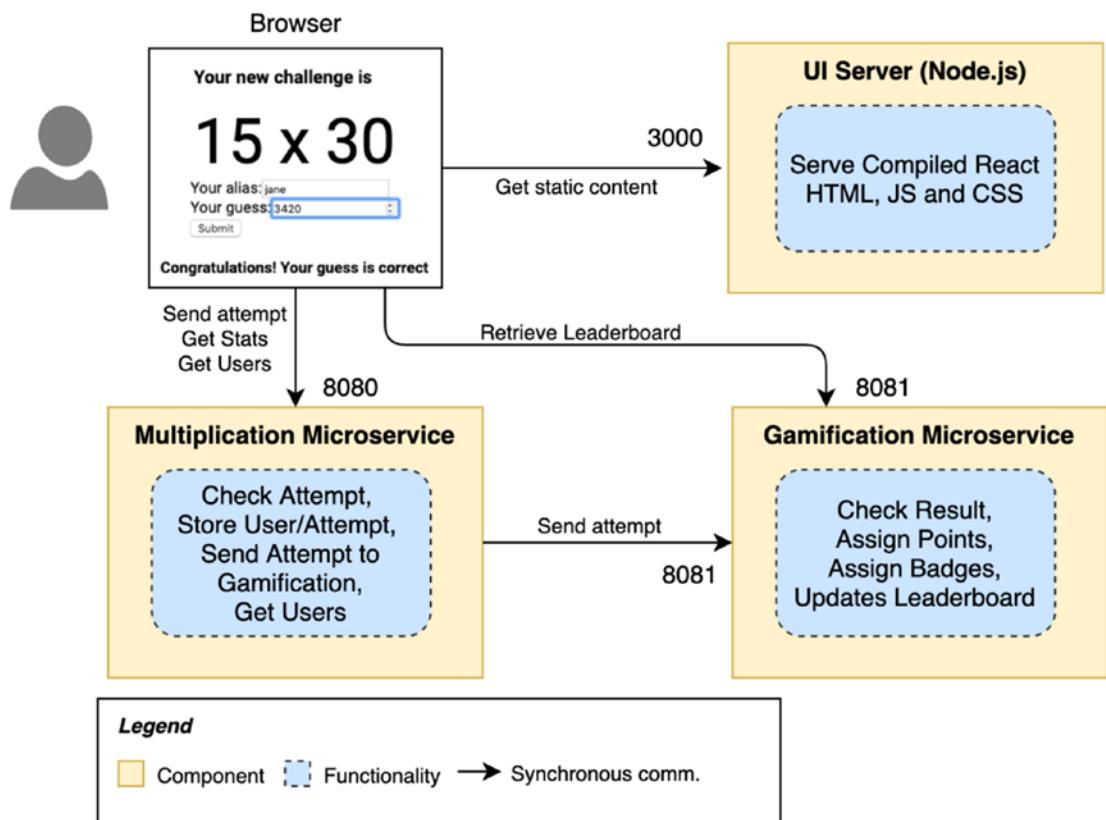


Figure 6-2. Logical view

Let's review the new additions in this design.

- There will be a new microservice, Gamification. To prevent local port conflicts, we'll deploy it on port 8081.
- The Multiplication microservice will send each attempt to the Gamification microservice to process new score, badges, and update the leaderboard.
- There will be a new component in our React UI to render the leaderboard with the score and badges. As you can see in the drawing, our UI will call both microservices.

Some notes about this design:

- We could also deploy the UI from one of the embedded web servers. However, it's better to treat the UI server as a different deployment unit. The same advantages apply here: independent workflow, flexible scalability, etc.
- It might look weird that the UI needs to call two services. You could consider that it's better to put a reverse proxy in front of the other two, to do the routing and keep the API client unaware of the back end's software architecture (see https://en.wikipedia.org/wiki/Reverse_proxy). That's actually the Gateway pattern, and we'll cover it in detail later in this book. Let's keep it simple for now.
- If you paid attention to this book's summary, the synchronous call from Multiplication to Gamification surely caught your attention. That's not the best design indeed, but let's keep the evolving approach example and learn first why it isn't the best idea.

Designing and Implementing the New Service

In this section, we'll design and implement the Gamification microservice following a similar approach to what we did with our first Spring Boot service, Multiplication.

Interfaces

When working with a modular system in general, we must pay attention to the *contract* between the modules. With microservices, this is even more relevant since, as a team, we want to clarify as soon as possible all the expected dependencies.

In our case, the Gamification microservice needs to expose an interface to accept new attempts. It needs that data to compute the statistics for users. For now, this interface will be a REST API. The exchanged JSON object can simply contain the same fields as the attempt we store on the Multiplication microservice: the attempt's data and the user's data. On the Gamification side, we'll use only the data we need.

On the other hand, the UI needs to collect the leaderboard details. We'll also create new REST endpoints in the Gamification microservice to access this data.

Info

Starting here, this chapter's section goes through the source code of the new Gamification microservice. It's good to take a look at it since we'll use some new small features across the different layers. However, we saw the main concepts already in previous chapters, so you may decide to take a shortcut. That's also possible. If you don't want to dive into the development of the Gamification microservice, you can jump directly to the section "Playing with Our Services" and use the code for this chapter, available at <https://github.com/Book-Microservices-v2/chapter06>.

The Spring Boot Skeleton for Gamification

We can use again the Spring Initializr at <https://start.spring.io/> to create the basic skeleton for our new application. This time, we know in advance that we'll need some extra dependencies, so we can add them directly from here: Lombok, Spring Web, Validation, Spring Data JPA, and the H2 Database. Fill in the details as shown in Figure 6-3.

The screenshot shows the Spring Initializr interface for creating a new project. The configuration is as follows:

- Project:** Maven Project
- Language:** Java
- Spring Boot:** 2.3.3
- Project Metadata:**
 - Group: microservices.book
 - Artifact: gamification
 - Name: gamification
 - Description: Gamification Microservice
 - Package name: microservices.book.gamification
 - Packaging: Jar
 - Java: 14
- Dependencies:**
 - Lombok: DEVELOPER TOOLS
 - Spring Web: WEB
 - Validation: I/O
 - Spring Data JPA: SQL
 - H2 Database: SQL
- Buttons at the bottom:**
 - GENERATE ⌘ + ↵
 - EXPLORE CTRL + SPACE
 - SHARE...

Figure 6-3. Creating the Gamification application

Download the zip file and extract it as a `gamification` folder next to the existing `multiplication` one. You can add this new project as a separate module within the same workspace, to keep everything organized within the same IDE instance.

Domain

Let's model our gamification domain trying to respect the context boundaries and minimizing the coupling with the existing functionality.

- We create a *score card* object, which holds the amount of score that a user obtains for a given challenge attempt.
- Similarly, we have a *badge card* object, representing a specific type of badge that has been won at a given time by a user. It doesn't need to be tied to a score card since you may win a badge when you surpass a given score threshold.
- To model the leaderboard, we create a *leaderboard position*. We'll display an ordered list of these domain objects to show the ranking to the users.

In this model, there are some relationships between our existing domain objects and the new ones, as shown in Figure 6-4.

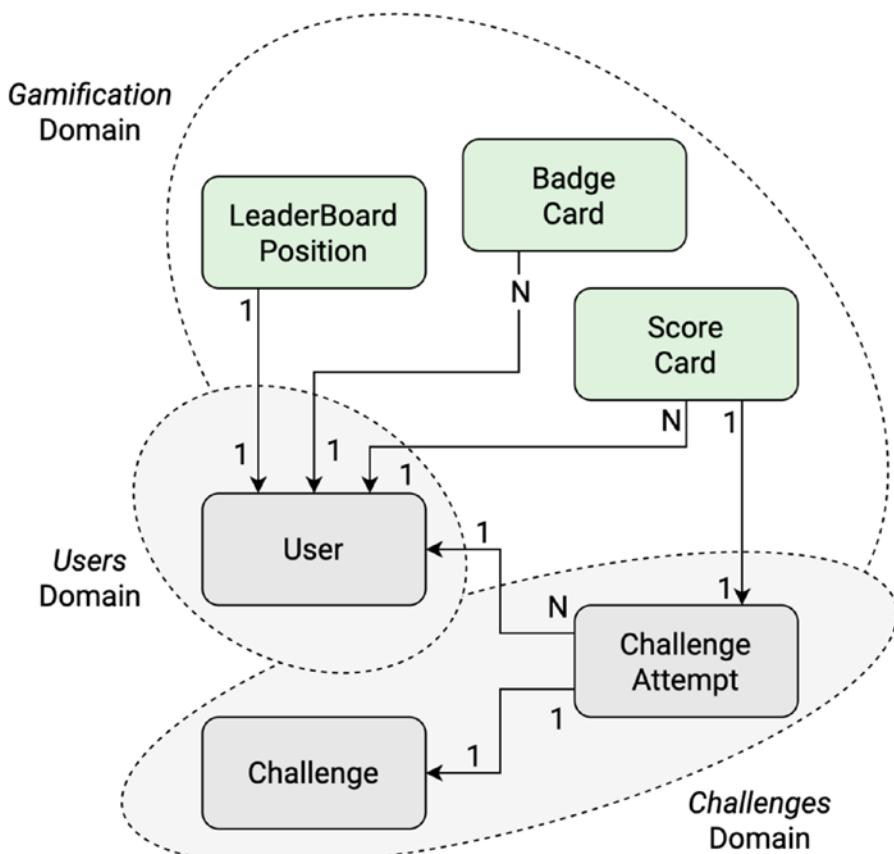


Figure 6-4. The new Gamification domain

As you see, we still keep these domains loosely coupled:

- The Users domain remains completely isolated. It doesn't keep any reference to any other object.
- The Challenges domain only needs to know about Users. We don't need to link their objects to the gamification concepts.
- The Gamification domain needs to reference Users and Challenge Attempts. We planned to get this data after an attempt is sent, so we'll store some references locally (the identifiers of the user and the attempt).

The domain objects can be mapped to Java classes easily. We'll use JPA/Hibernate also for this service so we can already add the JPA annotations. First, Listing 6-1 shows the ScoreCard class, with an extra constructor that will set some default values.

Source Code

You can find all the source code for this chapter on GitHub, in the `chapter06` repository.

See <https://github.com/Book-Microservices-v2/chapter06>.

Listing 6-1. The ScoreCard Domain/Data Class

```
package microservices.book.gamification.game.domain;

import lombok.*;
import javax.persistence.*;

/**
 * This class represents the Score linked to an attempt in the game,
 * with an associated user and the timestamp in which the score
 * is registered.
 */
@Entity
@Data
@AllArgsConstructor
```

```

@NoArgsConstructor
public class ScoreCard {

    // The default score assigned to this card, if not specified.
    public static final int DEFAULT_SCORE = 10;

    @Id
    @GeneratedValue
    private Long cardId;
    private Long userId;
    private Long attemptId;
    @EqualsAndHashCode.Exclude
    private long scoreTimestamp;
    private int score;

    public ScoreCard(final Long userId, final Long attemptId) {
        this(null, userId, attemptId, System.currentTimeMillis(), DEFAULT_SCORE);
    }

}

```

We used a new Lombok annotation this time, `@EqualsAndHashCode.Exclude`. As the name suggests, this will make Lombok not to include that field in the generated equals and hashCode methods. The reason is that this will make our tests easier when we compare objects, and in fact, we don't need the timestamp to determine whether two cards are equal.

The different badges are defined in an enum, `BadgeType`. We'll add a `description` field to have a friendly name for each one. See Listing 6-2.

Listing 6-2. The `BadgeType` Enum

```

package microservices.book.gamification.game.domain;

import lombok.Getter;
import lombok.RequiredArgsConstructor;

/**
 * Enumeration with the different types of Badges that a user can win.
 */
@RequiredArgsConstructor
@Getter

```

```
public enum BadgeType {

    // Badges depending on score
    BRONZE("Bronze"),
    SILVER("Silver"),
    GOLD("Gold"),

    // Other badges won for different conditions
    FIRST_WON("First time"),
    LUCKY_NUMBER("Lucky number");

    private final String description;
}
```

As you can see in the previous code, we also benefit from some Lombok annotations in enums. In this case, we use them to generate a constructor and the getter for the description field.

The BadgeCard class uses BadgeType, and it's also a JPA entity. See Listing 6-3.

Listing 6-3. The BadgeCard Domain/Data Class

```
package microservices.book.gamification.game.domain;

import lombok.*;
import javax.persistence.*;

@Entity
@Data
@AllArgsConstructor
@NoArgsConstructor
public class BadgeCard {

    @Id
    @GeneratedValue
    private Long badgeId;

    private Long userId;
    @EqualsAndHashCode.Exclude
    private long badgeTimestamp;
    private BadgeType badgeType;
```

```
public BadgeCard(final Long userId, final BadgeType badgeType) {
    this(null, userId, System.currentTimeMillis(), badgeType);
}
}
```

We also added a constructor to set some default values. Note that we don't need to add any specific JPA annotation to the enum type. Hibernate will map the values to the enum's ordinal value (an integer) by default. This works just fine if we keep in mind that we should only append new enum values at the end, but we could also configure the mapper to use the string values.

To model the leaderboard position, we create the class `LeaderBoardRow`. See Listing 6-4. We don't need to persist this object in our database since it's going to be created on the fly by aggregating scores and badges from our users.

Listing 6-4. The LeaderBoardRow Class

```
package microservices.book.gamification.game.domain;

import lombok.*;
import java.util.List;

@Value
@AllArgsConstructor
public class LeaderBoardRow {

    Long userId;
    Long totalScore;
    @With
    List<String> badges;

    public LeaderBoardRow(final Long userId, final Long totalScore) {
        this.userId = userId;
        this.totalScore = totalScore;
        this.badges = List.of();
    }
}
```

The `@With` annotation added to the `badges` field is provided by Lombok and generates a method for us to clone an object and add a new field value to the copy (in this case, `withBadges`). This is a good practice when we work with immutable classes since they don't have setters. We'll use this method when we create the business logic to merge the score and the badges for each leaderboard row.

Service

We'll divide the business logic in this new Gamification microservice into two.

- The game logic, responsible for processing the attempt and generating the resulting score and badges
- The leaderboard logic, which aggregates data and builds the ranking based on score

The game logic will reside in the class `GameServiceImpl`, which implements the `GameService` interface. The specification is simple: based on an attempt, it computes the score and badges and stores them. This business logic is accessible to the Multiplication microservice via a controller named `GameController`, which will expose a POST endpoint to send the attempt to. On the persistence layer, our business logic will require a `ScoreRepository` to save scorecards and a `BadgeRepository` to do the same thing with badge cards. Figure 6-5 shows a UML diagram with all the classes needed to build the game logic functionality.

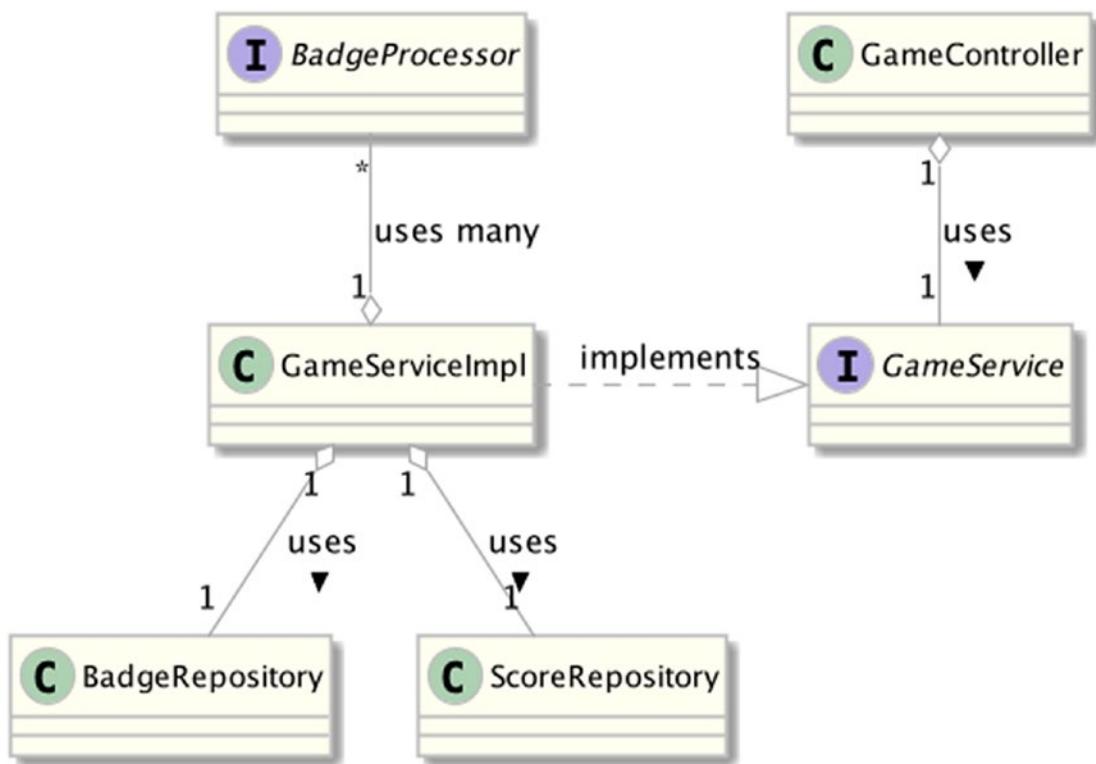


Figure 6-5. UML: game logic

We can define the GameService interface as shown in Listing 6-5.

Listing 6-5. The GameService Interface

```

package microservices.book.gamification.game;

import java.util.List;
import lombok.Value;
import microservices.book.gamification.challenge.ChallengeSolvedDTO;
import microservices.book.gamification.game.domain.BadgeCard;
import microservices.book.gamification.game.domain.BadgeType;
import microservices.book.gamification.game.domain.ScoreCard;

public interface GameService {
    /**
     * Process a new attempt from a given user.
     *
  
```

```

    * @param challenge the challenge data with user details, factors, etc.
    * @return a {@link GameResult} object containing the new score and badge
      cards obtained
  */
GameResult newAttemptForUser(ChallengeSolvedDTO challenge);

@Value
class GameResult {
  int score;
  List<BadgeType> badges;
}
}

```

The output after processing the attempt is a `GameResult` object, defined within the interface. It groups the score obtained from that attempt together with any new badge that the user may get. We could also consider not returning anything since it'll be the leaderboard logic showing the results. However, it's better to have a response from our method so we can test it.

The `ChallengeSolvedDTO` class defines the contract between the Multiplication and Gamification microservices, and we'll create it in both projects to keep them independent. For now, let's focus on the Gamification codebase. See Listing 6-6.

Listing 6-6. The ChallengeSolvedDTO Class

```

package microservices.book.gamification.challenge;

import lombok.Value;

@Value
public class ChallengeSolvedDTO {

  long attemptId;
  boolean correct;
  int factorA;
  int factorB;
  long userId;
  String userAlias;

}

```

Now that we have defined the domain classes and the skeleton of our service layer, we can use TDD and create some test cases for our business logic, using an empty interface implementation and the DTO class.

Exercise

Create the GameServiceTest with two test cases: a correct attempt and a wrong one. You'll find the solution in this chapter's code sources.

For now, focus only on the score calculation and not on the badges. We'll create a separate interface and the tests for that part.

A valid implementation of the GameService interface would be the one shown in Listing 6-7. It creates a ScoreCard object only if the challenge is correctly solved, and stores it. The badges are processed in a separate method for better readability. We also need some repository methods to save the score and badges and to retrieve the previously created records. For now, we can just assume these methods work; we'll explain them in detail later in the “Data” section.

Listing 6-7. Implementing the GameService Interface in the GameServiceImpl Class

```
package microservices.book.gamification.game;

import java.util.*;
import java.util.stream.Collectors;
import org.springframework.stereotype.Service;
import lombok.RequiredArgsConstructor;
import lombok.extern.slf4j.Slf4j;
import microservices.book.gamification.challenge.ChallengeSolvedDTO;
import microservices.book.gamification.game.badgeprocessors.BadgeProcessor;
import microservices.book.gamification.game.domain.BadgeCard;
import microservices.book.gamification.game.domain.BadgeType;
import microservices.book.gamification.game.domain.ScoreCard;

@Service
@Slf4j
@RequiredArgsConstructor
class GameServiceImpl implements GameService {
```

```

private final ScoreRepository scoreRepository;
private final BadgeRepository badgeRepository;
// Spring injects all the @Component beans in this list
private final List<BadgeProcessor> badgeProcessors;

@Override
public GameResult newAttemptForUser(final ChallengeSolvedDTO challenge) {
    // We give points only if it's correct
    if (challenge.isCorrect()) {
        ScoreCard scoreCard = new ScoreCard(challenge.getUserId(),
            challenge.getAttemptId());
        scoreRepository.save(scoreCard);
        log.info("User {} scored {} points for attempt id {}",
            challenge.getUserAlias(), scoreCard.getScore(),
            challenge.getAttemptId());
        List<BadgeCard> badgeCards = processForBadges(challenge);
        return new GameResult(scoreCard.getScore(),
            badgeCards.stream().map(BadgeCard::getBadgeType)
                .collect(Collectors.toList()));
    } else {
        log.info("Attempt id {} is not correct. " +
            "User {} does not get score.",
            challenge.getAttemptId(),
            challenge.getUserAlias());
        return new GameResult(0, List.of());
    }
}

/**
 * Checks the total score and the different score cards obtained
 * to give new badges in case their conditions are met.
 */
private List<BadgeCard> processForBadges(
    final ChallengeSolvedDTO solvedChallenge) {
    Optional<Integer> optTotalScore = scoreRepository.
        getTotalScoreForUser(solvedChallenge.getUserId());
    if (optTotalScore.isEmpty()) return Collections.emptyList();
    int totalScore = optTotalScore.get();
}

```

```

// Gets the total score and existing badges for that user
List<ScoreCard> scoreCardList = scoreRepository
    .findByUserIdOrderByScoreTimestampDesc(solvedChallenge.
        getUserId());
Set<BadgeType> alreadyGotBadges = badgeRepository
    .findByUserIdOrderByBadgeTimestampDesc(solvedChallenge.
        getUserId())
    .stream()
    .map(BadgeCard::getBadgeType)
    .collect(Collectors.toSet());

// Calls the badge processors for badges that the user doesn't have yet
List<BadgeCard> newBadgeCards = badgeProcessors.stream()
    .filter(bp -> !alreadyGotBadges.contains(bp.badgeType()))
    .map(bp -> bp.processForOptionalBadge(totalScore,
        scoreCardList, solvedChallenge))
    .flatMap(Optional::stream) // returns an empty stream if empty
// maps the optionals if present to new BadgeCards
    .map(badgeType ->
        new BadgeCard(solvedChallenge.getUserId(), badgeType)
    )
    .collect(Collectors.toList());
badgeRepository.saveAll(newBadgeCards);

return newBadgeCards;
}

}

```

As we can conclude from this implementation, the `BadgeProcessor` interface is accepting some contextual data and the solved attempt, and it's deciding whether to assign a given type of badge. Listing 6-8 shows the source code of that interface.

Listing 6-8. The `BadgeProcessor` Interface

```

package microservices.book.gamification.game.badgeprocessors;

import java.util.List;
import java.util.Optional;
import microservices.book.gamification.challenge.ChallengeSolvedDTO;

```

```

import microservices.book.gamification.game.domain.BadgeType;
import microservices.book.gamification.game.domain.ScoreCard;

public interface BadgeProcessor {

    /**
     * Processes some or all of the passed parameters and decides if the user
     * is entitled to a badge.
     *
     * @return a BadgeType if the user is entitled to this badge, otherwise empty
     */
    Optional<BadgeType> processForOptionalBadge(
        int currentScore,
        List<ScoreCard> scoreCardList,
        ChallengeSolvedDTO solved);

    /**
     * @return the BadgeType object that this processor is handling. You can use
     * it to filter processors according to your needs.
     */
    BadgeType badgeType();

}

```

Since we use constructor injection in GameServiceImpl with a list of BadgeProcessor objects, Spring will find all the beans that implement this interface and pass them to us. This is a flexible way of extending our game without interfering with other existing logic. We just need to add new BadgeProcessor implementations and annotate them with @Component so they are loaded in the Spring context.

Listings 6-9 and 6-10 are two of the five badge implementations that we need to fulfill our functional requirements, BronzeBadgeProcessor and FirstWonBadgeProcessor.

Listing 6-9. BronzeBadgeProcessor Implementation

```

package microservices.book.gamification.game.badgeprocessors;

import microservices.book.gamification.challenge.ChallengeSolvedDTO;
import microservices.book.gamification.game.domain.BadgeType;
import microservices.book.gamification.game.domain.ScoreCard;
import org.springframework.stereotype.Component;

```

```

import java.util.List;
import java.util.Optional;

@Component
class BronzeBadgeProcessor implements BadgeProcessor {

    @Override
    public Optional<BadgeType> processForOptionalBadge(
        int currentScore,
        List<ScoreCard> scoreCardList,
        ChallengeSolvedDTO solved) {
        return currentScore > 50 ?
            Optional.of(BadgeType.BRONZE) :
            Optional.empty();
    }

    @Override
    public BadgeType badgeType() {
        return BadgeType.BRONZE;
    }
}

```

Listing 6-10. FirstWonBadgeProcessor Implementation

```

package microservices.book.gamification.game.badgeprocessors;

import microservices.book.gamification.challenge.ChallengeSolvedDTO;
import microservices.book.gamification.game.domain.BadgeType;
import microservices.book.gamification.game.domain.ScoreCard;
import org.springframework.stereotype.Component;
import java.util.List;
import java.util.Optional;

@Component
class FirstWonBadgeProcessor implements BadgeProcessor {

    @Override
    public Optional<BadgeType> processForOptionalBadge(
        int currentScore,
        List<ScoreCard> scoreCardList,
        ChallengeSolvedDTO solved) {

```

```
    return scoreCardList.size() == 1 ?
        Optional.of(BadgeType.FIRST_WON) : Optional.empty();
}

@Override
public BadgeType badgeType() {
    return BadgeType.FIRST_WON;
}
}
```

Exercise

Implement the other three badge processors and all the unit tests to verify they work as expected. If you need help, you may consult this chapter’s source code.

1. The Silver badge. Won if the score exceeds 150.
 2. The Gold badge. Won if the score exceeds 400.
 3. The “Lucky number” badge. Won if any of the factors of the attempt is 42.
-

Once we finish the first block of the business logic, we can move to the second one: the leaderboard functionality. Figure 6-6 shows the UML diagram of the three layers we’ll implement in this chapter to build the leaderboard.

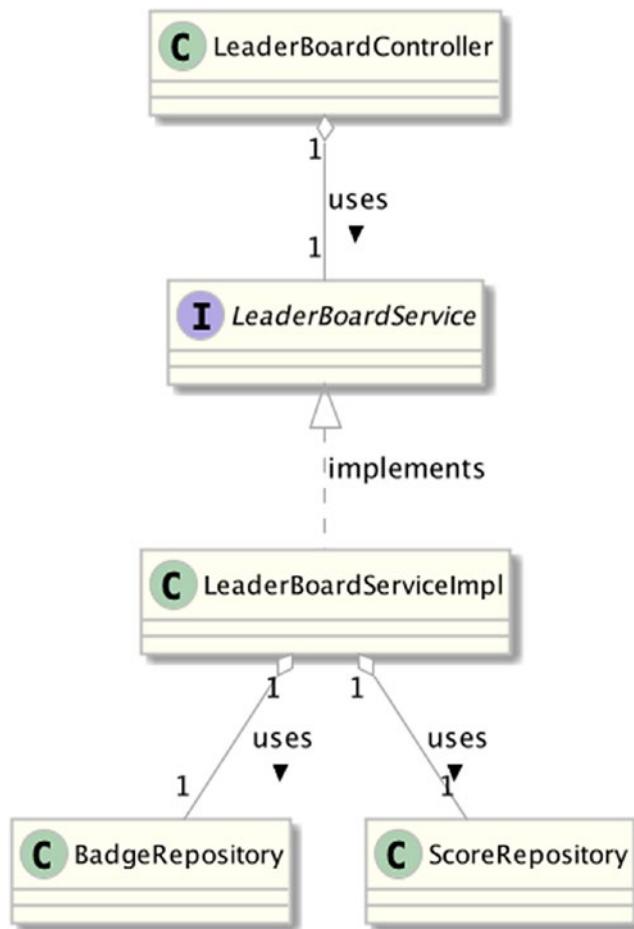


Figure 6-6. Leaderboard, UML diagram

The interface `LeaderBoardService` has a single method to return a sorted list of `LeaderBoardRow` objects. See Listing 6-11.

Listing 6-11. The `LeaderBoardService` Interface

```
package microservices.book.gamification.game;

import java.util.List;
import microservices.book.gamification.game.domain.LeaderBoardRow;

public interface LeaderBoardService {
```

```

    /**
     * @return the current leader board ranked from high to low score
     */
    List<LeaderBoardRow> getCurrentLeaderBoard();
}

```

Exercise

Create LeaderBoardServiceImplTest to verify that the implementation should query ScoreCardRepository to find the users with the top score and should query BadgeCardRepository to merge the score with their badges. As before, the repository classes are not there yet, but you can create some dummy methods and mock them for the tests.

The implementation of the leaderboard service can remain simple if we have the ability to aggregate the score and sort the resulting rows in the database. We'll see how in the next section. For now, we assume that we can get the score ranking from ScoreRepository (the `findFirst10` method). Then, we query the database to retrieve the badges for the users included in the ranking. See Listing 6-12.

Listing 6-12. The LeaderBoardService Implementation

```

package microservices.book.gamification.game;

import java.util.List;
import java.util.stream.Collectors;
import org.springframework.stereotype.Service;
import lombok.RequiredArgsConstructor;
import microservices.book.gamification.game.domain.LeaderBoardRow;

@Service
@RequiredArgsConstructor
class LeaderBoardServiceImpl implements LeaderBoardService {

    private final ScoreRepository scoreRepository;
    private final BadgeRepository badgeRepository;
}

```

```

@Override
public List<LeaderBoardRow> getCurrentLeaderBoard() {
    // Get score only
    List<LeaderBoardRow> scoreOnly = scoreRepository.findFirst10();
    // Combine with badges
    return scoreOnly.stream().map(row -> {
        List<String> badges =
            badgeRepository.findByIdOrderByBadgeTimestampDesc(
                row.getUserId()).stream()
                .map(b -> b.getBadgeType().getDescription())
                .collect(Collectors.toList());
        return row.withBadges(badges);
    }).collect(Collectors.toList());
}
}

```

Note that we used the method `withBadges` to copy an immutable object with a new value. The first time we generate the leaderboard, all rows have an empty list of badges. When we collect the badges, we can replace (using stream's `map`) each object with a copy with the corresponding badge list.

Data

In the business logic layer, we made some assumptions about `ScoreRepository` and `BadgeRepository` methods. It's time to build these repositories.

Remember that we get basic CRUD functionality just by extending Spring Data's `CrudRepository`, so we can save badges and score cards easily. For the rest of the queries, we'll make use of both query methods and JPQL.

The `BadgeRepository` interface defines a query method to find badges for a given user, ordered by date with most recent ones on top. See Listing 6-13.

Listing 6-13. The `BadgeRepository` Interface with a Query Method

```

package microservices.book.gamification.game;

import microservices.book.gamification.game.domain.BadgeCard;
import microservices.book.gamification.game.domain.BadgeType;
import org.springframework.data.repository.CrudRepository;
import java.util.List;

```

```

/**
 * Handles data operations with BadgeCards
 */
public interface BadgeRepository extends CrudRepository<BadgeCard, Long> {

    /**
     * Retrieves all BadgeCards for a given user.
     *
     * @param userId the id of the user to look for BadgeCards
     * @return the list of BadgeCards, ordered by most recent first.
     */
    List<BadgeCard> findByUserIdOrderByBadgeTimestampDesc(Long userId);
}

```

For scorecards, we need other query types. We identified three requirements thus far.

1. Calculate the total score of a user.
2. Get a list of users with the highest score, ordered, as LeaderBoardRow objects.
3. Read all ScoreCard records by user ID.

[Listing 6-14](#) shows the complete source code for the ScoreRepository.

Listing 6-14. The ScoreRepository Interface, Using Query Methods and JPQL Queries

```

package microservices.book.gamification.game;

import java.util.List;
import java.util.Optional;
import org.springframework.data.jpa.repository.Query;
import org.springframework.data.repository.CrudRepository;
import org.springframework.data.repository.query.Param;
import microservices.book.gamification.game.domain.LeaderBoardRow;
import microservices.book.gamification.game.domain.ScoreCard;

/**
 * Handles CRUD operations with ScoreCards and other related score queries
 */
public interface ScoreRepository extends CrudRepository<ScoreCard, Long> {

```

```

/**
 * Gets the total score for a given user: the sum of the scores of all
 * their ScoreCards.
 *
 * @param userId the id of the user
 * @return the total score for the user, empty if the user doesn't exist
 */
@Query("SELECT SUM(s.score) FROM ScoreCard s WHERE s.userId = :userId GROUP BY
s.userId")
Optional<Integer> getTotalScoreForUser(@Param("userId") Long userId);

/**
 * Retrieves a list of {@link LeaderBoardRow}s representing the Leader Board
 * of users and their total score.
 *
 * @return the leader board, sorted by highest score first.
 */
@Query("SELECT NEW microservices.book.gamification.game.domain.
LeaderBoardRow(s.userId, SUM(s.score)) " +
        "FROM ScoreCard s " +
        "GROUP BY s.userId ORDER BY SUM(s.score) DESC")
List<LeaderBoardRow> findFirst10();

/**
 * Retrieves all the ScoreCards for a given user, identified by his user id.
 *
 * @param userId the id of the user
 * @return a list containing all the ScoreCards for the given user,
 * sorted by most recent.
 */
List<ScoreCard> findByUserIdOrderByScoreTimestampDesc(final Long userId);
}

```

Unfortunately, Spring Data JPA's query methods don't support aggregations. The good news is that JPQL, the JPA Query Language, does support them, so we can use standard syntax to keep our code as database-agnostic as possible. We can get the total score for a given user with this query:

```
SELECT SUM(s.score) FROM ScoreCard s WHERE s.userId = :userId GROUP BY s.userId
```

Like in standard SQL, the `GROUP BY` clause indicates how to sum up the values. We can define parameters with the `:param` notation. Then, we annotate the corresponding method arguments with `@Param`. We could also use the approach we followed in the previous chapter, with argument position's placeholders like `?1`.

The second query is a bit special. In JPQL, we can use the constructors available in our Java classes. What we do in our example is an aggregation based on the total score, and we construct `LeaderBoardRow` objects using the two-argument constructor we defined (which sets an empty list of badges). Keep in mind that we have to use the full qualified name of the class in JPQL as shown in the source code.

Controller

While designing our Gamification domain, we agreed on a *contract* with the Multiplication service. It'll send each attempt to a REST endpoint on the gamification side. It's time to build that controller. See Listing 6-15.

Listing 6-15. The GameController Class

```
package microservices.book.gamification.game;

import org.springframework.http.HttpStatus;
import org.springframework.web.bind.annotation.*;
import lombok.RequiredArgsConstructor;
import microservices.book.gamification.challenge.ChallengeSolvedDTO;

@RestController
@RequestMapping("/attempts")
@RequiredArgsConstructor
public class GameController {

    private final GameService gameService;

    @PostMapping
    @ResponseStatus(HttpStatus.OK)
    void postResult(@RequestBody ChallengeSolvedDTO dto) {
        gameService.newAttemptForUser(dto);
    }
}
```

There is a REST API available on POST /attempts that accepts a JSON object containing data about the user and the challenge. In this case, we don't need to return any content, so we make use of the `ResponseStatus` annotation to configure Spring to return a 200 OK status code. Actually, this is the default behavior when a controller's method returns `void` and has been processed without errors. In any case, it's good to add it explicitly for better readability. Remember that if there is an error like a thrown exception, for example, Spring Boot's default error handling logic will intercept it and return an error response with a different status code.

We could also add validation to the DTO class to make sure other services don't send invalid data to the Gamification microservice, but, for now, let's keep it simple. We'll change this API in the next chapter anyway.

Exercise

Don't forget to add the tests for this first controller and the next one. You can find these tests in this chapter's source code.

Our second controller is for the leaderboard functionality and exposes a GET /leaders method that returns a JSON array of serialized `LeaderBoardRow` objects. This data is coming from the service layer, which uses the badge and score repositories to merge users' scores and badges. Therefore, the presentation layer remains simple. See the code in Listing 6-16.

Listing 6-16. The LeaderBoardController Class

```
package microservices.book.gamification.game;

import lombok.RequiredArgsConstructor;
import microservices.book.gamification.game.domain.LeaderBoardRow;
import org.springframework.web.bind.annotation.*;
import java.util.List;

/**
 * This class implements a REST API for the Gamification LeaderBoard service.
 */
@RestController
@RequestMapping("/leaders")
```

```

@RequiredArgsConstructor
class LeaderBoardController {

    private final LeaderBoardService leaderBoardService;

    @GetMapping
    public List<LeaderBoardRow> getLeaderBoard() {
        return leaderBoardService.getCurrentLeaderBoard();
    }
}

```

Configuration

We went through the three layers of our application: business logic, data, and presentation. We're still missing some Spring Boot configuration that we defined in the Multiplication microservice as well.

First, let's add some values to the `application.properties` file in the Gamification microservice. See Listing 6-17.

Listing 6-17. The `application.properties` File for the Gamification App

```

server.port=8081
# Gives us access to the H2 database web console
spring.h2.console.enabled=true
# Creates the database in a file
spring.datasource.url=jdbc:h2:file:~/gamification;DB_CLOSE_ON_EXIT=FALSE
# Creates or updates the schema if needed
spring.jpa.hibernate.ddl-auto=update
# For educational purposes we will show the SQL in console
spring.jpa.show-sql=true

```

The only new addition is the `server.port` property. We change it since we can't use the same default 8080 in our second application when we run them locally. We also set a different H2 file name in the datasource URL to create a separate database for this microservice, named `gamification`.

Besides, we'll need to enable CORS for this microservice too since the UI needs to be able to access the leaderboard API. Take a look at the section "Running Our Front End for the First Time" in Chapter 4 if you don't remember what CORS does. This file's contents are identical to the one we added in Multiplication. See Listing 6-18.

Listing 6-18. Adding CORS Configuration to the Gamification App

```
package microservices.book.gamification.configuration;

import org.springframework.context.annotation.Configuration;
import org.springframework.web.servlet.config.annotation.CorsRegistry;
import org.springframework.web.servlet.config.annotation.WebMvcConfigurer;

@Configuration
public class WebConfiguration implements WebMvcConfigurer {

    @Override
    public void addCorsMappings(final CorsRegistry registry) {
        registry.addMapping("/**").allowedOrigins("http://localhost:3000");
    }

}
```

Given that we also want to use the Hibernate's Jackson module, we have to add this dependency in Maven. Remember that we also need to inject the module in the context to be picked up by autoconfiguration. See Listings 6-19 and 6-20.

Listing 6-19. Adding the Jackson's Hibernate Module to Gamification's pom.xml File

```
<dependencies>
<!-- ... -->
<dependency>
    <groupId>com.fasterxml.jackson.datatype</groupId>
    <artifactId>jackson-datatype-hibernate5</artifactId>
</dependency>
</dependencies>
```

Listing 6-20. Defining the Bean for JSON's Hibernate Module to Be Used for Serialization

```
package microservices.book.gamification.configuration;

import com.fasterxml.jackson.databind.Module;
import com.fasterxml.jackson.datatype.hibernate5.Hibernate5Module;
import org.springframework.context.annotation.Bean;
import org.springframework.context.annotation.Configuration;
```

```

@Configuration
public class JsonConfiguration {

    @Bean
    public Module hibernateModule() {
        return new Hibernate5Module();
    }

}

```

Changes in Multiplication Microservice

We finished the first version of the Gamification microservice. Now we have to integrate both microservices together by communicating Multiplication with the new one.

Previously, we created a few REST APIs on the server's side. This time, we have to build a REST API Client instead. The Spring Web module offers a tool with that purpose: the `RestTemplate` class. Spring Boot provides an extra layer on top: the `RestTemplateBuilder`. This builder is injected by default when we use the Spring Boot Web starter, and we can use its methods to create `RestTemplate` objects in a fluent way with multiple configuration options. We can add specific message converters, security credentials if we need them to access the server, HTTP interceptors, etc. In our case, we can use the default settings since both applications are using Spring Boot's predefined configuration. That means that the serialized JSON object sent by our `RestTemplate` can be deserialized without problems on the server's side (the Gamification microservice).

To keep our implementation modular, we create the Gamification's REST client in a separate class: `GamificationServiceClient`. See Listing 6-21.

Listing 6-21. The `GamificationServiceClient` Class, in the Multiplication App

```

package microservices.book.multiplication.serviceclients;

import org.springframework.beans.factory.annotation.Value;
import org.springframework.boot.web.client.RestTemplateBuilder;
import org.springframework.http.ResponseEntity;
import org.springframework.stereotype.Service;
import org.springframework.web.client.RestTemplate;
import lombok.extern.slf4j.Slf4j;
import microservices.book.multiplication.challenge.ChallengeAttempt;
import microservices.book.multiplication.challenge.ChallengeSolvedDTO;

```

```

@Slf4j
@Service
public class GamificationServiceClient {

    private final RestTemplate restTemplate;
    private final String gamificationHostUrl;

    public GamificationServiceClient(final RestTemplateBuilder builder,
                                     @Value("${service.gamification.host}") final
                                     String gamificationHostUrl) {
        restTemplate = builder.build();
        this.gamificationHostUrl = gamificationHostUrl;
    }

    public boolean sendAttempt(final ChallengeAttempt attempt) {
        try {
            ChallengeSolvedDTO dto = new ChallengeSolvedDTO(attempt.getId(),
                    attempt.isCorrect(), attempt.getFactorA(),
                    attempt.getFactorB(), attempt.getUser().getId(),
                    attempt.getUser().getAlias());
            ResponseEntity<String> r = restTemplate.postForEntity(
                gamificationHostUrl + "/attempts", dto,
                String.class);
            log.info("Gamification service response: {}", r.getStatusCode());
            return r.getStatusCode().is2xxSuccessful();
        } catch (Exception e) {
            log.error("There was a problem sending the attempt.", e);
            return false;
        }
    }
}

```

This new Spring @Service can be injected in our existing ones. It uses the builder to initialize the RestTemplate with defaults (just calling build()). It also accepts in the constructor the host URL of the gamification service, which we want to extract as a configuration parameter.

In Spring Boot, we can create our own configuration options in the application.properties file and inject their values in components with the @Value annotation. The gamificationHostUrl argument will be set to the value of this new property, which we have to add to the Multiplication's properties file. See Listing 6-22.

Listing 6-22. Adding the URL of the Gamification Microservice as a Property in Multiplication

```
# ... existing properties

# Gamification service URL
service.gamification.host=http://localhost:8081
```

The rest of the implementation of the service client is simple. It constructs a (new) ChallengeSolvedDTO based on data from the domain object, the ChallengeAttempt. Then, it uses the postForEntity method in RestTemplate to send the data to the /attempts endpoint in Gamification. We don't expect a response body, but the method's signature requires it, so we can set it to String, for example.

We also wrapped the complete logic inside a try/catch block. The reason is that we don't want an error trying to reach the Gamification microservice to end up breaking our main business logic in the Multiplication microservice. This decision is further explained at the end of this chapter.

The ChallengeSolvedDTO class is a copy of the one we created on the Gamification side. See Listing 6-23.

Listing 6-23. The ChallengeSolvedDTO Class Needs to Be Included in the Multiplication Microservice Too

```
package microservices.book.multiplication.challenge;

import lombok.Value;

@Value
public class ChallengeSolvedDTO {

    long attemptId;
    boolean correct;
    int factorA;
    int factorB;
    long userId;
    String userAlias;

}
```

Now we can inject this service in the existing `ChallengeServiceImpl` class and use it to send the attempt after it has been processed. See Listing 6-24 for the modifications required in this class.

Listing 6-24. Adding Logic to `ChallengeServiceImpl` to Send an Attempt to the Gamification Microservice

```
@Slf4j
@RequiredArgsConstructor
@Service
public class ChallengeServiceImpl implements ChallengeService {

    private final UserRepository userRepository;
    private final ChallengeAttemptRepository attemptRepository;
    private final GamificationServiceClient gameClient;

    @Override
    public ChallengeAttempt verifyAttempt(ChallengeAttemptDTO attemptDTO) {
        // ... existing logic

        // Stores the attempt
        ChallengeAttempt storedAttempt = attemptRepository.save(attemptDTO);

        // Sends the attempt to gamification and prints the response
        HttpStatus status = gameClient.sendAttempt(storedAttempt);
        log.info("Gamification service response: {}", status);

        return storedAttempt;
    }

    // ...
}
```

Our test should be also updated to check that the call happens for each attempt. We can add a new mocked class to `ChallengeServiceTest`.

```
@Mock private GamificationServiceClient gameClient;
```

Then, we use Mockito's `verify` in our test cases to make sure this call is performed with the same data as stored in the database.

```
verify(gameClient).sendAttempt(resultAttempt);
```

Besides the REST API client, we want to add a second change to the Multiplication microservice: a controller to retrieve a collection of user aliases based on their identifiers. We need this because the leaderboard API we implemented in the `LeaderBoardController` class returns the score, badges, and position based on user IDs. The UI will need a way to map each ID to a user alias, to render the table in a friendlier manner. See the new `UserController` class in Listing 6-25.

Listing 6-25. The New UserController Class

```
package microservices.book.multiplication.user;

import java.util.List;
import org.springframework.web.bind.annotation.*;
import lombok.RequiredArgsConstructor;
import lombok.extern.slf4j.Slf4j;

@Slf4j
@RequiredArgsConstructor
@RestController
@RequestMapping("/users")
public class UserController {

    private final UserRepository userRepository;

    @GetMapping("/{idList}")
    public List<User> getUsersByIdList(@PathVariable final List<Long> idList) {
        return userRepository.findAllByIdIn(idList);
    }
}
```

This time we use a list of identifiers as a path variable, which Spring splits and passes to us as a standard `List`. In practice, this means that the API call can include one or more numbers separated by commas, e.g., `/users/1,2,3`.

As you see, we're injecting a repository in the controller, so we're not following the three-layer architecture principle here. The reason is that we don't need business logic for this specific use case, so, in these situations, it's better to keep our code simple. If we need business logic any time in the future, we could benefit from the loose coupling between layers and create the service layer in between these two.

The repository interface uses a new query method to perform a select in the users table, filtering those whose identifiers are in the passed list. See the source code in Listing 6-26.

Listing 6-26. The New Query Methods in the UserRepository Interface

```
package microservices.book.multiplication.user;

import java.util.List;
import java.util.Optional;
import org.springframework.data.repository.CrudRepository;

public interface UserRepository extends CrudRepository<User, Long> {

    Optional<User> findByAlias(final String alias);

    List<User> findAllByIdIn(final List<Long> ids);

}
```

Exercise

Update the tests in the Multiplication microservice to cover the new calls to the REST client and create a new one for the UserController. You can find the solutions in this chapter's source code.

UI

The back-end logic is ready, so we can move to the front-end part. We need two new JavaScript classes:

- A new API client to retrieve the leaderboard data from the Gamification microservice
- An additional React component to render the leaderboard

We'll also add a new method to the existing API client to retrieve a list of users based on their IDs.

The GameApiClient class in Listing 6-27 defines a different host and uses the fetch API to retrieve the JSON array of objects. For clarity, we also rename the existing ApiClient to ChallengesApiClient. Then, we include in this one a new method to retrieve the users. See Listing 6-28.

Listing 6-27. The GamiApiClient Class

```
class GameApiClient {
    static SERVER_URL = 'http://localhost:8081';
    static GET_LEADERBOARD = '/leaders';

    static leaderBoard(): Promise<Response> {
        return fetch(GameApiClient.SERVER_URL +
            GameApiClient.GET_LEADERBOARD);
    }
}

export default GameApiClient;
```

Listing 6-28. Renaming the Former Apiclient Class and Including the New Call

```
class ChallengesApiClient {

    static SERVER_URL = 'http://localhost:8080';
    // ...
    static GET_USERS_BY_IDS = '/users';

    // existing methods...

    static getUsers(userIds: number[]): Promise<Response> {
        return fetch(ChallengesApiClient.SERVER_URL +
            ChallengesApiClient.GET_USERS_BY_IDS +
            '/' + userIds.join(',');
    }
}

export default ChallengesApiClient;
```

The returned promises will be used in the new LeaderBoardComponent, which retrieves the data and updates its state's leaderboard attribute. Its render() method should map the array of objects to an HTML table with a row per position. We'll use JavaScript's Timing Events (see <https://tpd.io/timing-events>) to refresh the leaderboard every five seconds with the function setInterval.

See the complete source code of LeaderBoardComponent in Listing 6-29. Then, we'll dive a bit more into its logic.

Listing 6-29. The New LeaderBoardComponent in React

```
import * as React from 'react';
import GameApiClient from '../services/GameApiClient';
import ChallengesApiClient from '../services/ChallengesApiClient';

class LeaderBoardComponent extends React.Component {

  constructor(props) {
    super(props);
    this.state = {
      leaderboard: [],
      serverError: false
    }
  }

  componentDidMount() {
    this.refreshLeaderBoard();
    // sets a timer to refresh the leaderboard every 5 seconds
    setInterval(this.refreshLeaderBoard.bind(this), 5000);
  }

  getLeaderBoardData(): Promise {
    return GameApiClient.leaderBoard().then(
      lbRes => {
        if (lbRes.ok) {
          return lbRes.json();
        } else {
          return Promise.reject("Gamification: error response");
        }
      }
    );
  }
}
```

```
getUserAliasData(userIds: number[]): Promise {
    return ChallengesApiClient.getUsers(userIds).then(
        usRes => {
            if(usRes.ok) {
                return usRes.json();
            } else {
                return Promise.reject("Multiplication: error response");
            }
        }
    )
}

updateLeaderBoard(lb) {
    this.setState({
        leaderboard: lb,
        // reset the flag
        serverError: false
    });
}

refreshLeaderBoard() {
    this.getLeaderBoardData().then(
        lbData => {
            let userIds = lbData.map(row => row.userId);
            this.getUserAliasData(userIds).then(data => {
                // build a map of id -> alias
                let userMap = new Map();
                data.forEach(idAlias => {
                    userMap.set(idAlias.id, idAlias.alias);
                });
                // add a property to existing lb data
                lbData.forEach(row =>
                    row['alias'] = userMap.get(row.userId)
                );
                this.updateLeaderBoard(lbData);
            })
        }
    )
}
```

```
}).catch(reason => {
    console.log('Error mapping user ids', reason);
    this.updateLeaderBoard(lbData);
});
}

).catch(reason => {
    this.setState({ serverError: true });
    console.log('Gamification server error', reason);
});
}

render() {
    if (this.state.serverError) {
        return (
            <div>We're sorry, but we can't display game statistics at this
            moment.</div>
        );
    }
    return (
        <div>
            <h3>Leaderboard</h3>
            <table>
                <thead>
                    <tr>
                        <th>User</th>
                        <th>Score</th>
                        <th>Badges</th>
                    </tr>
                </thead>
                <tbody>
{this.state.leaderboard.map(row => <tr key={row.userId}>
    <td>{row.alias ? row.alias : row.userId}</td>
    <td>{row.totalScore}</td>
    <td>{row.badges.map(
        b => <span className="badge" key={b}>{b}</span>)}
    </td>
)</tr>)
}

```

```

        </tbody>
      </table>
    </div>
  );
}

export default LeaderBoardComponent;

```

The main logic is included in the `refreshLeaderBoard` function. First, it tries to fetch the leaderboard rows from the Gamification server. If it can't (the `catch` clause), it sets the `serverError` flag to true, so we'll render a message instead of the table. In case the data is retrieved normally, the logic performs a second call, this time to the Multiplication microservice. If we get a proper response, we map the user identifiers included in the data to their corresponding aliases and add a new field alias to each position in the leaderboard. In case of a failure in this second call, we still use the original data without the extra field.

The `render()` function differentiates between the error case and the standard case. If there is an error, we show a message instead of the table. This way, we make our application resilient because the main functionality (solving challenges) is working even when the Gamification microservice fails. The leaderboard data is displayed in rows with the user alias (or the ID if it couldn't be fetched), the total score, and the badge list.

We use the `badge` CSS class in the rendering logic. Let's create this custom style in the `App.css` stylesheet. See Listing 6-30.

Listing 6-30. Adding the Badge Style to App.css

```

/* ... existing styles ... */

.badge {
  font-size: x-small;
  border: 2px solid dodgerblue;
  border-radius: 4px;
  padding: 0.2em;
  margin: 0.1em;
}

```

Now, we should include the leaderboard component in our root container, the ChallengeComponent class. See the modifications made to the source code in Listing 6-31.

Listing 6-31. Adding the LeaderBoardComponent Inside the ChallengeComponent

```
import LeaderBoardComponent from './LeaderBoardComponent';

class ChallengeComponent extends React.Component {

    // ...existing methods...

    render() {
        return (
            <div className="display-column">
                {/* we add this just before closing the main div */}
                <LeaderBoardComponent />
            </div>
        );
    }
}

export default ChallengeComponent;
```

Playing with the System

We implemented the new Gamification microservice, connected the Multiplication application to it via a REST API client service, and built the UI to fetch the leaderboard and render it every five seconds.

It's time to play with our complete system. Use your IDE or the command line to start both back-end applications and the Node.js server. If you use the terminal, open three separate instances and run one of the commands in Listing 6-32 in each of them so you have access to all the logs separately.

Listing 6-32. Starting the Apps from the Console

```
/multiplication $ mvnw spring-boot:run
...
/gamification $ mvnw spring-boot:run
...
/challenges-frontend $ npm start
...
```

If everything goes well, we'll see the UI running in our browser. There will be an empty leaderboard (unless you already experimented a bit while coding). If we send a correct attempt, we should see something similar to Figure 6-7.

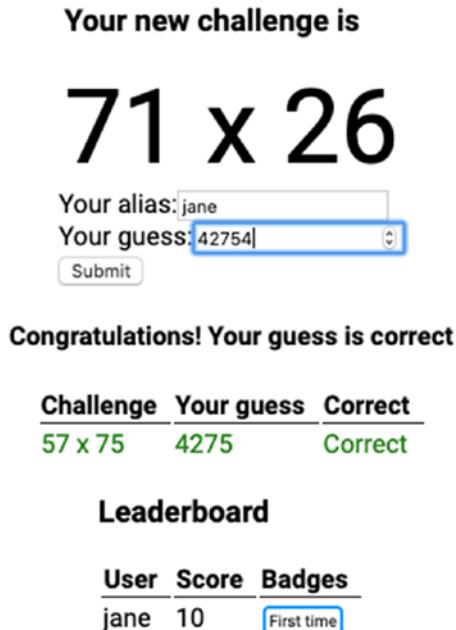


Figure 6-7. UI connected to both microservices

The moment we send a first correct attempt, we'll get 10 points and the "First time" badge. The game works! You can keep playing to see whether you score any metal badges or the Lucky number one. Thanks to the automatic rendering every five seconds, you can even play in multiple browser tabs and the leaderboard will be refreshed in each one of them.

Now let's take a look at the logs. On the Multiplication side, we'll see this line in the logs when we send a new attempt:

```
INFO 36283 --- [nio-8080-exec-4] m.b.m.challenge.ChallengeServiceImpl      :  
Gamification service response: 200 OK
```

The Gamification app will output either a line saying that the attempt was not correct and therefore there is no new score, or this line if you were right:

```
INFO 36280 --- [nio-8081-exec-9] m.b.gamification.game.GameServiceImpl      : User  
jane scored 10 points for attempt id 2
```

We'll also see many repeated log lines showing queries since we configured the app to show all JPA statements, and the UI is making periodic calls to retrieve the leaderboard and the user aliases.

Fault Tolerance

While refining our requirements, we established that the Gamification features are not critical and therefore we could accept some downtime in that part of the system. Let's bring this new microservice down and see what happens. If you still have the applications running, stop the Gamification application. Otherwise, start the UI server and Multiplication only.

We'll see the fallback message of the leaderboard component displayed on the screen, as shown in Figure 6-8. As we can verify using the Network tab in the developer tools, the HTTP calls to the gamification service (on port 8081) are failing.

Your new challenge is

15 x 30

Your alias:

Your guess:

Congratulations! Your guess is correct

Challenge	Your guess	Correct
45 x 76	3420	Correct

We're sorry, but we can't display game statistics at this time.

Figure 6-8. Gamification microservice is down

Besides, if we try to send an attempt, it'll still work. The error causes an exception that is captured within the `GamificationServiceClient` class.

```
ERROR 36666 --- [nio-8080-exec-2] m.b.m.s.GamificationServiceClient : There was a problem sending the attempt.
```

The core functionality is still working even with half of the back end being down. But keep in mind that, in that case, we would miss data, so users will not get any score for successful attempts.

As an alternative implementation, we could have used retry logic. We could implement a loop to keep trying to post the attempt until we get an OK response from the gamification microservice or until a certain amount of time passes. But, even though there are libraries that we can use to implement this pattern, the complexity of our system increases. What if the Multiplication microservice also goes down while retrying? Should we keep track of the attempts that we didn't send yet in the database? In that case, when the Gamification app comes back to life at a random moment, should we send the attempts in the same order as they happened? As you see, distributed systems like our microservices architecture introduce new challenges.

The Challenges Ahead

The system we built is working, so we should be proud of it. Even in the case of a failure in the Gamification microservice, the application keeps responding. See Figure 6-9 for the updated logical view of our system.

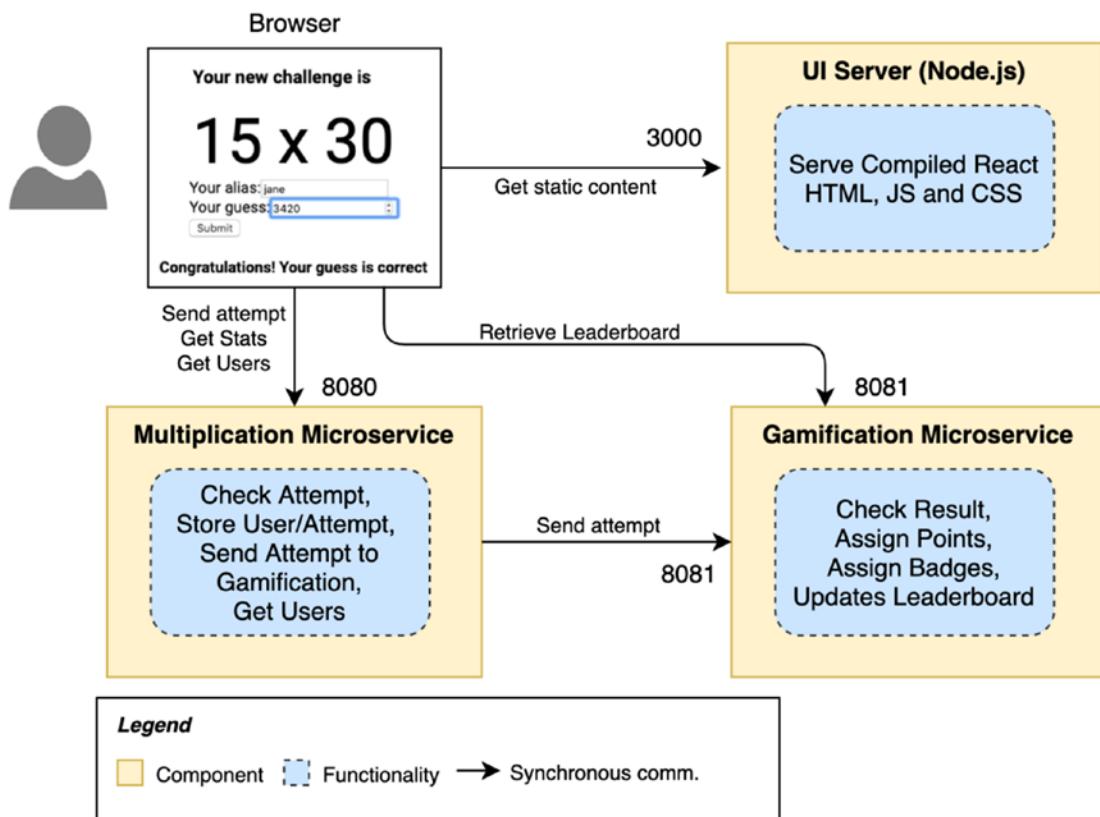


Figure 6-9. Logical view

Our back-end logic is now distributed across these two Spring Boot applications. Let's review the implications of building distributed systems, focusing on our microservice architecture and the new challenges we're facing.

Tight Coupling

When we modeled our domains, we argue that they're loosely coupled because we use only minimal references between the domain objects. However, we introduced awareness of the Gamification logic in the Multiplication microservice. The latter is explicitly calling the Gamification API to send an attempt and is responsible for delivering the message. We're using an imperative style that isn't that bad in a monolith, but it might become a big issue in a microservices architecture because it introduces tight coupling between microservices.

In our current design, the Gamification microservice becomes orchestrated by the Multiplication microservice, which is actively triggering the action. Instead of using this Orchestration pattern, we could use a Choreography pattern and let the Gamification microservice decide when to trigger its logic. We'll detail the differences between Orchestration and Choreography in the next chapter, when we cover event-driven architectures.

Synchronous Interfaces vs. Eventual Consistency

The Multiplication microservice expects the Gamification server to be available when an attempt is sent, as we detailed a bit earlier. If it isn't, that part of the process remains incomplete. All of this happens within the request's lifecycle. By the time the Multiplication server sends the response to the UI, score and badges are either updated or something went wrong. We built synchronous interfaces: the requests remain blocked until they're fully done or they failed.

When you have many microservices, you'll unavoidably have flows that span across them, like in our example, even when they have nicely designed context boundaries. To depict this, let's create a more sophisticated scenario as a hypothetical evolution of our back end. As a first addition, we want to send an email to the user when they reach 1,000 points. Without being judgmental about domain boundaries, let's say we have a dedicated microservice for that, which needs to be updated after assigning a new score. We also add a microservice that collects data for reporting and needs to be connected to both Multiplication and Gamification. See Figure 6-10 for a complete view of this hypothetical system.

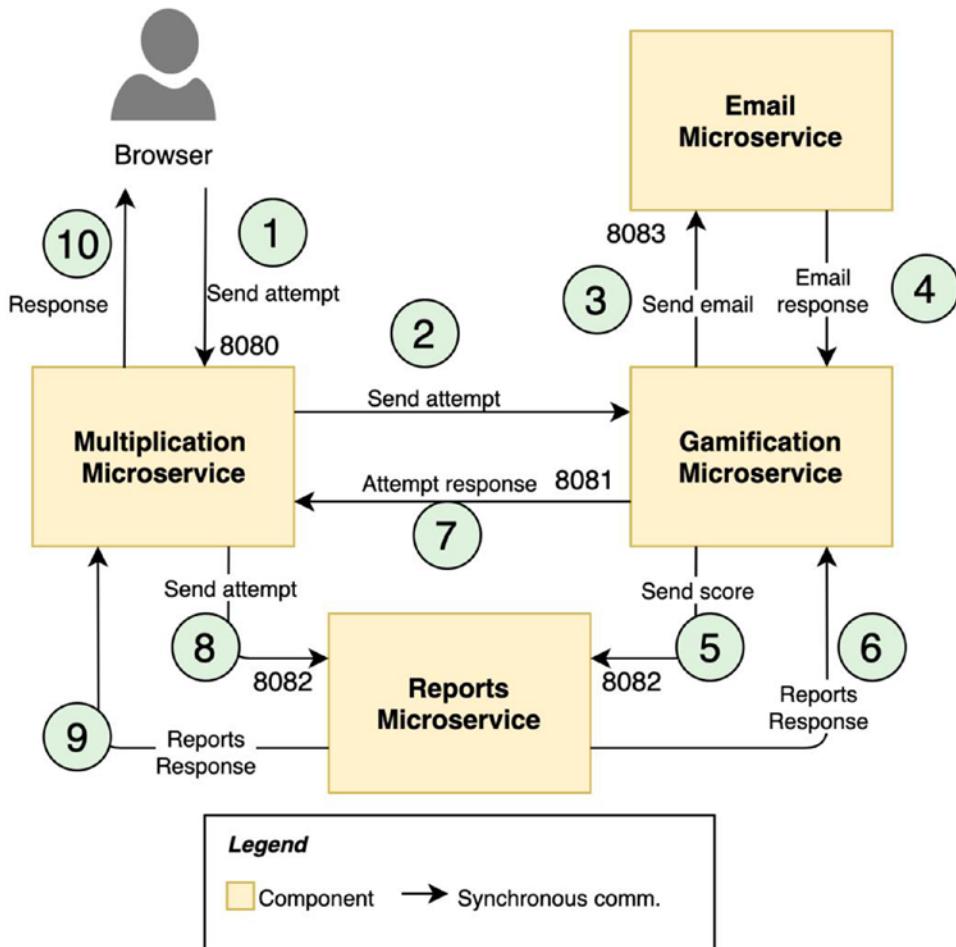


Figure 6-10. Hypothetical evolution of our system

We could keep building synchronous interfaces with REST API calls. Then, we would have a chain of calls as shown in the figure's sequence of numbers. The request from the browser would need to wait until all of them are completed. The more services we have in the chain, the longer the request will be blocked. If one microservice is slow, the whole chain becomes slow. The overall performance of the system is at least as bad as the worst microservice in the chain.

Synchronous dependencies are even worse when we don't build the microservices with fault tolerance in mind. In our example, a simple failing update operation from the Gamification microservice to the Reports microservice would potentially crash the entire flow. If we implement a retry mechanism within the same blocking thread, the performance degrades even more. If we let them fail too easily, we may end up with a lot of partially completed operations.

There is a clear conclusion thus far: synchronous interfaces introduce a strong dependency between microservices.

As an advantage, we know that the reports are updated in the back end by the time the user gets a response. So, it's the score. We even know if we could send an email or not, so we can give immediate feedback.

In a monolith, we wouldn't face this challenge because all these modules would live within the same deployable unit. We don't have issues due to network latency or errors if we're just calling other methods. Besides, if something goes down, it's going to be the whole system, so we don't need to design it while taking into account fine-grained fault tolerance.

So, if synchronous interfaces are bad, the important question is: do we need to block the complete request in the first place? Do we need to know that everything was completed before returning a response? To answer that question, let's modify our hypothetical case to detach the subsequent interactions between microservices. See Figure 6-11.

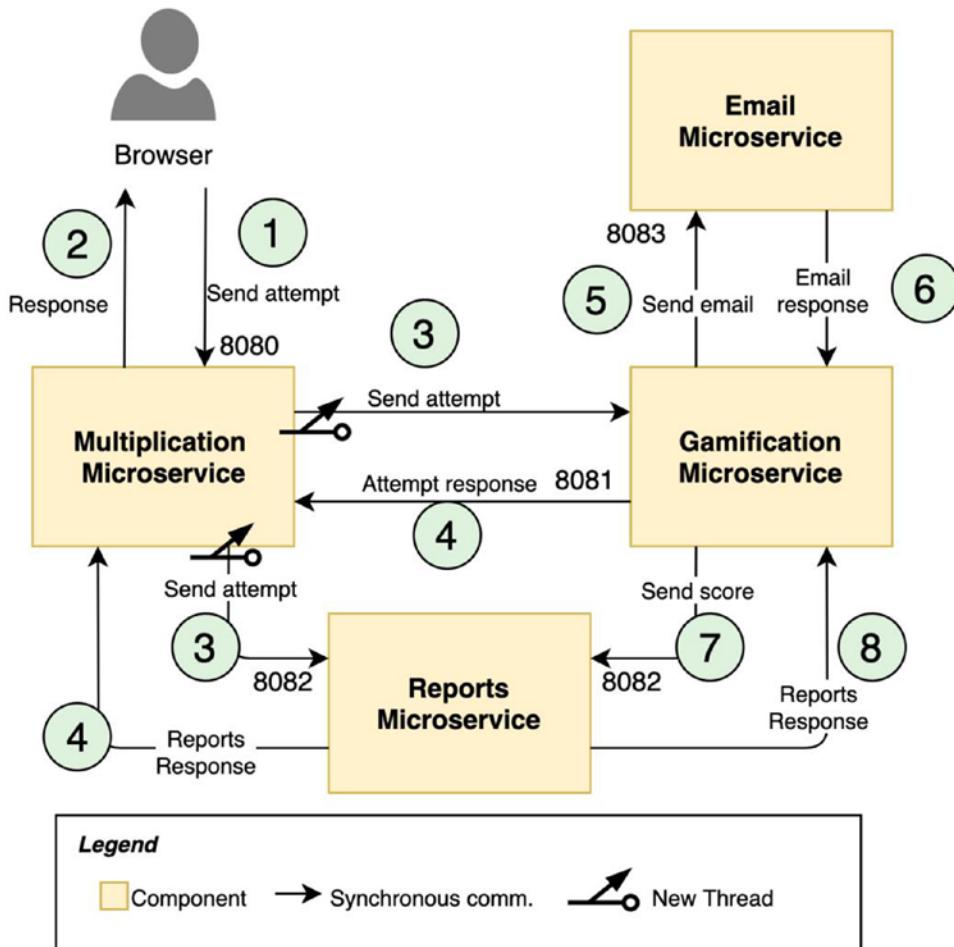


Figure 6-11. Asynchronous processing

This new design initiates some requests in new threads, unblocking the main one. We could use, for example, Java Futures. That would cause the response to be delivered to the client much earlier so we solve all the problems described before. But, as a consequence, we introduced eventual consistency. Imagine that, on the API client's side, there is a single sequential thread waiting for the response of sending the attempt. Then, this client's process will try to collect the score and the report. In the blocking-thread scenario, our API clients (e.g., the UI) know for sure that, after getting the response from Multiplication, the score in Gamification is consistent with the attempt. In this new asynchronous landscape, we can't guarantee that. If our network latency is good, the client might get the updated score. But maybe it takes one second to complete, or we have services down for a longer period, and it's updated only after some retries. We can't predict it.

Therefore, one of the hardest challenges we face when building microservice architectures is embracing eventual consistency. We should accept that the Gamification microservice's data might not be consistent with the Multiplication microservice's data at a given moment in time. It'll only be eventually consistent. In the end, with a proper design that makes our system robust, the Gamification microservice will be up-to-date. In the meantime, our API clients can't assume consistency between different API calls. And that's the key: it's not only our back-end system; it's also about our API clients. If we're the only ones consuming our APIs, that might not be a big issue: we can develop our REST clients having eventual consistency in mind. However, if we offer our APIs as a service, we have to educate our clients too. They have to know what to expect.

Thus, our original question about whether we need to block the request or not can be replaced by a more important question: can our system be eventually consistent? Of course, the answer depends on the functional and technical requirements we have.

For example, in some cases, the functional description of a system might imply strong consistency, but you can adapt it without a big impact. As a practical case, if we detach the email subflow as an asynchronous step, we could change the message prompted to the user from "You should have received an email with instructions" to "You will receive an email with instructions in a few moments. If you don't, please contact customer support." But being able to make changes like this always depends on the requirements and the appetite of the organization to embrace eventual consistency.

Microservices Are Not Always the Best Solution (Part I)

If your project's requirements are not compatible with eventual consistency across domains, a modular monolithic application might suit you better.

On the other hand, we don't need to go fully asynchronous everywhere. There are some cases where it makes sense to have synchronous calls between microservices. That's not a problem nor a reason to make a drama of our software architecture. We just need to keep an eye on those interfaces since sometimes it's a symptom of tight coupling between domains. In that case, we could consider merging them into the same microservice.

Looking back to our current system status, we can conclude that it's ready for eventual consistency. Since we don't rely on the response to refresh the leaderboard, we could switch to an asynchronous call between our microservices without any impact.

As you can imagine, there is a better way to implement asynchronous communication between microservices than having REST API calls with a retry pattern. We'll cover it in the next chapter.

Transactions

In a monolith, we could use the same relational database to store users, attempts, scores, and badges. Then, we could benefit from database transactions. We would get the ACID guarantees we covered briefly in the previous chapter: atomicity, consistency, isolation, and durability. In the case of an error saving scorecards, we could revert all previous commands within the transaction so the attempt wouldn't be stored either. That operation is known as a *rollback*. We could ensure data integrity at all times since we could avoid partial updates.

We can't have ACID guarantees across microservices because we can't achieve real transactions in a microservices architecture. They are deployed independently, so they live in different processes, and their databases should also be decoupled. Besides, to avoid interdependencies, we also concluded that we should accept eventual consistency.

Atomicity, or making sure that either all related data is stored or nothing, is hard to achieve across microservices. In our system, the first request stores the attempt, and then the Multiplication microservice calls the Gamification microservice to do its part. Even if we would keep that request synchronous, we never know if the score and badges were stored if we don't receive a response. What do we do then? Do we roll back the transaction? Do we always store the attempt no matter what happens in Gamification (as we did)?

In fact, there are imaginative—and complex—ways to try to achieve transaction rollbacks in a distributed system.

- *Two-phased commits (2PC)*: In this approach, we could send the attempt from Multiplication to Gamification, but we wouldn't store the data yet on either side. Then, once we get the response indicating that data is ready to be stored, we send a second request as a signal to store the score and badges on Gamification, and we store the attempt on Multiplication. With these two phases (prepare and commit), we minimized the time when something can go wrong. Yet we didn't eliminate the possibility since the second phase might fail anyway. In my opinion, this is a horrible idea since we have to stick to synchronous interfaces, and the complexity grows exponentially.

- *Sagas*: This design pattern involves two-way communication. We could build an asynchronous interface between both microservices, and if something goes wrong on the Gamification side, this microservice should be able to reach the Multiplication microservice to let it know. In our example, Multiplication would then delete the attempt that was just saved. That way we *compensate* a transaction. This comes with a high price in terms of complexity as well.

Undoubtedly, the best solution is to try to keep the functional flows that must use database transactions within the same microservice. If we can't split a transaction because it's critical in our system, it looks like the process should belong to the same domain anyway. For other flows, we can try to split the transaction boundaries and embrace eventual consistency.

We can also apply patterns to make our system more robust, so we'll minimize the risk of having partially executed operations. Any design pattern that can ensure the delivery of data between microservices will help with that goal. That will be covered in the next chapter as well.

Our system doesn't use distributed transactions. It doesn't require them either since we don't need immediate consistency between the attempts and the score. But there is still a design flaw: the Multiplication microservice ignores errors from Gamification so we might get successfully solved attempts without their corresponding score and badges. We'll improve that soon without needing to implement a retry mechanism ourselves.

Microservices Are Not Always the Best Solution (Part II)

If you find yourself implementing distributed transactions with 2PC or sagas all over the place, you should take some time to reflect on your requirements and your microservice boundaries. You might want to merge some of them or make a better distribution of functionalities. If you can't fix it in a simpler way, consider a modular monolithic application with a single relational database.

API Exposure

We created a REST endpoint in the Gamification microservice that was intended for the Multiplication microservice. But the UI also needs access to the Gamification microservice so, in fact, anybody could access it. Smart users could send fake data to the Gamification microservice if they use an HTTP client (like HTTPie). We would be in a bad situation since that would break our data integrity. Users could score points and get badges without the corresponding attempts stored on the Multiplication side.

There are multiple ways of solving this problem. We could think of adding a security layer to our endpoints and make sure that internal APIs are available only for other back-end services. A simpler option is to use a reverse proxy (with the Gateway pattern) to make sure that we expose only the public endpoints. We'll cover this option in more detail in Chapter [8](#).

Summary and Achievements

In this chapter, we looked into the reasons to move to a microservices architecture. We started detailing the approach we have followed so far, a small-monolith architecture, and analyzed what would be the pros and cons of continuing our journey toward a modular, monolithic application, compared to making the transition to microservices.

We also examined how a small monolith can help you define your domains better and complete faster the first version of our product to get early feedback from your users. The list of good practices to structure your code in modules should help you make a split in case you need it. But we also saw how sometimes a small monolith is not the best idea, especially if the development team is large from the beginning.

Making the decision to move to microservices (or start with them) requires a deep analysis of the functional and nonfunctional characteristics of your system to figure out requirements in terms of scalability, fault-tolerance, transactionality, eventual consistency, etc. That decision can be crucial for the success or the failure of a software project. I hope that all the considerations included in this chapter, and supported by the practical case, help you scrutinize all factors that are present in your project and make a sound decision and a good plan if you make the move.

As expected from this book, we decided to go for a microservices architecture. On the practical side, we navigated through the layers of the new Gamification application: services, repositories, controllers, and the new React component. We connected

Multiplication to Gamification using a simple, *imperative* approach, and we used this interface between our microservices to discover some of the new challenges we face with a microservices architecture.

By the end of this chapter, we also concluded that the synchronous interface we chose for communicating both microservices was the wrong decision. It introduces tight coupling and makes our architecture fragile against errors. This is a perfect baseline for the next chapter, where we'll introduce the advantages of an event-driven architecture.

Chapter's Achievements:

- You saw how a small-monolith approach can help you when starting new projects.
- You had a first contact with the pros and cons of microservice architectures (you'll keep learning them over the next chapters).
- You understood the differences between synchronous and asynchronous processing in distributed systems and how they relate to eventual consistency.
- You learned why it's important to embrace those new paradigms—asynchronous processes, eventual consistency—in microservice architectures to avoid tight coupling and domain pollution.
- You saw why microservices are not the best solution for all cases (e.g., if you need transactionality and immediate data consistency).
- You identified the first challenges we're facing in our practical case and saw how the current implementation is not the right way of implementing microservices.

CHAPTER 7

Event-Driven Architectures

In the previous chapter, we analyzed how the interface between our microservices plays a key role concerning tight coupling. The Multiplication microservice calls the Gamification microservice, becoming the orchestrator of the process. If there were other services that also needed to retrieve the data for each attempt, we would need to add extra calls from the Multiplication application to these services, thus creating a distributed monolith with a central brain. We covered this problem in detail when we examined a hypothetical extension of our back end.

We'll look in this chapter at a different way of designing these interfaces, based on the Publish-Subscribe pattern. The method is called an *event-driven architecture*. Instead of addressing the data to a specific destination, *publishers* classify and send events without any knowledge of the parts of the system that receive them, the *subscribers*. These event consumers don't need to be aware of the publishers' logic either. This change of paradigm makes our system loosely coupled and scalable, but it also brings new challenges to our system.

Our goal for this chapter is to understand the core concepts of event-driven architectures, their advantages, and the consequences of working with them. As usual, we'll apply this knowledge to our system following a hands-on approach.

Core Concepts

This section highlights the core concepts of event-driven architectures.

The Message Broker

A key element in an event-driven architecture is the *message broker*. In this type of architecture, the system components communicate with the broker instead of connecting directly to each other. That's how we keep the loose coupling between them.

Message brokers normally include routing functionalities. They allow creating multiple “channels,” so we can separate messages based on our requirements. One or more publishers may generate messages in each of these channels, and these messages may be consumed by one or more subscribers (or even none). Within this chapter, we'll see in more detail what a message is and the different messaging topologies you may want to use. See Figure 7-1 for a conceptual view of some typical scenarios using a message broker.

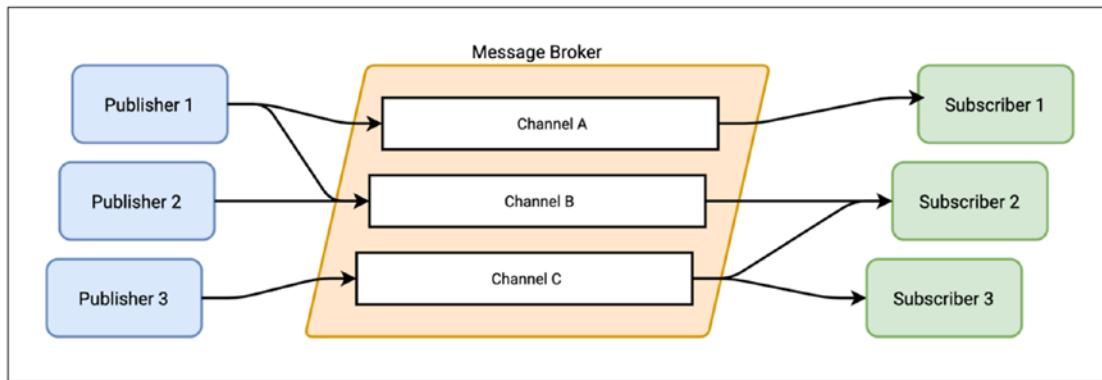


Figure 7-1. Message brokers: high-level view

These concepts are not new at all. Developers who have been active for a while now surely identify similar patterns in enterprise service bus (ESB) architectures. The bus pattern facilitates the communication between the different pieces of a system, providing data transformation and mapping, message queuing and sorting, routing, etc.

There is still a bit of controversy about the exact differences between ESB architectures and those based on message brokers. A broadly accepted distinction is that, in ESB, the channel itself has much more relevance in the system. The service bus sets the protocol standards for communication, and it transforms and routes the data to the specific target. Some implementations can take care of distributed transactions. In some cases, they even have a sophisticated UI to model business processes and

transform these rules into configuration and code. Typically, ESB architectures tend to concentrate a big part of the system's business logic inside the bus, so they become the system's orchestration layer. See Figure 7-2.

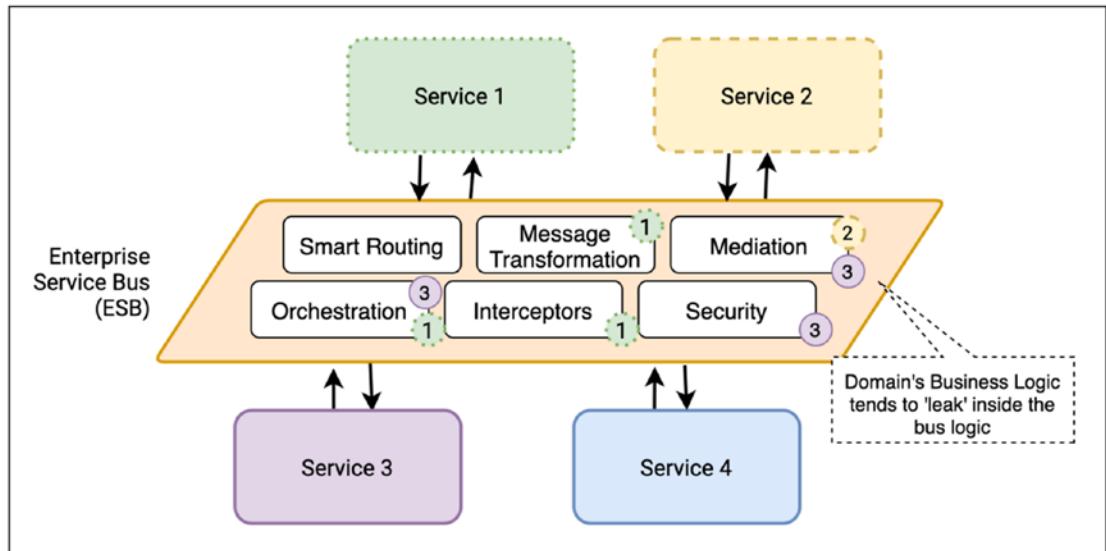


Figure 7-2. ESB architectures concentrate business logic inside the bus

Moving all the business logic inside the same component and having a central orchestrator in the system are software architecture patterns that tend to fail. Systems that follow that route have a single point of failure, and their core part (the bus in this case) becomes harder to maintain and evolve over time since the whole organization depends on it. The logic embedded in the bus tends to become a big mess. That's one of the reasons why the ESB architectures got such a bad reputation over the last years.

Based on these bad experiences, many people tend now to move away from this centrally orchestrating, too-smart messaging channel and implement a simpler approach with a message broker, using it just for the communication between different components.

At this point, you might be picturing a clear line between ESB as complex channels and message brokers as simple channels. However, I mentioned earlier that there is a bit of controversy, so it's not that easy to draw that line. On one hand, you could use an ESB platform but keep the business logic properly isolated. On the other hand, some modern messaging platforms such as Kafka offer tools that allow you to embed some logic in the channel. You can transform messages with functions that may include the business logic

if you want. You can also query data in the channel like you would do with a database, and you can process the output as you want. For example, based on some data included in a message, you could decide to take it out of a specific channel and move it to another one, with a different format. Therefore, you can switch between tools that are normally associated with different architecture patterns (ESB/message brokers) but still use them similarly. This idea already gives us an early introduction of a core takeaway of the coming chapters: first you need to understand the patterns, and then you can choose the tool that best suits your needs.

I recommend you avoid including business logic in the communication channel as much as possible. Keep that logic where it belongs in your distributed system, respecting a domain-driven design approach. That's what we'll do in our system: we'll introduce a message broker to keep our services loosely coupled and scalable, keeping the business processes inside each microservice.

Events and Messages

In event-driven architectures, an *event* indicates that something happened in the system. Events get published to a messaging channel (e.g., a message broker) by the business logic that owns the domain where those events happened. Other components in the architecture that are interested in a given event type subscribe to the channel to consume all the subsequent event instances. As you can see, events relate to the publish-subscribe pattern, so they are linked to message brokers or buses too. We'll implement an event-driven architecture using a message broker, so let's focus on that specific case.

A *message*, on the other hand, is a more generic term. Many people make a distinction between messages, as elements that are directly addressed to a system component, and events, as pieces of information that reflect facts that happened at a given domain and don't have a specific addressee. However, an event is actually a message from a technical perspective when we send it via a message broker (because there is no such thing as an event broker). To keep it simple, we'll use in the book the term *message* to refer to a generic piece of information that goes through a message broker, and we'll use *event* when we refer to a message that follows an event-driven design.

Note that there is nothing that prevents us from modeling events and sending them using REST APIs (similarly to what we did in our application). However, that doesn't help reduce tight coupling: the producers need to be aware of the consumers to target the events to them.

When we use events with a message broker, we can better isolate all the components in our software architecture. Publishers and subscribers don't need to be aware of each other. This fits nicely into a microservice architecture because we want to keep microservices as independent as possible. With this strategy, we could introduce new microservices that consume events from the channels without needing to modify the microservice that publishes those events, nor other subscribers.

Thinking in Events

Bear in mind that the introduction of a message broker and some classes with the suffix Event don't make our architecture "event driven" automatically. We have to design our software thinking in events, and that requires effort if we're not used to it. Let's use our application to analyze this a bit deeper.

In the first scenario, imagine that we would have created a Gamification API to assign scores and badges to a given user. See the top part of Figure 7-3. Then, the Multiplication microservice would call this API, updateScore, not only becoming aware of this microservice's existence but also becoming the owner of part of its business logic (by assigning a score for a solved attempt). This is a common mistake that people make when they start with microservice architectures and come from an imperative programming style. They tend to change method calls by API calls between microservices, implementing a remote procedure call (RPC) pattern, sometimes without even noticing it. In an attempt to improve the coupling between microservices, we could introduce a message broker. Then, we replace the REST API call with a message directed to the Gamification microservice, the UpdateScore message. But, would we improve the system with this change? Not much. The message still has a specific destination, so it can't be reused by any new microservices. Besides, both parts of the system remain tightly coupled, and, as a side effect, we replaced the synchronous interface with an asynchronous one, introducing extra complexity (as we saw in the previous chapter and will also elaborate further in this one).

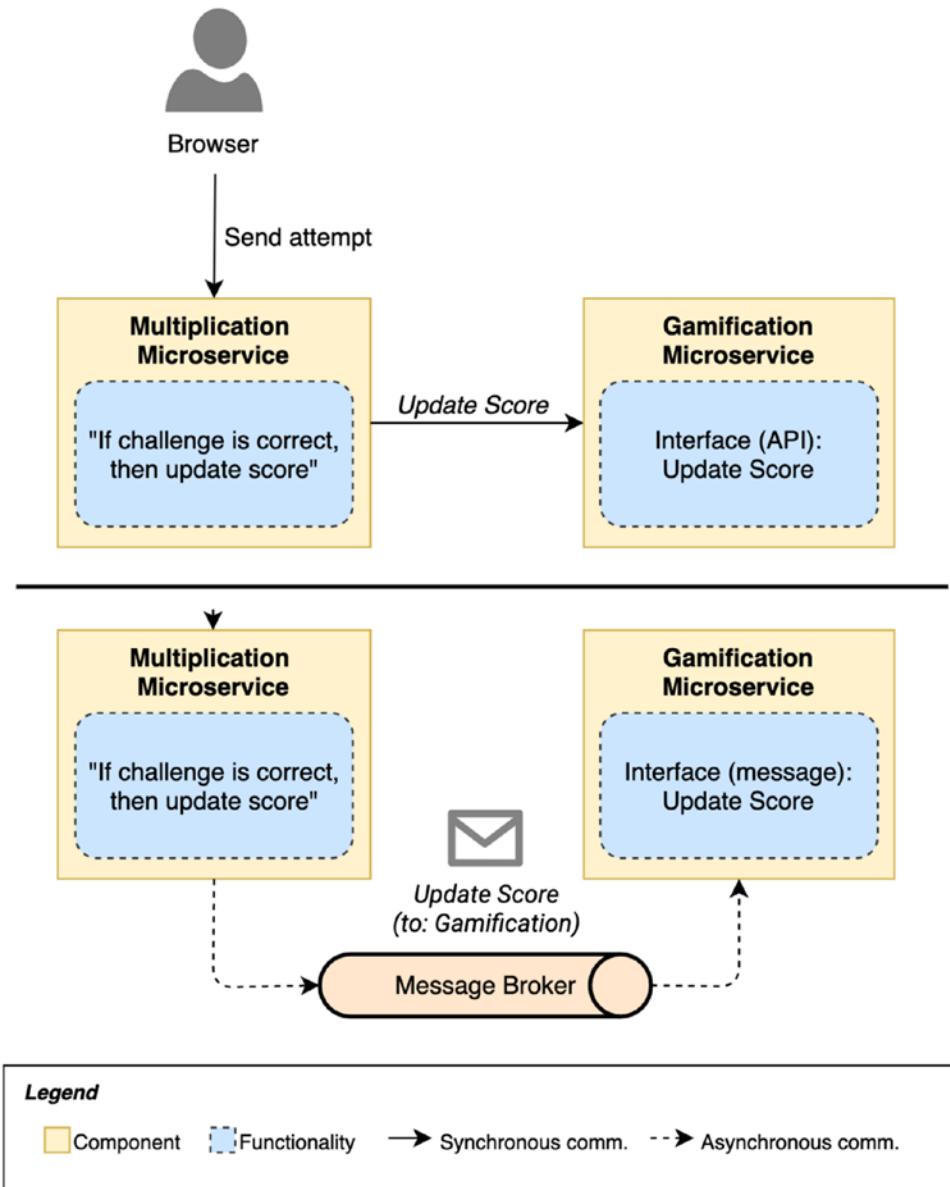


Figure 7-3. Imperative approach: REST vs. message

The second scenario is based on our current implementation. See Figure 7-4. We pass a `ChallengeSolvedDTO` object from Multiplication to Gamification, so we respect our domain boundaries. We don't include gamification logic in the first service. However, we still need to address Gamification directly, so the tight coupling remains. With the introduction of a message broker, we could solve this problem. The Multiplication

microservice could send a `ChallengeSolvedDTO` to a generic channel and continue doing its logic. Our second microservice could subscribe to this channel and process the message (which is already an event conceptually) to calculate the new score and badges. New microservices added to the system could transparently subscribe to the channel if they are also interested in the `ChallengeSolvedDTO` message to, for example, generate reports or send messages to the user.

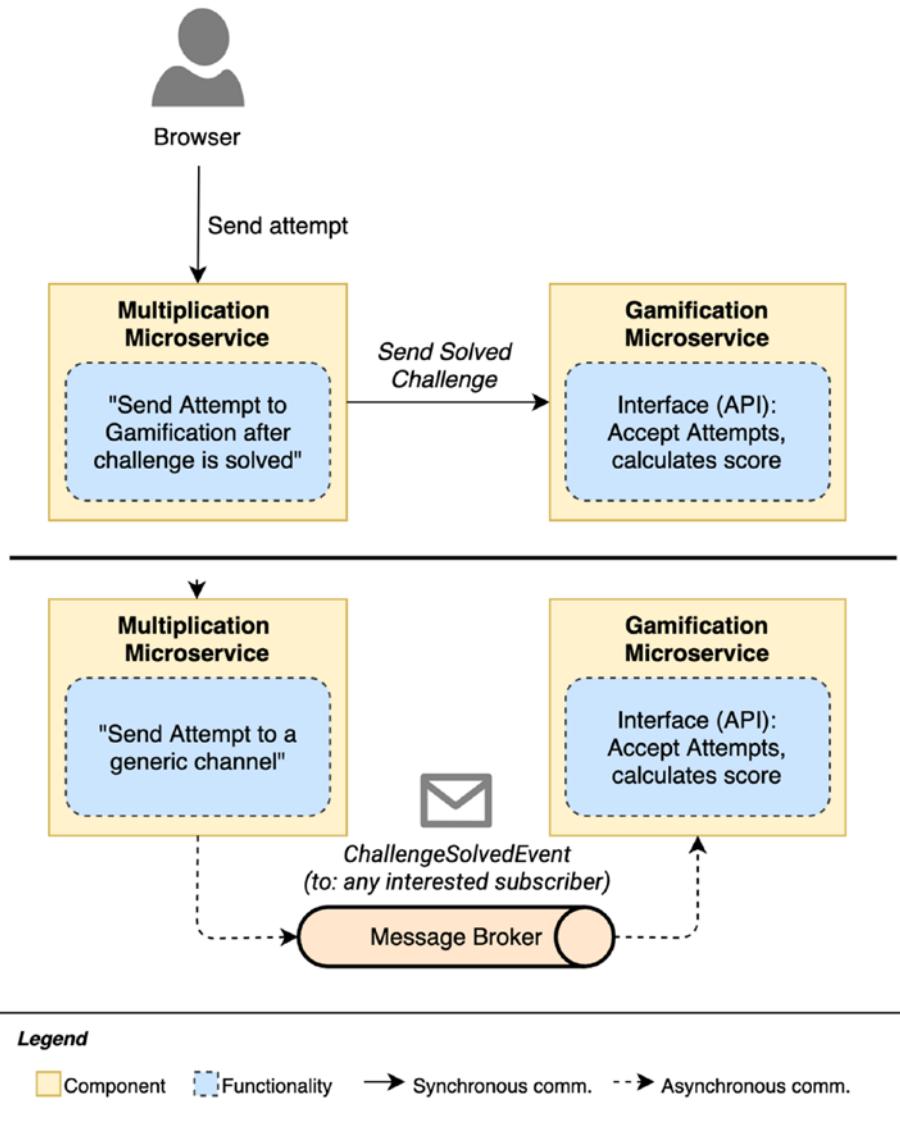


Figure 7-4. Events: REST vs. message

Our first scenario implements a command pattern, where the Multiplication microservice is instructing what to do to the Gamification microservice (a.k.a. orchestration). The second scenario implements the event pattern, by sending a notification about something that already happened, together with contextual data. The consumers will process this data, which may trigger their business logic and possibly other events as a result. This approach is sometimes called *choreography*, as opposed to orchestration. When we base our software architecture on these event-driven designs, we refer to it as an *event-driven architecture*.

As you see, to achieve a real event-driven architecture, we have to rethink business processes that may be expressed in an imperative style and define them instead as (re) actions and events. Not only we should define domains using DDD, but we should also model interactions between them as events. If you want to know more about a technique to help you conduct these design sessions, check <https://tpd.io/event-storming>.

Before moving forward, let me insist again on an important remark: you don't need to change every single communication interface in your system to follow an event-driven style. You probably need to implement the command and the request/response patterns in some cases where events don't fit. Don't try to force a business requirement that only fits as a command to artificially behave as an event. On the technical side, don't be afraid of using REST APIs for use cases where they make more sense, like commands that need a synchronous response.

Microservices Are Not Always the Best Solution (III)

When you build a microservice architecture that uses mostly imperative, targeted interfaces, you have a lot of hard dependencies between all these system components. Many people refer to this scenario as a *distributed monolith*, since you still have the disadvantages of a monolithic application: tight coupling and therefore less flexibility to modify microservices.

If you need some time to build an event-driven mindset in your organization, you could rather set up a modular system and start implementing event patterns across the modules. Then, you benefit from learning one thing at a time and keeping a manageable complexity. Once you achieve loose coupling, you can split the modules into microservices.

Asynchronous Messaging

In the previous chapter, we dedicated a section to analyzing the impact of changing synchronous interfaces into asynchronous ones. With the introduction of a message broker as a tool to build an event-driven architecture, the adoption of asynchronous messaging is implicit. Publishers send events and don't wait for a response from any event consumer. That will keep our architecture loosely coupled and scalable. See Figure 7-5.

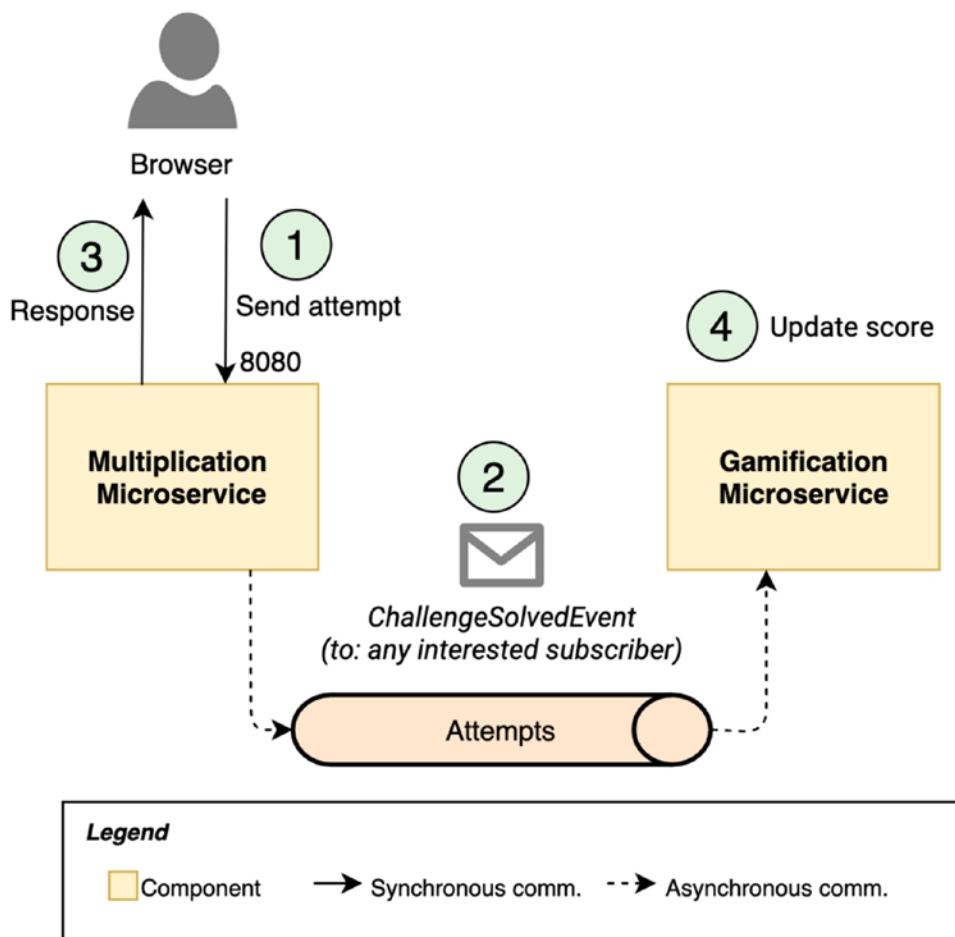


Figure 7-5. Asynchronous process with a message broker

However, we could also use a message broker and keep our processes synchronous. Let's use our system as an example again. We plan to replace the REST API interface by a message broker. Yet, instead of creating a single channel to send our events, we could create two and use the second one to receive a response from the Gamification microservice. See Figure 7-6. In our code, we could then block the request's thread and wait for that acknowledgment before continuing the process.

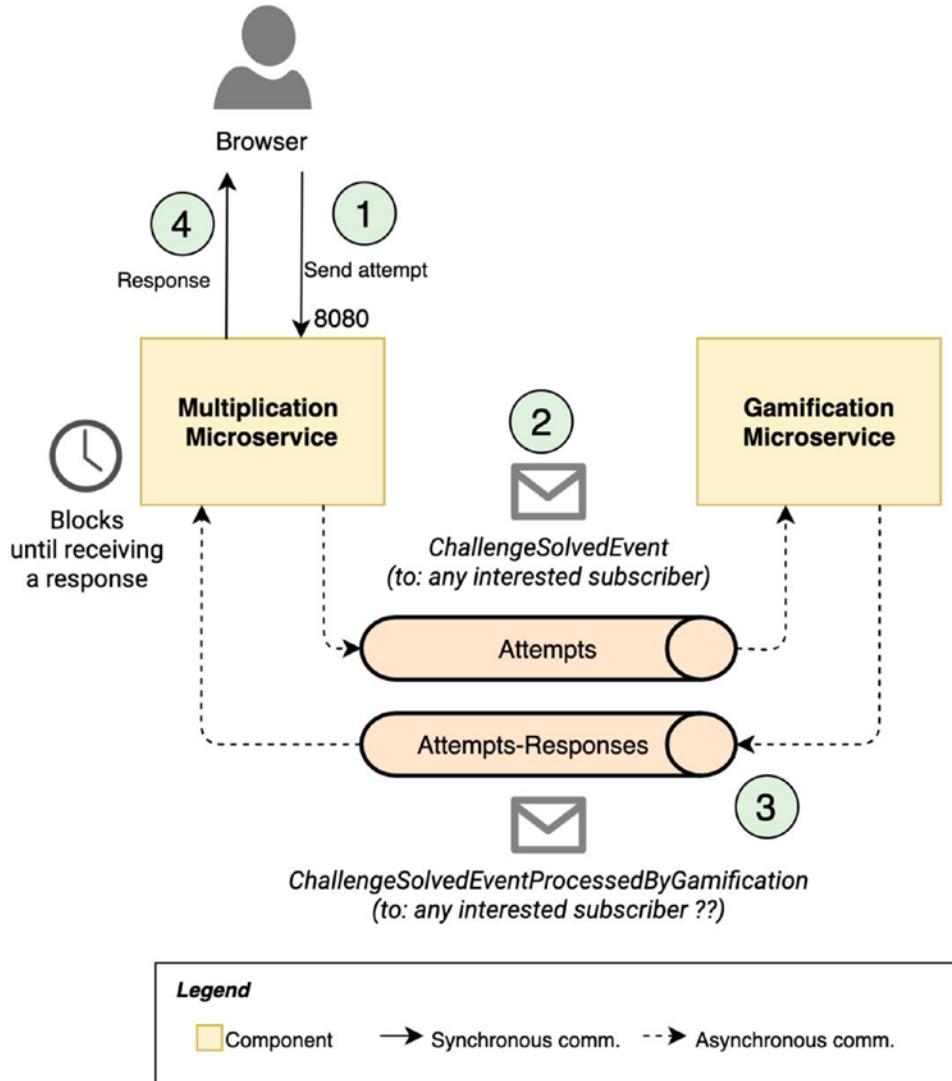


Figure 7-6. Synchronous processing with a message broker

That's actually a request/response pattern on top of a message broker. This combination can be useful in some use cases, but it's not recommended in an event-driven approach. The main reason is that we get the tight coupling again: the Multiplication microservice needs to know about the subscribers and how many they are to make sure it receives all the responses. We still get some advantages like scalability (as we'll detail later), but there are other patterns we can apply to improve the scalability with synchronous interfaces like a load balancer (as we'll see in the next chapter). Therefore, we could consider using a simpler synchronous interface like a REST API, in situations where our process needs to be synchronous anyway. See Table 7-1 for a summary of how you could combine patterns and tools. Keep in mind that this is just a recommendation. As we have analyzed already, you may have your own preferences to implement these patterns using different tooling.

Table 7-1. Combining Patterns and Tools

Pattern	Type	Implementation
Request/response	Synchronous	REST API
Commands that require blocking	Synchronous	REST API
Commands that don't require blocking	Asynchronous	Message broker
Events	Asynchronous	Message broker

It's worth noting that, even though the end-to-end communication can be asynchronous, we'll get a synchronous interface with the message broker from our applications. That's an important characteristic. When we publish a message, we want to be sure the broker received it before continuing with something else. The same applies to subscribers, where the broker requires acknowledgment after consuming messages to mark them as processed and move to the next ones. These two steps are critical to keep our data safe and make our system reliable. We'll explain these concepts using our practical case later in this chapter.

Reactive Systems

The word *Reactive* can be used in multiple contexts, having a different meaning depending on the technical layer that refers to. The most accepted definition of a *reactive system* describes it as a set of design principles to apply in software architecture to make the system responsive (responds on time), resilient (stays responsive if there are failures), elastic (adapts to be responsive under different workloads), and message-driven (ensures loose coupling and boundary isolation). These design principles are listed in the Reactive Manifesto (<https://tpd.io/rmanifesto>). We'll follow these patterns while building our system, so we can claim we're building a reactive system.

On the other hand, *reactive programming* refers to a set of techniques used in programming languages around patterns such as futures (or promises), reactive streams, backpressure, etc. There are popular libraries that help you implement these patterns in Java, like Reactor or RxJava. With reactive programming, you can split your logic into a set of smaller blocks that can run asynchronously and later compose or transform the result. That brings a concurrency improvement since you can go faster when you do tasks in parallel.

Switching to reactive programming doesn't make your architecture reactive. They work at different levels: reactive programming helps achieve improvements inside components and in terms of concurrency. Reactive systems are changes at a higher level, between components, that help build loosely coupled, resilient, and scalable systems. See <https://tpd.io/react-sys-prg> for more details on the differences between both techniques.

Pros and Cons of Going Event-Driven

In the previous chapter, we went through the pros and cons of moving to microservices. We gain in flexibility and scalability, but we face new challenges such as eventual consistency, fault tolerance, and partial updates.

Going event-driven with a message broker pattern help with these challenges. Let's briefly describe how, using our practical example.

- *Loose coupling between microservices:* We already figured out how we can make the Multiplication service unaware of the Gamification service. The first sends an event to the broker, and Gamification subscribes and reacts to the event, updating score and badges for a user.

- *Scalability:* As we'll see in this chapter, it's easy to add new instances of a given application to scale up our system horizontally. Moreover, it would be easy to introduce new microservices in our architecture. They could subscribe to events and work independently, for example in the hypothetical situation we analyzed: we could generate reports or send emails based on events triggered by existing services.
- *Fault tolerance and eventual consistency:* If we make the message broker reliable enough, we can use it to guarantee eventual consistency even when system components fail. If the Gamification microservice goes down for some time, it could catch up with the events later when it comes back since the broker can persist the messages. That gives us some flexibility. We'll see this in practice at the end of this chapter.

On the other hand, adopting event-based design patterns confirms our choice for eventual consistency. We avoid creating blocking, imperative flows. Instead, we use asynchronous processes that simply notify other components. This, as we saw, requires a different mindset, so we (and possibly our API clients) accept that the state of the data might not be consistent across microservices all the time.

Additionally, with the introduction of the message broker, we're adding a new component to our system. We can't simply say that the message broker doesn't fail, so we have to prepare the system for new potential errors.

- *Dropped messages:* It might be the case that the `ChallengeSolvedEvent` never reaches Gamification. If you're building a system where you shouldn't miss an event, you should configure the broker to fulfill the *at-least-once* guarantee. This policy ensures the messages are delivered at least once by the broker, although they could be duplicated.
- *Duplicated messages:* The message broker, under certain situations, may send more than once some messages that were published only once. In our system, if we get the event twice, we'll increment the score incorrectly. Therefore, we have to think about making the event consumption *idempotent*. In computing, an operation is idempotent if it can be called more than once without different outcomes. A possible solution in our case would be marking the events that we

already processed on the Gamification side (e.g., in the database) and ignore any repeated ones. Some brokers like RabbitMQ and Kafka also offer a good *at-most-once* guarantee if we configure them properly, and that helps prevent duplicates.

- *Unordered messages:* Even though the broker can try its best to avoid unordered messages, this can still happen if something fails or maybe due to a bug in our software. We have to prepare our code to cope with that. When possible, try to avoid the assumption that the events will be consumed in the same order in which they were published in time.
- *Broker's downtime:* In the worst case, the broker can also become unavailable. Both publishers and subscribers should try to deal with that situation (e.g., with a retry strategy or a cache). We could also flag the services as *unhealthy* and stop accepting new operations (as we'll cover in the next chapter). That could imply downtime for the complete system, but might be a better option than accepting partial updates and inconsistent data.

These example solutions proposed in each of the previous bullets are *resilience patterns*. Some of them could be translated into coding tasks that we should do to make our system work even in case of failures, such as idempotency, retries, or health checks. As we mentioned already, a good resilience is important in distributed systems such as a microservice architecture, so it's always handy to know these patterns to bring solutions for the unhappy flows during your design sessions.

Another liability of an event-driven system is that *traceability* becomes harder. We call a REST API, and that might trigger events; then, there could be components reacting to those events, subsequently publishing some other events, and the chain continues. Knowing what event caused what action in a different microservice might not be a problem when we only have a few distributed processes. However, when the system grows, having an overall view of these chains of events and actions is a big challenge in an event-driven microservice architecture. We need this view because we want to be able to debug operations that go wrong and find out why we trigger a given process. Fortunately, there are tools to implement *distributed tracing*: a way for us to link events and actions and visualize them as a chain of actions/reactions. For example, the Spring family has Spring Cloud Sleuth, a tool that automatically injects some identifiers (span

IDs) in the logs and propagates those identifiers when we make/receive HTTP calls, publish/consume messages via RabbitMQ, etc. Then, if we use centralized logging, we can link all these processes using the identifiers. We'll cover some of these strategies in the next chapter.

Messaging Patterns

We can identify several patterns in messaging platforms that we can apply depending on what we want to accomplish. Let's detail them from a high-level perspective, without going into implementation details for any specific platform. You can use Figure 7-7 as a guide to understanding these concepts, detailed in the next pages.

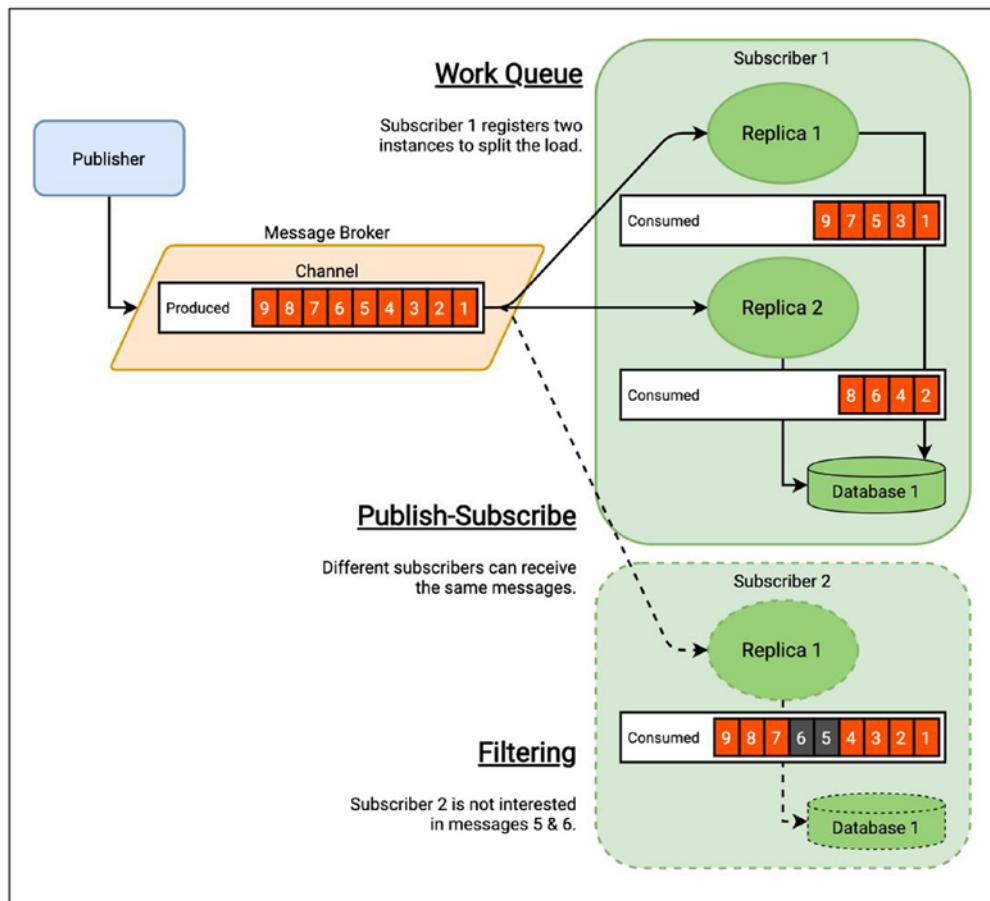


Figure 7-7. Messaging patterns

Publish-Subscribe

In this pattern, different subscribers receive copies of the same messages. For example, we could have multiple components in our system that are interested in the ChallengeSolvedEvent, like the Gamification microservice and a hypothetical Reporting microservice. In this case, it's important to configure these subscribers so they both receive the same messages. Each subscriber would process the event with a different purpose so they don't cause duplicated operations.

Note that this pattern suits better for events and not for messages addressed to a specific service.

Work Queues

This pattern is also known as the *competing consumers* pattern. In this case, we want to split the work between multiple instances of the same application.

As shown in Figure 7-7, we could have multiple replicas of the same microservice. Then, the intention is to balance the load between them. Each instance would consume different messages, process them, and maybe store the result in a database. The database in the figure reminds us that multiple replicas of the same component should share the same data layer, so it's safe to split the work.

Filtering

It's also common that you have subscribers who are interested in all the published messages in a channel, and some others that are just interested in some of them. That's the case of the second subscriber in Figure 7-7. The simplest option we could think of is to discard those as soon as they're consumed, based on some filtering logic included within the application. Instead, some message brokers also offer filtering capabilities out of the box, so a component could register itself as a subscriber with a given filter.

Data Durability

If the broker persists the messages, the subscribers don't need to be running all the time to consume all the data. Each subscriber has an associated *marker* in the broker to know what's the last message that they consumed. If they aren't able to get messages at a given time, the data flow can continue later from where they left it.

Even after all subscribers retrieved a specific message, you may want to keep it stored in the broker for some time. This is useful in case you want new subscribers to get messages that were sent prior to their existence. Also, persisting all messages for a given period can be helpful if you want to “reset the marker” for a subscriber, causing all messages to be reprocessed. This can be used, for example, to repair corrupted data, but it might be also a risky operation when subscribers are not idempotent.

In a system that models all operations as events, you could benefit from event persistence even more. Imagine that you wipe all the data from any existing database. Theoretically, you could replay all events from the beginning and re-create the same state. Therefore, you don’t need to keep the last state of a given entity in the database at all, since you can see it as an “aggregate” of multiple events. This, in a nutshell, it’s the core concept of *event sourcing*. We won’t dive into the details of this technique since it adds an extra layer of complexity, but check <https://tpd.io/eventsrcif> if you want to know more about it.

Message Broker Protocols, Standards, and Tools

Over the years, a few messaging protocols and standards related to message brokers have arisen. This is a reduced list with some popular examples:

- *Advanced Message Queuing Protocol (AMQP)*: This is a wire-level protocol that defines the data format of messages as a stream of bytes.
- *Message Queuing Telemetry Transport (MQTT)*: This is also a protocol, and it has become popular for Internet of Things (IoT) devices since it can be implemented with little code, and it can work under limited bandwidth conditions.
- *Streaming Text Oriented Messaging Protocol (STOMP)*: This is a text-based protocol like HTTP but oriented to messaging middleware.
- *Java Message Service (JMS)*: Unlike the previous ones, JMS is an API standard. It focuses on the behavior that a messaging system should implement. Therefore, we can find different JMS client implementations that connect to message brokers using different underlying protocols.

The following are popular software tools that implement some of these protocols and standards, or have their own ones:

- RabbitMQ is an open source message broker implementation that supports the AMQP, MQTT, and STOMP protocols, among others. It also offers a JMS API client and has a powerful routing configuration.
- Mosquitto is an Eclipse message broker that implements the MQTT protocol, so it's a popular choice for IoT systems.
- Kafka was designed originally by LinkedIn, and it uses its own binary protocol over TCP. Even though the Kafka core features don't offer the same functionalities as a traditional message broker (e.g., routing), it's a powerful messaging platform when the requirements for the messaging middleware are simple. It's commonly used in applications that handle a big volume of data in streams.

As in any case where you need to choose between different tools, you should familiarize with their documentation and analyze how your requirements can benefit from its functionalities: the data volumes you're planning to handle, the delivery guarantees (at-least-once, at-most-once), the error handling strategies, the distributed setup possibilities, etc. Both RabbitMQ and Kafka are popular tools when building event-driven architectures with Java and Spring Boot. Besides, the Spring framework has integrations for these tools, so it's easy to work with them from a coding perspective.

In this book, we use RabbitMQ and the AMQP protocol. The main reason is that this combination offers a wide variety of configuration possibilities, so you can learn most of these options and later reuse this knowledge in any other messaging platform you choose.

AMQP and RabbitMQ

RabbitMQ has native support for the AMQP protocol version 0.9.1 and supports the AMQP 1.0 version via a plugin. We'll use the included 0.9.1 version since it's simpler and has better support; see <https://tpd.io/amqp1>.

We'll take a look now at the main AMQP 0.9.1 concepts. If you want to dive into concepts in more detail, I recommend you refer to <https://tpd.io/amqp-c> in the RabbitMQ documentation.

Overall Description

As described earlier in this chapter, publishers are components or applications in a system that publish messages to the broker. Consumers, also known as *subscribers*, receive and process these messages.

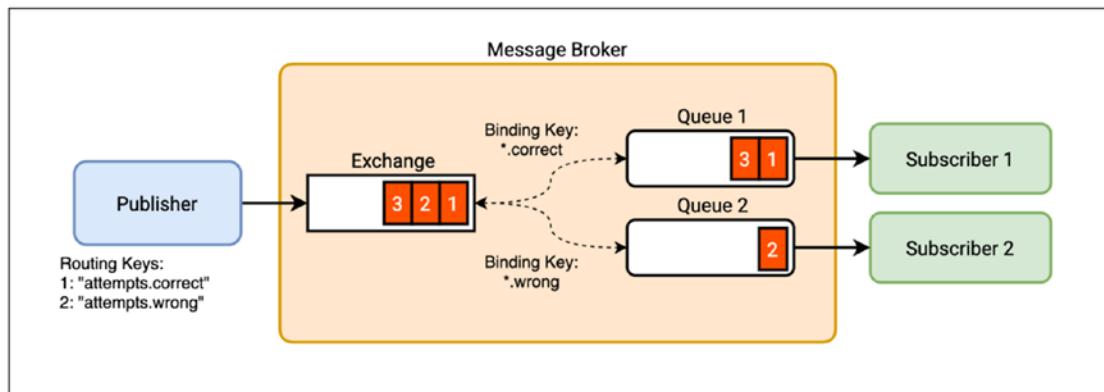


Figure 7-8. RabbitMQ: concepts

AMQP also defines exchanges, queues, and bindings. See Figure 7-8 to better understand these concepts.

- *Exchanges* are the entities where the messages are sent. They do the routing to the queues following a logic defined by the exchange type and rules, known as *bindings*. The exchanges can be durable if they persist after a broker restart, or transient if they don't.
- *Queues* are the objects in AMQP that store the messages to be consumed. Queues may have zero, one, or multiple consumers. A queue can also be durable or transient, but keep in mind that a durable queue doesn't mean that all its messages are persisted. To make messages survive a broker restart, they also have to be published as persistent messages.
- *Bindings* are rules to route messages published to the exchanges to certain queues. Therefore, we say that a queue is bound to a given exchange. Some exchange types support an optional *binding key* to determine which messages published to an exchange should end up in a given queue. In that sense, you can see a binding key as a filter.

On the other hand, publishers can specify *routing keys* when sending messages, so they can be filtered properly based on binding keys if these configurations are in use. Routing keys are composed of words delimited by dots, like `attempt.correct`. Binding keys have a similar format, but they may include pattern matchers, depending on the exchange type.

Exchange Types and Routing

There are several exchange types that we can use. Figure 7-9 shows examples of each of these exchange types, combined with different routing strategies that are defined by the binding keys, and the corresponding routing keys per message.

- The *default exchange* is predeclared by the broker. All created queues are bound to this exchange with a binding key equal to the queue name. From a conceptual perspective, it means that messages can be published with a destination queue in mind if we use that name as a routing key. Technically, these messages still go through the exchange. This setup is not commonly used since it defeats the whole routing purpose.
- The *direct exchange* is commonly used for unicast routing. The difference with the Default Exchange is that you can use your own binding keys, and you can also create multiple queues with the same binding key. Then, these queues will all get messages whose routing key matches the binding key. Conceptually, we use it when we publish messages knowing the destination (unicast), but we don't need to know how many queues will get the message.
- The *fanout exchange* doesn't use routing keys. It routes all the messages to all the queues that are bound to the exchange, so it's perfect for a broadcast scenario.
- The *topic exchange* is the most flexible. Instead of binding queues to this exchange using a given value, we can use a pattern. That allows subscribers to register queues to consume a filtered set of messages. Patterns can use # to match any set of words or * to match only one word.

- The *headers exchange* uses the message headers as routing keys for better flexibility since we can set up the match condition to one or many headers and for an all-match or any-match configuration. Standard routing keys are therefore ignored.

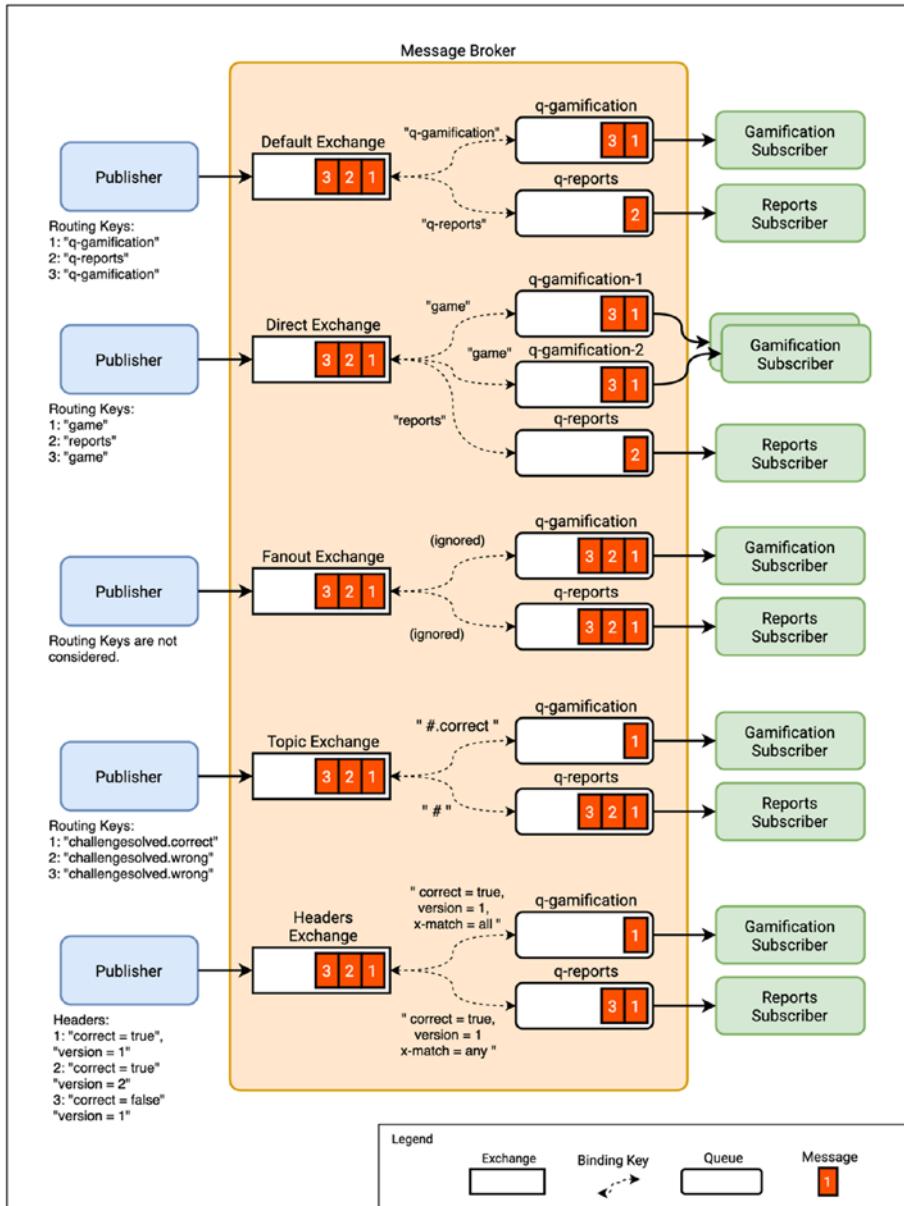


Figure 7-9. Exchange types: examples

As we can see, the publish-subscribe and filtering patterns we described earlier in this chapter are applied in these scenarios. The direct exchange example in the figure might look like a work queue pattern, but it's not. This example is there on purpose to demonstrate that, in AMQP 0.9.1, load balancing happens between consumers of the same queue, not between queues. To implement the work queue pattern, we normally subscribe more than once to the same queue. See Figure 7-10.

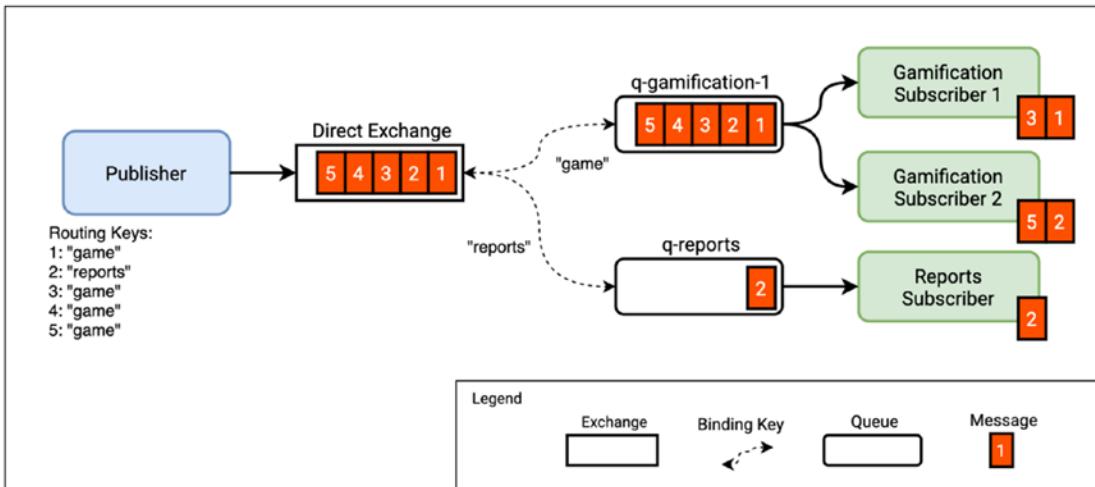


Figure 7-10. Work queue in AMQP

Message Acknowledgments and Rejection

AMQP defines two different acknowledgment modes for consumer applications. Understanding them is important since after a consumer sends an acknowledgment, the message is removed from the queue.

The first alternative is to use *automatic acknowledgment*. With this strategy, messages are considered as delivered when they're sent to the application. The second option is called *explicit acknowledgment*, and it consists of waiting until the application sends an ACK signal. This second option is much better to guarantee that all messages get processed. The consumer can read the message, run some business logic, persist related data, and even trigger a subsequent event before sending the acknowledgment signal to the broker. In this case, the message is removed from the queue only if it has been fully processed. If the consumer dies before sending the signal (or there is an error), the broker will try to deliver the message to another consumer or, if there is none, it'll wait until there is one available.

Consumers can also *reject* messages. For example, imagine that one of the consumer instances can't access the database due to network errors. In this case, the consumer can reject the message, specifying if it should be requeued or discarded. Note that, if the error that caused the message rejection remains for some time and there are no other consumers that can handle it successfully, we may end up in an infinite loop of requeue-rejection.

Setting Up RabbitMQ

Now that we learned the main AMQP concepts, it's time to download and install the RabbitMQ broker.

Go to the RabbitMQ download page (<https://tpd.io/rabbit-dl>) and select the appropriate version for your operating system. In this book, we'll use RabbitMQ version 3.8.3. RabbitMQ is written in Erlang, so you may need to install this framework separately if it's not included in the binary installation for your system.

Once we follow all instructions on the download page, we have to start the broker. The required steps should be also included in the download page for your OS. For example, in Windows, RabbitMQ is installed as a service that you can start/stop from the Start menu. In macOS, you have to run a command from the command line.

RabbitMQ includes some standard plugins, but not all of them are enabled by default. As an extra step, we will enable the management plugin, which gives us access to a Web UI, and an API to monitor and manage the broker. From the `sbin` folder inside the broker's installation folder, we have to execute the following:

```
$ rabbitmq-plugins enable rabbitmq_management
```

Then, when we restart the broker, we should be able to navigate to `http://localhost:15672` and see a login page. Since we're running locally, we can use the default username and password values: `guest/guest`. RabbitMQ supports customization of the access control to the broker; check <https://tpd.io/rmq-ac> if you want to know more details about user authorization. Figure 7-11 shows the RabbitMQ management plugin UI after we log in.

CHAPTER 7 EVENT-DRIVEN ARCHITECTURES

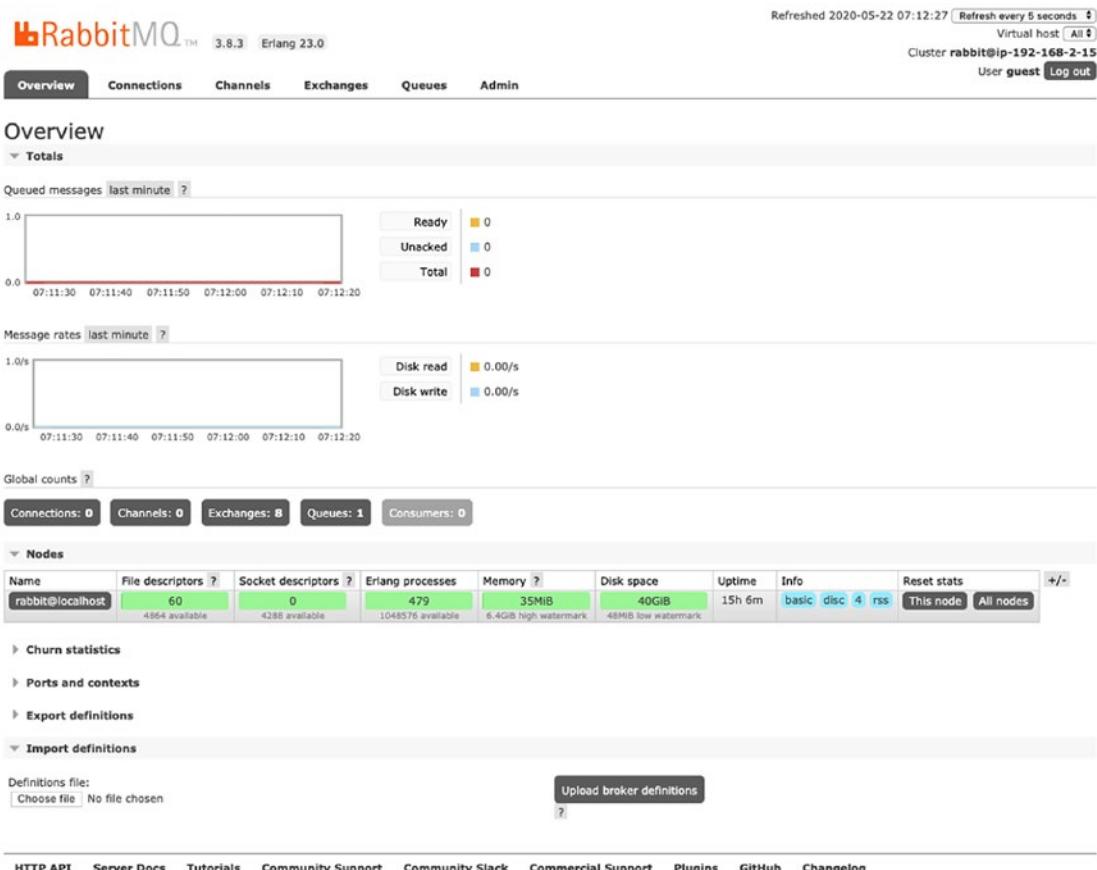


Figure 7-11. RabbitMQ management plugin UI

From this UI, we can monitor queued messages, processing rates, statistics about the different registered nodes, etc. The toolbar gives us access to many other features such as monitoring and management of queues and exchanges. We can even create or delete these entities from this interface. We'll create exchanges and queues programmatically instead, but this tool will be useful to understand how our application works with RabbitMQ.

In the main section, Overview, we can see a list of nodes. We just installed it locally, so there is only one node named rabbit@localhost. We could add more RabbitMQ broker instances over a network and then set up a distributed cluster over different machines. That would give us better availability and fault tolerance since the broker can replicate data, so we can still operate if nodes go down or in case of network partitions. The Clustering Guide (<https://tpd.io/rmq-cluster>) in the official RabbitMQ documentation describes the possible configuration options.

Spring AMQP and Spring Boot

Since we're building our microservices with Spring Boot, we'll use Spring modules to connect to the RabbitMQ message broker. In this case, the Spring AMQP project is what we're looking for. This module contains two artifacts: `spring-rabbit`, which is a set of utils to work with a RabbitMQ broker, and `spring-amqp`, which includes all the AMQP abstractions, so we can make our implementation vendor-independent. Currently, Spring only offers a RabbitMQ implementation of the AMQP protocol.

As with other modules, Spring Boot provides a starter for AMQP with extra utilities such as autoconfiguration: `spring-boot-starter-amqp`. This starter uses both artifacts described earlier, so it implicitly assumes that we'll use a RabbitMQ broker (since it's the only implementation available).

We'll use Spring to declare our exchanges, queues, and bindings and to produce and consume messages.

Solution Design

While describing the concepts in this chapter, we already had a quick preview of what we're going to build. See Figure 7-12. This diagram still includes the sequence numbers to make clear that the response from the Multiplication microservice to the client may happen before the Gamification microservice processes the message. It's an asynchronous, eventually consistent flow.

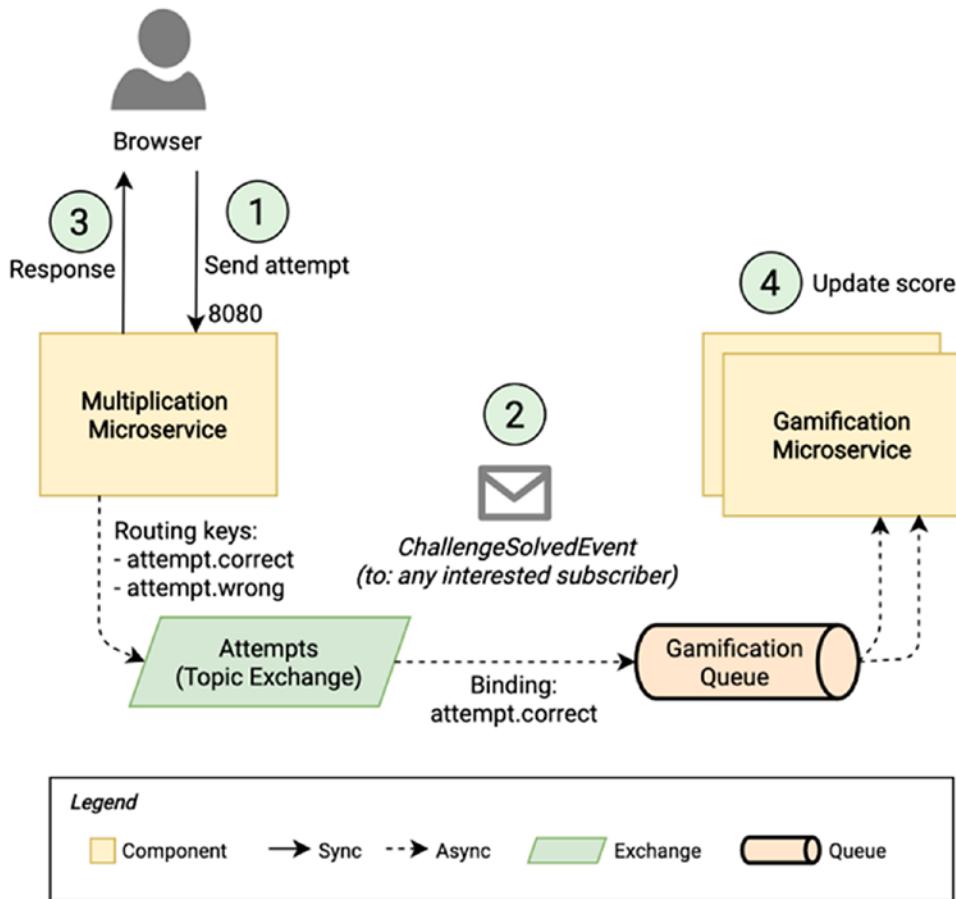


Figure 7-12. Asynchronous process with a message broker

As shown in the diagram, we'll create an attempts exchange, of type Topic. In an event-driven architecture like ours, this gives us the flexibility to send the events with certain routing keys and allow consumers to subscribe to all of them or set up their own filters in their queues.

Conceptually, the Multiplication microservice owns the attempts exchange. It'll use it to publish events that are related to attempts coming from the users. In principle, it'll publish both correct and wrong items, since it doesn't know anything about the consumers' logic. On the other hand, the Gamification microservice declares a queue with a binding key that suits its requirements. In this case, this routing key is used as a filter to receive only correct attempts. As you see in the previous figure, we may have multiple instances of the Gamification microservice consuming from the same queue. In this case, the broker will balance the load between all instances.

In the hypothetical situation of having a different microservice that is also interested in the `ChallengeSolvedEvent`, this one would need to declare its own queue to consume the same messages. For example, we could introduce the `Reports` microservice that creates a “reports” queue and uses a binding key `attempt.*` (or `#`) to consume both correct and wrong attempts.

As you see, we can nicely combine the publish-subscribe and work queue patterns so multiple microservices can process the same messages and multiple instances of the same microservice can share the load between them. Besides, by making publishers responsible for the exchanges and subscribers responsible for the queues, we build an event-driven microservice architecture that achieves loose coupling with the introduction of a message broker.

Let’s create a list of tasks that we need to do to accomplish our plan:

1. Add the new starter dependency to our Spring Boot applications.
2. Remove the REST API client that sends the challenge explicitly to Gamification and the corresponding controller.
3. Rename the `ChallengeSolvedDTO` as `ChallengeSolvedEvent`.
4. Declare the exchange on the Multiplication microservice.
5. Change the logic of the Multiplication microservice to publish an event instead of calling the REST API.
6. Declare the queue on the Gamification microservice.
7. Include the consumer logic to get the events from the queue and connect it to the existing service layer to process the correct attempts for score and badges.
8. Refactor the tests accordingly.

At the end of this chapter, we’ll also play with the new setup and experiment with the load balancing and fault-tolerance benefits that RabbitMQ introduces.

Adding the AMQP Starter

To use the AMQP and RabbitMQ features in our Spring Boot applications, let’s add the corresponding starter to our `pom.xml` files. Listing 7-1 shows this new dependency.

Listing 7-1. Adding the AMQP Starter to Both Spring Boot Projects

```
<dependencies>
    <!-- ... existing dependencies -->
    <dependency>
        <groupId>org.springframework.boot</groupId>
        <artifactId>spring-boot-starter-amqp</artifactId>
    </dependency>
</dependencies>
```

Source Code

You can find all the source code for this chapter on GitHub, in the `chapter07` repository.

See <https://github.com/Book-Microservices-v2/chapter07>.

This starter includes the aforementioned `spring-rabbit` and `spring-amqp` libraries. The transitive dependency `spring-boot-autoconfigure`, which we know from previous chapters, includes some classes that take care of the connection to RabbitMQ and the setup of some convenient defaults.

In this case, one of the most interesting classes is `RabbitAutoConfiguration` (see <https://tpd.io/rabbitautocfg>). It uses a group of properties defined in the `RabbitProperties` class (see <https://tpd.io/rabbitprops>) that we can override in our `application.properties` file. There, we can find for example the predefined port (15672), username (`guest`), and password (`guest`). The autoconfiguration class builds the connection factory and the *configurer* for `RabbitTemplate` objects, which we can use to send (and even receive) messages to RabbitMQ. We'll use the abstraction interface, `AmqpTemplate` (see <https://tpd.io/amqp-temp-doc>).

The autoconfiguration package also includes some default configuration for receiving messages using an alternative mechanism: the `RabbitListener` annotation. We'll cover this in more detail while coding our RabbitMQ subscriber.

Event Publishing from Multiplication

Let's focus first on our publisher, the Multiplication microservice. After we added the new dependency, we can include some extra configuration.

- *The name of the exchange:* It's useful to have it in the configuration in case we need to modify it later depending on the environment we're running our application, or share it across applications as we'll see in the next chapter.
- *Logging settings:* We add them to see extra logs when the app interacts with RabbitMQ. To do this, we'll change the log level of the `RabbitAdmin` class to `DEBUG`. This class interacts with the RabbitMQ broker to declare the exchanges, queues, and bindings.

Besides, we can remove the property that points to the Gamification service; we don't need to call it directly anymore. Listing 7-2 shows all property changes.

Listing 7-2. Adjusting application.properties in the Multiplication Microservice

```
# ... all properties above remain untouched

# For educational purposes we will show the SQL in console
# spring.jpa.show-sql=true <- it's time to remove this

# Gamification service URL <-- We remove this block
# service.gamification.host=http://localhost:8081

amqp.exchange.attempts=attempts.topic

# Shows declaration of exchanges, queues, bindings, etc.
logging.level.org.springframework.amqp.rabbit.core.RabbitAdmin = DEBUG
```

Now we add the Exchange declaration to a separate configuration file for AMQP. The Spring module has a convenient builder for this, `ExchangeBuilder`. What we do is to add a bean of the topic type we want to declare in the broker. Besides, we'll use this configuration class to switch the predefined serialization format to JSON. See Listing 7-3 before we move to the explanation.

Listing 7-3. Adding AMQP Configuration Beans

```

package microservices.book.multiplication.configuration;

import org.springframework.amqp.core.ExchangeBuilder;
import org.springframework.amqp.core.TopicExchange;
import org.springframework.amqp.support.converter.Jackson2JsonMessageConverter;
import org.springframework.beans.factory.annotation.Value;
import org.springframework.context.annotation.Bean;
import org.springframework.context.annotation.Configuration;

<**
 * Configures RabbitMQ via AMQP abstraction to use events in our application.
 */

@Configuration
public class AMQPConfiguration {

    @Bean
    public TopicExchange challengesTopicExchange(
        @Value("${amqp.exchange.attempts}") final String exchangeName) {
        return ExchangeBuilder.topicExchange(exchangeName).durable(true).build();
    }

    @Bean
    public Jackson2JsonMessageConverter producerJackson2MessageConverter() {
        return new Jackson2JsonMessageConverter();
    }

}

```

We make the topic *durable*, so it'll remain in the broker after RabbitMQ restarts. Also, we declare it as a topic exchange since that's the solution we envisioned in our event-driven system. The name is picked up from configuration thanks to the already known `@Value` annotation.

By injecting a bean of type `Jackson2JsonMessageConverter`, we're overriding the default Java object serializer by a JSON object serializer. We do this to avoid various pitfalls of the Java object serialization.

- It's not a proper standard that we can use between programming languages. If we would introduce a consumer that's not written in Java, we have to look for a specific library to perform cross-language deserialization.
- It uses a hard-coded fully qualified type name in the header of the message. The deserializer expects the Java bean to be located in the same package and to have the same name and fields. This is not flexible at all, since we may want to deserialize only some properties and keep our own version of the event data, following good domain-driven design practices.

The `Jackson2JsonMessageConverter` uses a Jackson's `ObjectMapper` preconfigured in Spring AMQP. Our bean will be used then by the `RabbitTemplate` implementation, the class that serializes and send objects as AMQP messages to the broker. On the subscriber side, we can benefit from the popularity of the JSON format to deserialize the contents using any programming language. We could also use our own object representation and ignore properties we don't need on the consumer side, thereby reducing the coupling between microservices. If the publisher includes new fields in the payload, the subscribers don't need to change anything.

JSON is not the only standard supported by Spring AMQP message converters. You can also use XML or Google's Protocol Buffers (a.k.a. *protobuf*). We'll stick to JSON in our system since it's an extended standard, and it's also good for educational purposes because the payload is readable. In real systems where performance is critical, you should consider an efficient binary format (e.g., *protobuf*). See <https://tpd.io/dataser> for a comparison of data serialization formats.

Our next step is to remove the `GamificationServiceClient` class. Then, we also want to rename our existing `ChallengeSolvedDTO` to make it an event. We don't need to modify any field, just the name. See Listing 7-4.

Listing 7-4. Renaming `ChallengeSolvedDTO` as `ChallengeSolvedEvent`

```
package microservices.book.multiplication.challenge;

import lombok.Value;

@Value
public class ChallengeSolvedEvent {
```

```

long attemptId;
boolean correct;
int factorA;
int factorB;
long userId;
String userAlias;

}

```

The naming convention shown here is a good practice for events. They represent a fact that already happened, so the name should use the past tense. Also, by adding the Event suffix, it's really clear we're using an event-driven approach.

Next, we create a new component in our service layer to publish the event. This is the equivalent to the REST client we already removed, but this time we communicate with the message broker. We annotate this new class, ChallengeEventPub, with the @Service stereotype, and use constructor injection to wire an AmqpTemplate object and the name of the exchange. See Listing 7-5 for the complete source code.

Listing 7-5. The ChallengeSolvedEvent's Publisher

```

package microservices.book.multiplication.challenge;

import org.springframework.amqp.core.AmqpTemplate;
import org.springframework.beans.factory.annotation.Value;
import org.springframework.stereotype.Service;

@Service
public class ChallengeEventPub {

    private final AmqpTemplate amqpTemplate;
    private final String challengesTopicExchange;

    public ChallengeEventPub(final AmqpTemplate amqpTemplate,
                           @Value("${amqp.exchange.attempts}")
                           final String challengesTopicExchange) {
        this.amqpTemplate = amqpTemplate;
        this.challengesTopicExchange = challengesTopicExchange;
    }
}

```

```

public void challengeSolved(final ChallengeAttempt challengeAttempt) {
    ChallengeSolvedEvent event = buildEvent(challengeAttempt);
    // Routing Key is 'attempt.correct' or 'attempt.wrong'
    String routingKey = "attempt." + (event.isCorrect() ?
        "correct" : "wrong");
    amqpTemplate.convertAndSend(challengesTopicExchange,
        routingKey,
        event);
}

private ChallengeSolvedEvent buildEvent(final ChallengeAttempt attempt) {
    return new ChallengeSolvedEvent(attempt.getId(),
        attempt.isCorrect(), attempt.getFactorA(),
        attempt.getFactorB(), attempt.getUser().getId(),
        attempt.getUser().getAlias());
}
}

```

AmqpTemplate is just an interface defining the AMQP standards. The underlying implementation is RabbitTemplate, and it uses the JSON converter we configured earlier. We plan to call the challengeSolved method from the main Challenge service logic, within the ChallengeServiceImpl class. This method translates the domain object to the event object using the auxiliary method buildEvent, and it uses the amqpTemplate to convert (to JSON) and send the event with a given routing key. This one is either attempt.correct or attempt.wrong depending on whether the user was right or not.

As we can see, publishing a message to the broker with Spring and Spring Boot is simple thanks to the provided AmqpTemplate/RabbitTemplate and the default configuration, which abstracts the connection to the broker, message conversion, exchange declaration, etc.

The only part we're missing in our code is connecting the challenge logic with this publisher's class. We just need to replace the injected GamificationServiceClient service we use in ChallengeServiceImpl by the new ChallengeEventPub, and use the new method call. We can also rewrite the comment to clarify that we're not calling the Gamification service but sending an event for any component in our system that might be interested. See Listing 7-6.

Listing 7-6. Modifying the ChallengeServiceImpl Class to Send the New Event

```

@Slf4j
@RequiredArgsConstructor
@Service
public class ChallengeServiceImpl implements ChallengeService {

    private final UserRepository userRepository;
    private final ChallengeAttemptRepository attemptRepository;
    private final ChallengeEventPub challengeEventPub; // replaced

    @Override
    public ChallengeAttempt verifyAttempt(ChallengeAttemptDTO attemptDTO) {
        // ...
        // Stores the attempt
        ChallengeAttempt storedAttempt = attemptRepository.save(checkedAttempt);

        // Publishes an event to notify potentially interested subscribers
        challengeEventPub.challengeSolved(storedAttempt);

        return storedAttempt;
    }
    // ...
}

```

Exercise

Modify the existing ChallengeServiceTest to verify that it uses the new service instead of the removed REST client.

Instead of leaving aside the ChallengeEventPubTest as an exercise, let's cover it in the book since it poses a new challenge. We want to check that the AmqpTemplate, which we'll mock, is called with the desired routing key and event object, but we can't access that data from outside the method. Making the method return an object with these values seems like adapting the code too much to our tests. What we can do in this case is to use Mockito's ArgumentCaptor class (see <https://tpd.io/argcap>) to *capture* the arguments passed to a mock, so we can assert these values later.

Besides, since we made a quick break in our journey to visit a test again, we will introduce another JUnit feature: *parameterized tests* (see <https://tpd.io/param-tests>). Our test cases to verify correct and wrong attempts are similar, so we can write a generic test for both cases and use a parameter for the assertion. See the ChallengeEventPubTest source code in Listing 7-7.

Listing 7-7. A Parameterized Test to Check Behavior for Correct and Wrong Attempts

```
package microservices.book.multiplication.challenge;

import org.junit.jupiter.api.BeforeEach;
import org.junit.jupiter.api.extension.ExtendWith;
import org.junit.jupiter.params.ParameterizedTest;
import org.junit.jupiter.params.provider.ValueSource;
import org.mockito.ArgumentCaptor;
import org.mockito.Mock;
import org.mockito.junit.jupiter.MockitoExtension;
import org.springframework.amqp.core.AmqpTemplate;

import microservices.book.multiplication.user.User;

import static org.assertj.core.api.BDDAssertions.*;
import static org.mockito.Mockito.*;

@ExtendWith(MockitoExtension.class)
class ChallengeEventPubTest {

    private ChallengeEventPub challengeEventPub;

    @Mock
    private AmqpTemplate amqpTemplate;

    @BeforeEach
    public void setUp() {
        challengeEventPub = new ChallengeEventPub(amqpTemplate,
                "test.topic");
    }

    @ParameterizedTest
    @ValueSource(bools = {true, false})
    public void sendsAttempt(boolean correct) {
```

```

// given
ChallengeAttempt attempt = createTestAttempt(correct);

// when
challengeEventPub.challengeSolved(attempt);

// then
var exchangeCaptor = ArgumentCaptor.forClass(String.class);
var routingKeyCaptor = ArgumentCaptor.forClass(String.class);
var eventCaptor = ArgumentCaptor.forClass(ChallengeSolvedEvent.class);

verify(amqpTemplate).convertAndSend(exchangeCaptor.capture(),
                                    routingKeyCaptor.capture(), eventCaptor.capture());
then(exchangeCaptor.getValue()).isEqualTo("test.topic");
then(routingKeyCaptor.getValue()).isEqualTo("attempt." +
                                         (correct ? "correct" : "wrong"));
then(eventCaptor.getValue()).isEqualTo(solvedEvent(correct));
}

private ChallengeAttempt createTestAttempt(boolean correct) {
    return new ChallengeAttempt(1L, new User(10L, "john"), 30, 40,
                                correct ? 1200 : 1300, correct);
}

private ChallengeSolvedEvent solvedEvent(boolean correct) {
    return new ChallengeSolvedEvent(1L, correct, 30, 40, 10L, "john");
}

}

```

Gamification as a Subscriber

Now that we finished the publisher's code, we move to the subscriber's: the Gamification microservice. In a nutshell, we need to replace the existing controller that accepts attempts by an event subscriber. That implies creating an AMQP queue and binding it to the topic exchange that we declared earlier in the Multiplication microservice.

First, let's fill in the configuration settings. We remove also here the property to show the queries and add extra logging for RabbitMQ. Then, we set up the names of the new queue and the exchange, which matches the value we added to the previous service. See Listing 7-8.

Listing 7-8. Defining Queue and Exchange Names in Gamification

```
# ... all properties above remain untouched
amqp.exchange.attempts=attempts.topic
amqp.queue.gamification=gamification.queue

# Shows declaration of exchanges, queues, bindings, etc.
logging.level.org.springframework.amqp.rabbit.core.RabbitAdmin = DEBUG
```

To declare the new queue and the binding, we'll also use a configuration class named `AMQPConfiguration`. Bear in mind that we should also declare the exchange on the consumer's side. Even though the subscriber doesn't own the exchange conceptually, we want our microservices to be able to start in any given order. If we don't declare the exchange on the Gamification microservice and the broker's entities have not been initialized yet, we're forced to start the Multiplication microservice before. The exchange has to be there when we declare the queue. This applies only the first time since we make the exchange durable, yet it's a good practice to declare all exchanges and queues that a microservice requires within its code, so it doesn't rely on any other. Note that the declaration of RabbitMQ entities is an idempotent operation; if the entity is there, the operation doesn't have any effect.

We also need some configuration on the consumer side to deserialize the messages using JSON, instead of the format provided by the default's message converter. Let's have a look at the full source code of the configuration class in Listing 7-9, and we'll detail some parts later.

Listing 7-9. The AMQP Configuration for the Gamification Microservice

```
package microservices.book.gamification.configuration;

import com.fasterxml.jackson.annotation.JsonCreator;
import com.fasterxml.jackson.module.paramnames.ParameterNamesModule;

import org.springframework.amqp.core.*;
import org.springframework.amqp.rabbit.annotation.RabbitListenerConfigurer;
import org.springframework.beans.factory.annotation.Value;
import org.springframework.context.annotation.Bean;
import org.springframework.context.annotation.Configuration;
import org.springframework.messaging.converter.MappingJackson2MessageConverter;
```

CHAPTER 7 EVENT-DRIVEN ARCHITECTURES

```
import org.springframework.messaging.handler.annotation.support.  
DefaultMessageHandlerMethodFactory;  
import org.springframework.messaging.handler.annotation.support.  
MessageHandlerMethodFactory;  
  
@Configuration  
public class AMQPConfiguration {  
  
    @Bean  
    public TopicExchange challengesTopicExchange(  
        @Value("${amqp.exchange.attempts}") final String exchangeName) {  
        return ExchangeBuilder.topicExchange(exchangeName).durable(true).build();  
    }  
  
    @Bean  
    public Queue gamificationQueue(  
        @Value("${amqp.queue.gamification}") final String queueName) {  
        return QueueBuilder.durable(queueName).build();  
    }  
  
    @Bean  
    public Binding correctAttemptsBinding(final Queue gamificationQueue,  
                                         final TopicExchange attemptsExchange) {  
        return BindingBuilder.bind(gamificationQueue)  
            .to(attemptsExchange)  
            .with("attempt.correct");  
    }  
  
    @Bean  
    public MessageHandlerMethodFactory messageHandlerMethodFactory() {  
        DefaultMessageHandlerMethodFactory factory = new  
        DefaultMessageHandlerMethodFactory();  
  
        final MappingJackson2MessageConverter jsonConverter =  
            new MappingJackson2MessageConverter();  
        jsonConverter.getObjectMapper().registerModule(  
            new ParameterNamesModule(JsonCreator.Mode.PROPERTIES));  
  
        factory.setMessageConverter(jsonConverter);  
        return factory;  
    }  
}
```

```

@Bean
public RabbitListenerConfigurer rabbitListenerConfigurer(
    final MessageHandlerMethodFactory messageHandlerMethodFactory) {
    return (c) -> c.setMessageHandlerMethodFactory(messageHandlerMethodFactory);
}

}

```

The declaration of the exchange, queue, and binding are straightforward with the provided builders. We declare a durable queue to make it survive broker restarts, with a name coming from the configuration value. The Bean's declaration method for the Binding uses the two other beans, injected by Spring, and links them with the value attempt.correct. As mentioned already, we're interested only in the correct attempts to process scores and badges.

Next to that, we set up a `MessageHandlerMethodFactory` bean to replace the default one. We actually use the default factory as a baseline but then replace its message converter by a `MappingJackson2MessageConverter` instance, which handles the message deserialization from JSON to Java classes. We fine-tune its included `ObjectMapper` and add the `ParameterNamesModule` to avoid having to use empty constructors for our event classes. Note that we didn't need to do this when passing DTOs via REST APIs (our previous implementation) because Spring Boot configures this module within the web layer autoconfiguration. However, it doesn't do this for RabbitMQ because JSON is not the default option; therefore, we need to configure it explicitly.

This time, we won't use the `AmqpTemplate` to receive messages since that's based on polling, which consumes network resources unnecessarily. Instead, we want the broker to notify subscribers when there are messages, so we'll go for an asynchronous option. The AMQP abstraction doesn't support this, but the `spring-rabbit` component offers two mechanisms for consuming messages asynchronously. The simplest, most popular one is the `@RabbitListener` annotation, which we'll use to get the events from the queue. To configure the listeners to use a JSON deserialization, we have to override the bean `RabbitListenerConfigurer` with an implementation that uses our custom `MessageHandlerMethodFactory`.

Our next task is to rename the `ChallengeSolvedDTO` to `ChallengeSolvedEvent`. See Listing 7-10. Technically, there is no need to use the same class name since the JSON format only specifies the field names and values. However, this is a good practice because then you can easily find related event classes across your projects.

Listing 7-10. Renaming ChallengeSolvedDTO as ChallengeSolvedEvent in Gamification

```
package microservices.book.gamification.challenge;

import lombok.Value;

@Value
public class ChallengeSolvedEvent {

    long attemptId;
    boolean correct;
    int factorA;
    int factorB;
    long userId;
    String userAlias;

}
```

Following domain-driven design practices, we could adjust this event's deserialized fields. For instance, we don't need the `userAlias` for the Gamification's business logic, so we could remove it from the consumed event. Since Spring Boot configures the `ObjectMapper` to ignore unknown properties by default, that strategy would work without needing to configure anything else. Not sharing the code of this class across microservices is a good practice because it also allows for loose coupling, backward compatibility, and independent deployments. Imagine that the Multiplication microservice would evolve and store extra data, for example, a third factor for harder challenges. This extra factor would then be added to the published event's code. The good news is that, by using different representations of the event per domain and configuring the mapper to ignore unknown properties, the Gamification microservice would still work after such change without needing to update its event representation.

Let's code now the event consumer. As introduced earlier, we'll use the `@RabbitListener` annotation for this. We can add this annotation to a method to make it act as the processing logic of a message when it arrives. In our case, we only need to specify the queue name to subscribe to, since we already declared all RabbitMQ entities in a separate configuration file. There are options to embed these declarations within this annotation, but the code doesn't look that clean anymore (see <https://tpd.io/rmq-listener> if you're curious:).

Check the source of the consumer in Listing 7-11, and then we'll cover the most relevant parts.

Listing 7-11. The RabbitMQ Consumer's Logic

```
package microservices.book.gamification.game;

import org.springframework.amqp.AmqpRejectAndDontRequeueException;
import org.springframework.amqp.rabbit.annotation.RabbitListener;
import org.springframework.stereotype.Service;

import lombok.RequiredArgsConstructor;
import lombok.extern.slf4j.Slf4j;
import microservices.book.gamification.challenge.ChallengeSolvedEvent;

@RequiredArgsConstructor
@Slf4j
@Service
public class GameEventHandler {

    private final GameService gameService;

    @RabbitListener(queues = "${amqp.queue.gamification}")
    void handleMultiplicationSolved(final ChallengeSolvedEvent event) {
        log.info("Challenge Solved Event received: {}", event.getAttemptId());
        try {
            gameService.newAttemptForUser(event);
        } catch (final Exception e) {
            log.error("Error when trying to process ChallengeSolvedEvent", e);
            // Avoids the event to be re-queued and reprocessed.
            throw new AmqpRejectAndDontRequeueException(e);
        }
    }
}
```

As you see, the amount of code needed to implement a RabbitMQ subscriber is minimal. We can pass the queue name using the configuration property to the `RabbitListener` annotation. Spring processes this method and analyzes the arguments. Given that we specified a `ChallengeSolvedEvent` class as the expected input, Spring

automatically configures a deserializer to transform the message from the broker into this object type. It'll use JSON since we override the default `RabbitListenerConfigurer` in our `AMQPConfiguration` class.

From the consumer's code, you can also infer what our error handling strategy is. By default, the logic that Spring builds based on the `RabbitListener` annotations will send the acknowledgment to the broker when the method finalizes without exceptions. In Spring Rabbit, this is called the `AUTO` acknowledgment mode. We could change it to `NONE` if we want the ACK signal to be sent even before processing it, or to `MANUAL` if we want to be fully in control (then we have to inject an extra parameter to send this signal). We can set up this parameter and other configuration values at the factory level (global configuration) or at the listener level (via passing extra parameters to the `RabbitListener` annotation). Our error strategy here is to use the default value `AUTO` but catch any possible exception, log the error, and then rethrow an `AmqpRejectAndDontRequeueException`. This is a shortcut provided by Spring AMQP to reject the message and tell the broker not to requeue it. That means that if there is an unexpected error in the Gamification's consumer logic, we'll lose the message. That is acceptable in our case. If we would want to avoid this situation, we could also set up our code to retry a few times by rethrowing an exception with the opposite meaning, `ImmediateRequeueAmqpException`, or use some tools available in Spring AMQP like an error handler or *message recoverer* to process these failed messages. See the Exception Handling section (<https://tpd.io/spring-amqp-exc>) in the Spring AMQP docs for more detailed information.

We can do a lot with the `RabbitListener` annotation. These are a few of the included functionalities:

- Declare exchanges, queues, and bindings.
- Receive messages from multiple queues with the same method.
- Process the message headers by annotating extra arguments with `@Header` (for a single value) or `@Headers` (for a map).
- Inject a `Channel` argument, so we can control acknowledgments, for example.
- Implement a Request-Response pattern, by returning a value from the listener.
- Move the annotation to the class level and use `@RabbitHandler` for methods. This approach allows us to configure multiple methods to process different message types that are coming through the same queue.

For details about these use cases, check the Spring AMQP documentation (<https://tpd.io/samqp-docs>).

Exercise

Create a test for the new `GameEventHandler` class. Check that the service is called and that an exception in its logic causes the re-throwing of the expected AMQP exception. The solution is included in the provided source code for this chapter.

Now that we have the subscriber's logic, we can safely remove the `GameController` class. Then, we refactor the existing `GameService` interface and its implementation, `GameServiceImpl`, to accept the renamed `ChallengeSolvedEvent`. The rest of the logic can remain the same. See Listing 7-12 with the resulting `newAttemptForUser` method.

Listing 7-12. The Updated Newattemptforuser Method Using the Event Class

```

@Override
public GameResult newAttemptForUser(final ChallengeSolvedEvent challenge) {
    // We give points only if it's correct
    if (challenge.isCorrect()) {
        ScoreCard scoreCard = new ScoreCard(challenge.getUserId(),
            challenge.getAttemptId());
        scoreRepository.save(scoreCard);
        log.info("User {} scored {} points for attempt id {}",
            challenge.getUserAlias(), scoreCard.getScore(),
            challenge.getAttemptId());
        List<BadgeCard> badgeCards = processForBadges(challenge);
        return new GameResult(scoreCard.getScore(),
            badgeCards.stream().map(BadgeCard::getBadgeType)
                .collect(Collectors.toList()));
    } else {
        log.info("Attempt id {} is not correct. " +
            "User {} does not get score.",
            challenge.getAttemptId(),
            challenge.getUserAlias());
        return new GameResult(0, List.of());
    }
}

```

We could remove the check for the correct attempt, but then we would depend too much on a proper routing on the Multiplication microservice. If we keep it, it's easier for everyone to read the code and know what it does without having to figure out that there is a filter logic based on routing keys. We can benefit from the broker's routing, but remember that we don't want to embed too much behavior inside the channel.

With these changes, we finalized the required modifications to switch to an event-driven architecture in our microservices. Keep in mind that there are more classes affected by the renamed DTO as `ChallengeSolvedEvent`. We omitted them since your IDE should take care of these changes automatically. Once more, let's review the list of changes we made to our system:

1. We added the new AMQP starter dependency to our Spring Boot applications to use AMQP and RabbitMQ.
2. We removed the REST API client (in Multiplication) and the controller (in Gamification) because we switched to an event-driven architecture using RabbitMQ.
3. We renamed the `ChallengeSolvedDTO` as `ChallengeSolvedEvent`. The renaming caused the modification of other classes and tests, but those changes are not relevant.
4. We declared the new topic exchange in both microservices.
5. We changed the logic of the Multiplication microservice to publish an event instead of calling the REST API.
6. We defined the new queue on the Gamification microservice.
7. We implemented the RabbitMQ consumer logic in the Gamification microservice.
8. We refactored the tests accordingly to adapt them to the new interface.

Remember that you have all the code shown in this chapter available in the book's online repository.

Analysis of Scenarios

Let's try a few different scenarios with our new event-driven system. Our goal is to prove that the introduction of the new architecture design with the message broker brings real advantages.

To recap, see Figure 7-13 for the current state of our system.

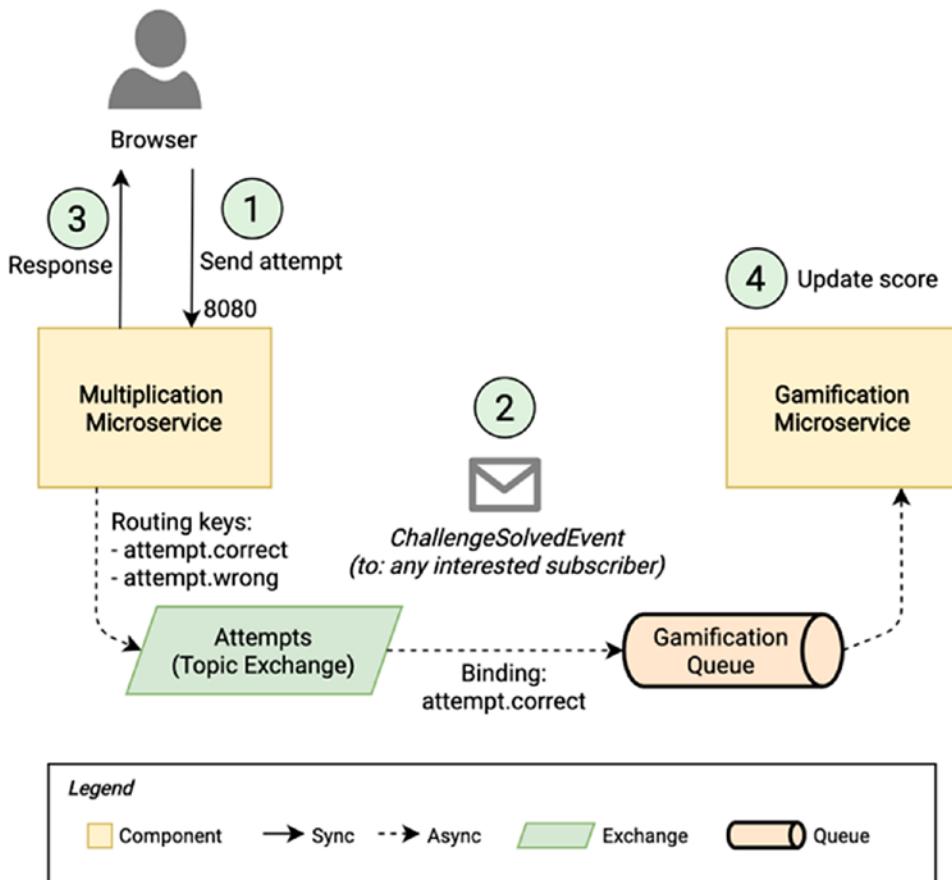


Figure 7-13. Logical view

All the scenarios in this section require us to start up the complete system following these steps:

1. Make sure the RabbitMQ service is running. Otherwise, start it.
2. Run both microservice applications: Multiplication and Gamification.
3. Run React's user interface.
4. From a browser, go to the RabbitMQ admin UI at `http://localhost:15672/` and log in using guest/guest.

Happy Flow

We didn't see our system working with the new message broker yet. That's the first thing we're going to try. Before that, let's check the logs of the Gamification microservice. You should see some new log lines, as shown in Listing 7-13.

Listing 7-13. Spring Boot Application Logs Showing the Initialization for Rabbitmq

```
INFO 11686 --- [main] o.s.a.r.c.CachingConnectionFactory: Attempting to connect to: [localhost:5672]
INFO 11686 --- [main] o.s.a.r.c.CachingConnectionFactory: Created new connection: rabbitConnectionFactory#7c7e73c5:0/SimpleConnection@2bf2d6eb [delegate=amqp://guest@127.0.0.1:5672/, localPort= 63651]
DEBUG 11686 --- [main] o.s.amqp.rabbit.core.RabbitAdmin : Initializing declarations
DEBUG 11686 --- [main] o.s.amqp.rabbit.core.RabbitAdmin : declaring Exchange 'attempts.topic'
DEBUG 11686 --- [main] o.s.amqp.rabbit.core.RabbitAdmin : declaring Queue 'gamification.queue'
DEBUG 11686 --- [main] o.s.amqp.rabbit.core.RabbitAdmin : Binding destination [gamification.queue (QUEUE)] to exchange [attempts.topic] with routing key [attempt.correct]
DEBUG 11686 --- [main] o.s.amqp.rabbit.core.RabbitAdmin : Declarations finished
```

The first two lines are usually logged when we use Spring AMQP. They indicate that the connection to the broker was successful. As mentioned earlier, we didn't need to add any connection properties such as the host or the credentials since we're using the defaults. The rest of the log lines are there because we changed the logging level of the RabbitAdmin class to DEBUG. These are self-explanatory, including the values of the exchange, queue, and binding we created.

On the Multiplication side, there are no RabbitMQ logs yet. The reason is that the connection and the declaration of the exchange happen only when we publish our first message. This means that the topic exchange is declared first by the Gamification microservice. It's good we prepared our code not to mind the booting sequence.

We can now have a look at the RabbitMQ UI to see the current status. On the Connections tab, we'll see one created by the Gamification microservice. See Figure 7-14.

The screenshot shows the RabbitMQ management interface. At the top, there are navigation tabs: Overview, **Connections**, Channels, Exchanges, Queues, and Admin. The Connections tab is active. In the top right corner, there are refresh controls (Refresh every 5 seconds), a dropdown for Virtual host (set to All), a Cluster status (rabbit@ip-192-168-2-15), and a User (guest) with a Log out button. Below the tabs, the title 'Connections' is displayed, followed by a section header 'All connections (1)'. There is a 'Pagination' section with a page number (1), a filter input field, a 'Regex?' checkbox, and a 'Displaying 1 item, page size up to: 100' message. The main content area shows a table with the following data:

Overview		Details				Network		+/-
Name	User name	State	SSL / TLS	Protocol	Channels	From client	To client	
127.0.0.1:63651 rabbitConnectionFactory@7c7e73c5:0	guest	running	o	AMQP 0-9-1	1	0kB/s	0kB/s	

Below the table, there is a footer with links: HTTP API, Server Docs, Tutorials, Community Support, Community Slack, Commercial Support, Plugins, GitHub, and Changelog.

Figure 7-14. RabbitMQ UI: single connection

If we switch to the Exchanges tab, we'll see the `attempts.topic` exchange, of type `topic`, and declared as *durable* (D). See Figure 7-15.

The screenshot shows the RabbitMQ management interface. At the top, there's a header with the RabbitMQ logo, version 3.8.3, Erlang 23.0, and connection information (Cluster rabbit@ip-192-168-2-15, User guest, Log out). Below the header, a navigation bar has tabs for Overview, Connections, Channels, Exchanges (which is selected and highlighted in grey), Queues, and Admin. The main content area is titled "Exchanges" and shows a table of exchanges. The table has columns: Name, Type, Features, Message rate in, and Message rate out. There are 8 rows:

Name	Type	Features	Message rate in	Message rate out
(AMQP default)	direct	D		
amq.direct	direct	D		
amq.fanout	fanout	D		
amq.headers	headers	D		
amq.match	headers	D		
amq.rabbitmq.trace	topic	D I		
attempts.topic	topic	D		

Below the table, there's a link to "Add a new exchange". At the bottom of the page, there's a footer with links: HTTP API, Server Docs, Tutorials, Community Support, Community Slack, Commercial Support, Plugins, GitHub, and Changelog.

Figure 7-15. RabbitMQ UI: exchange list

Now, clicking the exchange name takes us to the detail page, where we can even see a basic graph displaying what are the bound queues and the corresponding binding key. See Figure 7-16.

Refreshed 2020-05-24 10:11:47 Refresh every 5 seconds Virtual host All Cluster rabbit@ip-192-168-2-15 User guest Log out

Overview Connections Channels Exchanges Queues Admin

Exchange: attempts.topic

Message rates last hour ?

Currently idle

Details

Type	topic
Features	durable: true
Policy	

Bindings

To	Routing key	Arguments	Unbind
gamification.queue	attempt.correct		Unbind

Figure 7-16. RabbitMQ UI: exchange detail

The Queues tab shows the recently created Queue with its name, also configured as durable. See Figure 7-17.

Refreshed 2020-05-24 10:05:27 Refresh every 5 seconds Virtual host All Cluster rabbit@ip-192-168-2-15 User guest Log out

Overview Connections Channels Exchanges Queues Admin

Queues

All queues (1)

Pagination

Page 1 of 1 - Filter: Regex ? Displaying 1 item , page size up to: 100

Overview		Messages				Message rates			+/-
Name	Type	Features	State	Ready	Unacked	Total	Incoming	deliver / get	ack
gamification.queue	classic	D	Idle	0	0	0			

Add a new queue

HTTP API Server Docs Tutorials Community Support Community Slack Commercial Support Plugins GitHub Changelog

Figure 7-17. RabbitMQ UI: queue list

After we had a look at how everything has been initialized, let's navigate to our UI and send some correct and incorrect attempts. If you prefer, you can cheat a bit and run this command at least ten times, which will generate ten correct attempts.

```
$ http POST :8080/attempts factorA=15 factorB=20 userAlias=test1 guess=300
```

In the Multiplication logs, we should see now how it connects to the broker and declares the exchange (which has no effect since it was there already).

The logs of the Gamification app should then reflect the consumption of the events and the corresponding updated score. See Listing 7-14.

Listing 7-14. Logs of the Gamification Microservice After Receiving New Events

```
INFO 11686 --- [ntContainer#0-1] m.b.gamification.game.GameEventHandler :  
Challenge Solved Event received: 50  
INFO 11686 --- [ntContainer#0-1] m.b.gamification.game.GameServiceImpl : User  
test1 scored 10 points for attempt id 50  
INFO 11686 --- [ntContainer#0-1] m.b.gamification.game.GameEventHandler :  
Challenge Solved Event received: 51  
INFO 11686 --- [ntContainer#0-1] m.b.gamification.game.GameServiceImpl : User  
test1 scored 10 points for attempt id 51  
INFO 11686 --- [ntContainer#0-1] m.b.gamification.game.GameEventHandler :  
Challenge Solved Event received: 52  
INFO 11686 --- [ntContainer#0-1] m.b.gamification.game.GameServiceImpl : User  
test1 scored 10 points for attempt id 52  
INFO 11686 --- [ntContainer#0-1] m.b.gamification.game.GameEventHandler :  
Challenge Solved Event received: 53  
...  
...
```

The Connections tab in the RabbitMQ manager displays at this time the connections from both applications. See Figure 7-18.

The screenshot shows the RabbitMQ UI's 'Connections' tab. At the top, there are tabs for Overview, Connections (which is selected), Channels, Exchanges, Queues, and Admin. On the right, it shows 'User guest' and 'Log out'. Below the tabs, the title 'Connections' is displayed with a dropdown menu 'All connections (2)'. A 'Pagination' section includes a search bar 'Page 1 of 1 - Filter:' and a checkbox 'Regex ?'. To the right, it says 'Displaying 2 items, page size up to: 100'. The main table has columns: Name, User name, State, SSL / TLS, Protocol, Channels, From client, and To client. Two connections are listed:

Name	User name	State	SSL / TLS	Protocol	Channels	From client	To client
127.0.0.1:54536 rabbitConnectionFactory#796f632b:0	guest	running	○	AMQP 0-9-1	1	2kB/s	2kB/s
127.0.0.1:63651 rabbitConnectionFactory#7c7e73c5:0	guest	running	○	AMQP 0-9-1	1	0kB/s	0kB/s

At the bottom, there are links for HTTP API, Server Docs, Tutorials, Community Support, Community Slack, Commercial Support, Plugins, GitHub, and Changelog.

Figure 7-18. RabbitMQ UI: both connections

Besides, we can see some activity happening in the broker if we go to the Queues tab and click the queue name. You can change the filter to the last 10 minutes on the Overview panel to make sure you capture all the events. See Figure 7-19.

The screenshot shows the RabbitMQ UI's 'Queues' tab. At the top, there are tabs for Overview, Connections, Channels, Exchanges, Queues (selected), and Admin. On the right, it shows 'User guest' and 'Log out'. Below the tabs, the title 'Queue gamification.queue' is displayed with a dropdown menu 'Overview'. A 'Queued messages last ten minutes' chart shows 1.0 message at 10:22. To its right, a legend indicates Ready (yellow square), Unacked (light blue square), and Total (dark red square). Both Ready and Unacked counts are 0. A 'Message rates last ten minutes' chart shows rates from 0.0/s to 1.0/s. To its right, a legend indicates Publish (yellow square), Deliver (manual ack) (light blue square), Deliver (auto ack) (dark red square), Consumer ack (green square), Redelivered (purple square), and Get (manual ack) (dark blue square). All rates are 0.00/s. Below these charts, specific metrics are shown: Get (auto ack) at 0.00/s and Get (empty) at 0.00/s.

Figure 7-19. RabbitMQ UI: queue detail

This is great. Our system works perfectly with the message broker. The correct attempts are routed to the queue declared by the Gamification application. This microservice also subscribes to that queue, so it gets the events published to the exchange and processes them to assign new scores and badges. After that, as already happened before the new changes, our UI will get the updated statistics in the next request to the Gamification's REST endpoint to retrieve the leaderboard. See Figure 7-20.

The screenshot shows a web-based user interface for a challenge. At the top, it says "Your new challenge is" followed by a large "80 x 64". Below this, there are two input fields: "Your alias:" and "Your guess:", both containing the value "0". A "Submit" button is located below the guess field. At the bottom, there is a section titled "Leaderboard" displaying a table of user scores and badges:

User	Score	Badges
test1	100	Bronze First time
jane	20	First time
john	10	First time

Figure 7-20. UI: app working with the message broker

Gamification Becomes Unavailable

The previous implementation of our system, as we left it after the previous chapter, was resilient in the sense that it didn't fail if the Gamification microservice was unavailable. However, in that situation, we would miss all the attempts sent during the incident. Let's see what happens now with the introduction of the message broker.

First, make sure you stop the Gamification microservice. Then, we can send again another ten attempts using either the UI or the command-line trick. Let's use the alias test-g-down:

```
$ http POST :8080/attempts factorA=15 factorB=20 userAlias=test-g-down guess=300
```

The Queue detail view in the RabbitMQ UI shows now ten queued messages. This number doesn't go back to zero as before. This is because the queue is still there, but there are no consumers to dispatch these messages to. See Figure 7-21.

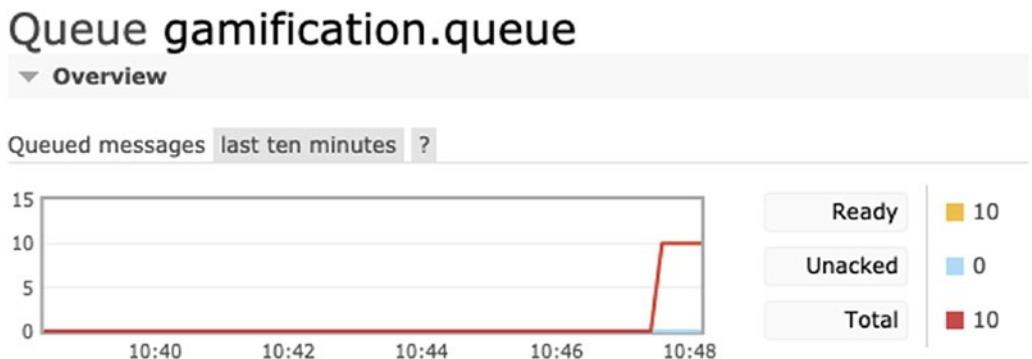


Figure 7-21. RabbitMQ UI: queued messages

We can also check the logs of the Multiplication microservice and verify that there are no errors. It published the messages to the broker and returned an OK response to the API client. We achieved loose coupling. The Multiplication app doesn't need to know if consumers are available. The whole process is asynchronous and event-driven now.

When we bring the Gamification service back again, we'll see in the logs how immediately after booting up it receives all the event messages from the broker. Then, this service just triggers its logic, and the score is updated accordingly. We didn't miss any data this time. Listing 7-15 shows an extract of the Gamification logs after you start it again.

Listing 7-15. The Application Consumes the Pending Events After Becoming Available Again

```

INFO 24808 --- [           main] m.b.g.GamificationApplication      :
Started GamificationApplication in 3.446 seconds (JVM running for 3.989)
INFO 24808 --- [ntContainer#0-1] m.b.gamification.game.GameServiceImpl : User
test-g-down scored 10 points for attempt id 61
INFO 24808 --- [ntContainer#0-1] m.b.gamification.game.GameEventHandler   :
Challenge Solved Event received: 62
INFO 24808 --- [ntContainer#0-1] m.b.gamification.game.GameServiceImpl : User
test-g-down scored 10 points for attempt id 62

```

```

INFO 24808 --- [ntContainer#0-1] m.b.gamification.game.GameEventHandler : 
Challenge Solved Event received: 63
INFO 24808 --- [ntContainer#0-1] m.b.gamification.game.GameServiceImpl : User
test-g-down scored 10 points for attempt id 63
INFO 24808 --- [ntContainer#0-1] m.b.gamification.game.GameEventHandler : 
Challenge Solved Event received: 64
...

```

You can also verify how the leaderboard shows up again with the updated score for the user test-g-down. We made our system not only resilient but also able to recover after a failure. The queue detail in the RabbitMQ interface also shows a zero counter for queued messages, since they all have been consumed already.

As you can imagine, RabbitMQ allows us to configure for how long the messages can be kept in a queue before discarding them (time-to-live, TTL). We can also configure a maximum length for the queue if we prefer so. By default, these parameters are not set, but we can enable them per message (at publishing time) or when we declare the queue. See Listing 7-16 for an example of how we could have configured our queue to have a custom TTL of six hours and a max length of 25000 messages. This is just an example of how important it is that you get familiar with the configuration of the broker, so you can adjust it to your needs.

Listing 7-16. An Example Queue Configuration Showing Some Extra Parameter Options

```

@Bean
public Queue gamificationQueue(
    @Value("${amqp.queue.gamification}") final String queueName) {
    return QueueBuilder.durable(queueName)
        .ttl((int) Duration.ofHours(6).toMillis())
        .maxLength(25000)
        .build();
}

```

The Message Broker Becomes Unavailable

Let's go one step further and bring the broker down while the queue has messages pending to be delivered. To test this scenario, we should follow these steps:

1. Stop the Gamification microservice.
2. Send a few correct attempts with the user alias `test-rmq-down` and verify in the RabbitMQ UI that the queue is keeping those messages.
3. Stop the RabbitMQ broker.
4. Send one extra correct attempt.
5. Start the Gamification microservice.
6. After about ten seconds, start again the RabbitMQ broker.

The outcome of this manual test is that only the attempt we sent while the broker is down fails to be processed. Actually, we'll get an HTTP error response from the server since we didn't catch any potential exception within the publisher, nor in the main service logic located at `ChallengeServiceImpl`. We could add a try/catch clause, so we are still able to respond. The strategy would be then to suppress the error silently. A possibly better approach is to implement a custom HTTP error handler to return a specific error response such as `503 SERVICE UNAVAILABLE` to indicate that the system is not operational when we lose connection with the broker. As you see, we have multiple options. In a real organization, the best approach is to discuss these alternatives and choose the one that better suits your nonfunctional requirements like *availability* (we want to have the challenge features available as much time as possible) or *data integrity* (we want to have always a score for every sent attempt).

A second observation from our test is that none of the two microservices crash when the broker goes unavailable. Instead, the Gamification microservice keeps retrying to connect every few seconds, and the Multiplication microservice does the same when a new attempt request comes. When we bring the broker up again, both microservices recover their connections. This is a nice feature included in the Spring AMQP project, to try to recover the connection when it's unavailable.

If you perform these steps, you also see how the consumer gets the messages even after the broker restarted while there were pending ones to be sent. The Gamification microservice reconnects to RabbitMQ, and this one sends the queued events. This is

not only because we declared a durable exchange and queue, but because the Spring implementation uses the *persistent delivery mode* while publishing all messages. This is one of the message properties that we could also set ourselves if we would use RabbitTemplate (instead of AmqpTemplate) to publish messages. See Listing 7-17 for an example of how we could change the delivery mode to make our messages not survive a broker restart.

Listing 7-17. Example of How to Change the Delivery Mode to Nonpersistent

```
MessageProperties properties = MessagePropertiesBuilder.newInstance()
    .setDeliveryMode(MessageDeliveryMode.NON_PERSISTENT)
    .build();
rabbitTemplate.getMessageConverter().toMessage(challengeAttempt, properties);
rabbitTemplate.convertAndSend(challengesTopicExchange,
    routingKey,
    event);
```

This example also demonstrates why it is important to know the configuration options of the tool we use. Sending all messages as persistent brings us a nice advantage, but it has an extra cost in performance. If we would configure a cluster of RabbitMQ instances that it's properly distributed, chances of the whole cluster going down would be minimal, so we might prefer to accept a potential loss of messages to improve performance. Again, this depends on your requirements; missing some score is not the same as missing a purchase order in an online store, for example.

Transactional

Our previous test exposed an undesired situation, but it was difficult to spot it. When we send the attempt while the broker is down, we get a server error with a 500 error code. That gives the API client the impression that the attempt wasn't processed correctly. However, it was partially processed.

Let's test that part again, but, this time, we'll check the database entries. We only need the Multiplication microservice running and the broker stopped. Then, we send an attempt with a user alias test-tx to get the error response again. See Listing 7-18.

Listing 7-18. Error Response When the Broker Is Unreachable

```
$ http POST :8080/attempts factorA=15 factorB=20 userAlias=test-tx guess=300
HTTP/1.1 500
[...]
{
    "error": "Internal Server Error",
    "message": "",
    "path": "/attempts",
    "status": 500,
    "timestamp": "2020-05-24T10:48:37.861+00:00"
}
```

Now, we navigate to the H2 console for the Multiplication database at `http://localhost:8080/h2-console`. Make sure you connect using the URL `jdbc:h2:file:~/multiplication`. Then, we run this query to get all data from both tables where the user's alias is `test-tx`:

```
SELECT * FROM USER u, CHALLENGE_ATTEMPT a WHERE u.ALIAS = 'test-tx' AND u.ID = a.USER_ID
```

The query gives us one result, as shown in Figure 7-22. That means the attempt was stored even though we got an error response. This is a bad practice because the API client doesn't know the result of the challenge, so it can't display a proper message. Yet the challenge has been saved. However, it's the expected outcome provided that our code persists the object before trying to send the message to the broker.

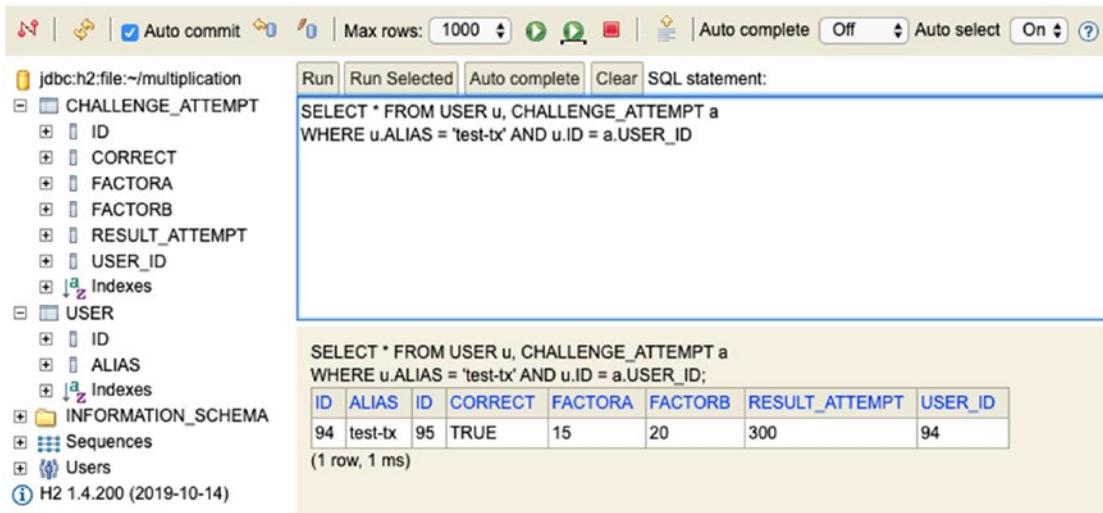


Figure 7-22. H2 console: record is stored despite the failure

Instead, we could treat the whole logic included in the service method `verifyAttempt` as a *transaction*. A database transaction can be rolled back (not executed). That's what we want if we get an error even after calling the `save` method in the repository. Doing this is easy with the Spring framework since we just need to add a Java Transaction API (JTA) annotation to our code, `javax.transaction.Transactional`. See Listing 7-19.

Listing 7-19. Adding the @Transactional Annotation to the Service Logic in Multiplication

```
@Transactional
@Override
public ChallengeAttempt verifyAttempt(ChallengeAttemptDTO attemptDTO) {
    // ...
}
```

If there is an exception in a method annotated with `@Transactional`, the transaction will be rolled back. If we would need all our methods within a given service to be transactional, we can add instead this annotation at the class level.

You can try again the same scenario steps after applying this change. Build and restart the Multiplication microservice and send a new attempt while the broker is down, this time with a different alias. If you run the corresponding query to see if the attempt was stored, you'll find that this time it isn't. Spring rolls back the database operation due to the thrown exception, so it's never performed.

Spring also supports RabbitMQ transactions, both on the publisher and subscriber sides. When we send messages with `AmqpTemplate` or its `RabbitTemplate` implementation within the scope of a method annotated with `@Transactional` and when we enable transactionality in the channel (RabbitMQ), these messages don't reach the broker if an exception happens even after the method call to send them. On the consumer's side, it's also possible to use transactions to reject messages that have been already processed. In this case, the queue would need to be set up to requeue the rejected messages (which is the default behavior). The section Transactions in the Spring AMQP documentation explains how they work in detail; see <https://tpd.io/rmq-tx>.

In many cases like ours, we can simplify the strategy for transactions and limit it only to the database.

- While publishing, if we have only one broker operation, we can publish the message at the end of the process. Any error that happens before or while sending the message will cause the rollback of the database operation.
- On the subscriber's side, the message will be rejected by default if there is an exception, and we can requeue it if that's what we want. Then, we could also use the `Transactional` annotation in our `newAttemptForUser`'s service method, so database operations will be rolled back too in case of a failure.

Local transactionality within a microservice is critical to keep data consistent and avoid partially completed processes inside a domain. Therefore, you should think of things that can go wrong in your business logic when it entails multiple steps with possible interactions with external components such as a database or a message broker.

Exercise

Add the `@Transactional` annotation to the `GameServiceImpl` service so either both scorecards and badges are stored or none of them if something goes wrong. We already decided to discard messages if we can't process them, so we don't need transactionality over the message broker's operation.

Scaling Up Microservices

Until now, we have been running a single instance of each microservice. As we described previously in the book, one of the main advantages of a microservice architecture is that we can scale up parts of our system independently. We also listed this feature as a benefit of introducing the event-driven approach with a message broker: we can add more instances of publishers and subscribers transparently. However, we can't claim that our architecture supports adding more copies of each microservice yet.

Let's focus on the first reason why our applications can't work with multiple instances: the database. When we scale up microservices horizontally, all the replicas should share the data tier and store the data in a common place, not isolated per instance. We say microservices must be stateless. The reason is that different requests or messages may end up in different microservice instances. For example, we should not assume that two attempts sent from the same user are going to be processed by the same Multiplication instance, so we can't keep any in-memory state across attempts. See Figure 7-23.

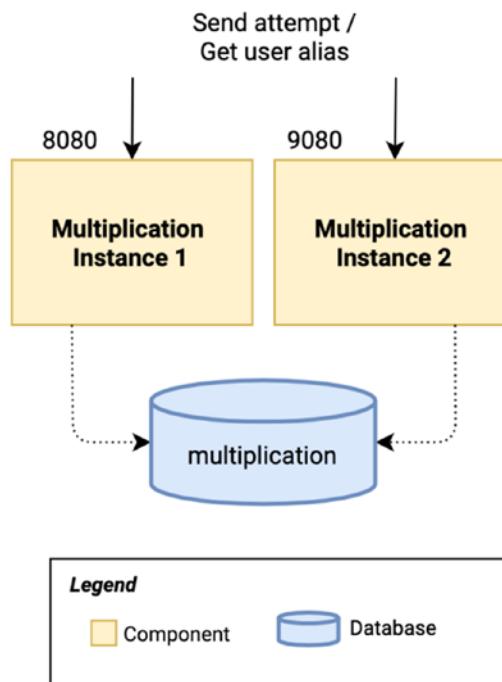


Figure 7-23. Scaling up: interface questions

The good news is that our microservices are already stateless, we process each request or message independently, and the result ends up in the database. However, we have a technical issue. If we boot up a second Multiplication instance on port 9080, it'll fail to start because it tries to create a new database instance. That's not what we want, since it should connect to a common database server shared across replicas. Let's reproduce this error. First, run the Multiplication microservice as usual (our first instance).

To start a second instance of a given service locally, we just need to override the `server.port` parameter, so we avoid port conflicts. You can do this from your IDE or using the command line from the multiplication microservice directory.

```
$ ./mvnw spring-boot:run -Dspring-boot.run.arguments="--server.port=9080"
```

When you start this second replica, the logs prompt the following error:

```
[...] Database may be already in use: null. Possible solutions: close all other connection(s); use the server mode [90020-200]
```

This error happened because we're using the H2 database engine, which is designed to behave as an embedded process by default, not as a server. Anyway, H2 supports the *server mode* as the error message suggests. The only thing we need to do is to add a parameter to both URLs that we used to connect to the database from our microservices. Then, the first time the engine starts, it'll allow other instances to use the same file and therefore the same database. Remember to apply this change to both the Multiplication and Gamification microservices. See Listing 7-20.

Listing 7-20. Enabling the Server Mode in H2 to Connect from Multiple Instances

```
# ... other properties
# Creates the database in a file (adding the server mode)
spring.datasource.url=jdbc:h2:file:~/multiplication;DB_CLOSE_ON_EXIT=FALSE;AUTO_SERVER=true;
```

Now, we can start multiple instances of each microservice, and they'll share the same data tier across replicas. First problem solved.

The second challenge we face is about load balancing. If we start two instances of each application, how do we connect to them from the user interface? This same question also applied to the REST API call we had at the end of the previous chapter

between both microservices: which Gamification instance would we call to send the attempt? If we want to balance the system's HTTP traffic across copies, we need something else. In the next chapter, we'll cover HTTP load balancing in detail.

For now, let's focus on how the message broker helped us achieve load balancing between RabbitMQ message subscribers. See Figure 7-24.

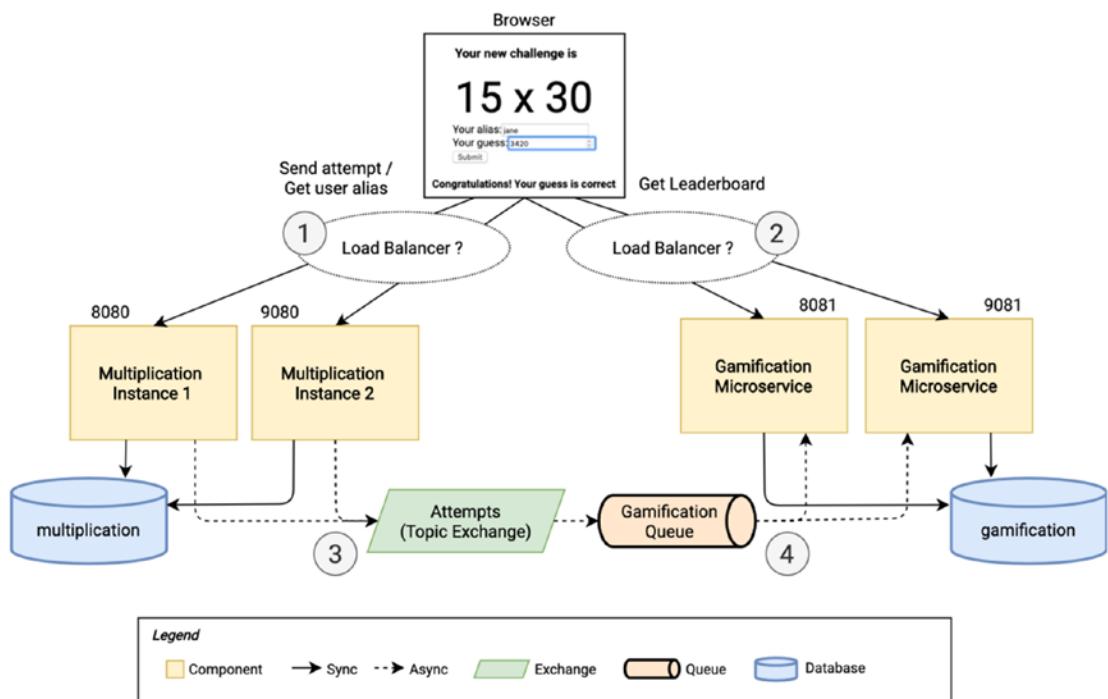


Figure 7-24. Scaling up: interface questions

There are four numbered interfaces in Figure 7-24. As we said, we'll see how to implement an HTTP load balancer pattern in the next chapter, so let's examine how interfaces 3 and 4 work with multiple sources.

Message brokers like RabbitMQ support message publication from multiple sources. That means we can have more than one copy of the Multiplication microservice publishing events to the same topic exchange. This works transparently: these instances open different connections, declare the exchange (it'll be created only the first time), and publish data without needing to know that there are other publishers. On the subscriber's side, we already learned how a RabbitMQ queue can be shared by multiple

consumers. When we start more than one instance of the Gamification microservice, all declare the same queue and binding, and the broker is smart enough to do the load balancing between them.

So, we solved load balancing at the message level. It turns out we don't need to do anything. Now, let's see this working in practice.

Follow the same steps as in previous scenarios to boot up one instance of each microservice, the UI, and the RabbitMQ service. Then, run the commands in Listing 7-21 in two separate terminals to have the same setup as shown in the previous Figure 7-23, with two replicas of each microservice. Keep in mind that you need to execute them from each corresponding microservice's home folder.

Listing 7-21. Starting a Second Instance of Each Microservice

```
multiplication $ ./mvnw spring-boot:run -Dspring-boot.run.arguments="--server.  
port=9080"  
[... logs ...]  
gamification $ ./mvnw spring-boot:run -Dspring-boot.run.arguments="--server.  
port=9081"  
[... logs ...]
```

Once we get all instances up and running, enter four correct attempts with the same new alias from the UI. Note that the attempts will hit only the first instance of our Multiplication microservice, but the event consumption is balanced across both Gamification copies. Check the logs to verify how each application should have processed two events. Besides, since the database is shared across instances, it doesn't matter that the UI is requesting the leaderboard from the instance running at port 8081. This instance will aggregate all the scorecards and badges stored by all replicas. See Figure 7-25.

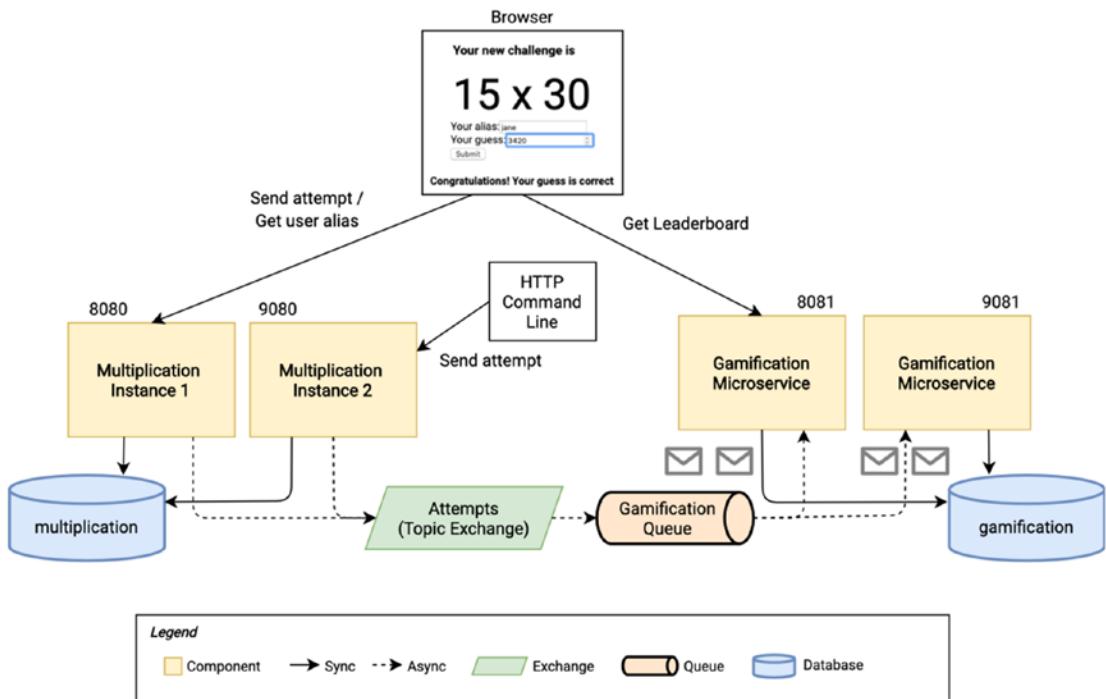


Figure 7-25. Scaling up: first test

As shown in Figure 7-25, we can also verify that multiple publishers work together using the command line to send correct attempts to the second instance of the Multiplication microservice. Let's send a few calls to the instance located at port 9080 and check how they're also processed. As expected, in this case, the messages are also balanced across subscribers. See Listing 7-22 for an example call to the second instance.

Listing 7-22. Sending a Correct Attempt to the Second Instance of Multiplication

```
$ http POST :9080/attempts factorA=15 factorB=20 userAlias=test-multi-pub
guess=300
```

That's a great achievement. We demonstrated how the message broker helps reach a good system scalability, and we implemented a worker queue pattern with multiple subscriber instances sharing the load between them.

As a consequence, we also improved resilience. In the previous section "Gamification Becomes Unavailable," we stopped the Gamification instance and saw how it'll catch up with the pending events when it becomes alive again. With the introduction of multiple instances, if one of them becomes unavailable, the broker will

automatically direct all messages to the other ones. You can try this by stopping now the first instance of Gamification (running on port 8081). Then, send two correct attempts and check in the logs how they're both processed successfully by the second instance. With this test, you can also verify that this resilience's improvement is limited—for now—to the event consumer interface. The UI can't balance the load or detect that one replica is down. Consequently, the UI doesn't display the leaderboard because the browser is trying to access the first Gamification instance. We'll fix these problems in the next chapter.

Summary and Achievements

This chapter introduced an important concept that is usually tied to microservice architectures: the event-driven software patterns. To give you a complete background, we focused first on one of the most popular tools we can use to implement it: the message broker.

We learned how a message broker can help us achieve loose coupling between our microservices, the same as similar patterns have helped other service-oriented architectures in the past years. The event pattern goes an extra step toward loose coupling by modeling a type of message that is not directed to any specific target since it just represents a fact that happened in a particular domain. Different consumers can then subscribe to these event streams and react upon them, possibly triggering their own business logic, which could produce additional events, and so forth. We saw how we can combine event-driven strategies with the publish-subscribe and worker queue patterns to have clean cuts between domains and improve the scalability of our system.

RabbitMQ, with its AMQP implementation, has some tools that we used to build our new architecture: exchanges to publish event messages, queues to subscribe to them, and bindings to link them with an optional filter. Not only did we learn the core concepts about these messaging entities, but also we covered some configuration options around message acknowledgment, message rejection, and persistence. Remember that you may need to fine-tune the RabbitMQ configuration to adjust it to your functional and nonfunctional requirements.

The coding part in this chapter remained simple thanks to the Spring Boot abstractions. We integrated RabbitMQ in our Spring Boot applications via Spring AMQP. We declared the broker entities as beans, made use of the `AmqpTemplate` to publish messages, and used the `@RabbitListener` annotation to consume them. The

Multiplication microservice is no longer aware of the Gamification microservice; it just publishes an event when an attempt is processed. We finally achieved loose coupling with our new event-driven software architecture.

A relevant part of this chapter is the last section, where we went through different scenarios to demonstrate that the patterns we implemented really helped us improve resilience and scalability, provided that we build our code with these nonfunctional requirements in mind.

The good news about the concepts in this chapter is that, once you grasp them, you can apply them to other systems using different technologies. An event-driven software architecture based on Scala and Kafka, for example, faces the same challenges and usually requires similar patterns: multiple subscribers for the same Kafka topic, load balancing between consumers (using consumer groups), configuring delivery guarantees like at-least-once and at-most-once, etc. Just remember that, with different tools, you may get different pros and cons.

At this stage, I hope you observed already that the important part of building a good software architecture is understanding the design patterns and how they relate to the functional and nonfunctional requirements. Only after you get to know these patterns can you analyze the tools and frameworks that implement them, comparing the characteristics they offer.

Sometimes, we may want to build an event-driven architecture because we think it's the best technical solution, but it might not be the best pattern for our business requirements. We should avoid these situations because the software will tend to evolve to adapt to the real business case, and that can cause many problems. I've seen microservice architectures that are plagued with synchronous calls between them, either because the requirements were not adapted to embrace eventual consistency or simply because that's not even possible according to the functional requirements. Take sufficient time to analyze the problem you want to solve before jumping into technical solutions, and be skeptical about new architectural patterns that promise to solve all possible requirements.

While we were developing our system, we've encountered a few new challenges that we couldn't figure out yet. We need HTTP load balancing with unavailable instance detection. Besides, our UI is pointing directly to each microservice, so it's aware of the back-end structure. Then, we also feel how managing our system is getting harder in

aspects such as starting it up or checking logs in multiple places. The complexity of a microservice architecture is starting to become more noticeable. In the next chapter, we'll cover some patterns and tools that help us deal with this complexity.

Chapter's Achievements:

- You learned the core concepts of event-driven architectures. For that, you got a good knowledge base of how message brokers work.
- You went through the pros and cons of event-driven architectures to know when it makes sense to apply this pattern in your future projects.
- You understood how to implement different messaging patterns depending on your use cases.
- You applied all the learned concepts using a RabbitMQ message broker in our practical case.
- You learned how Spring Boot abstracts many functionalities of RabbitMQ, allowing you to do a lot with just some little code additions.
- You refactored a tight-coupled system and converted it into a proper event-driven architecture.
- You played with the application to understand how resilience works, how you can scale up your consumers, and how you deal with transactionality.

CHAPTER 8

Common Patterns in Microservice Architectures

Surely, you realized already how we're transitioning in the previous two chapters from solutions that fulfill functional requirements to pattern implementations that don't add any business functionality. Yet, they're essential for our system to be more scalable, to be more resilient, or to have better performance.

When we finished the Gamification microservice and connected its logic to the web client, we completed the new requested feature in user story 3. However, the new functional requirements came together with other nonfunctional requirements: system capacity, availability, and organizational flexibility.

After some analysis, we decided to move to a distributed system architecture based on microservices because that approach would bring advantages to our case study.

We started as simple as possible, connecting services synchronously over HTTP and pointing the user's interface to our two microservices. Then, using our practical example, we saw how that approach would undermine our plan, due to the tight coupling between microservices and the inability to scale. To fix that situation, we adopted asynchronous communication and eventual consistency, and we introduced a message broker to implement an event-driven architecture design. As a result, we solved the tight-coupling challenge, and now our microservices are nicely isolated. We even covered load balancing partially since the broker is taking care of it for event consumers.

While we were progressing with our architecture, we briefly described other patterns that would help us achieve our target, like the gateway pattern or the load balancer for HTTP interfaces. Besides these clear needs, there are some other essential practices for microservice architectures that we didn't introduce yet: service discovery, health detection, configuration management, logging, tracing, end-to-end testing, etc.

In this chapter, we'll go through all these patterns using our system as an example. That will help you understand the problem before we dive into the solution. Our journey through the microservice's common patterns and tools has just started.

Gateway

We know already some problems that the gateway pattern would solve in our architecture.

- Our React application needs to point to multiple back-end microservices to interact with their APIs. This is wrong because the front end should treat the back end as a single server with multiple APIs. Then, we don't expose our architecture, thus making it more flexible in case we want to make changes in the future.
- If we introduce multiple instances of our back-end services, our UI doesn't know how to balance the load between them. Neither would it know how to redirect all requests to a different back-end instance if one of the instances is not available. Even though it's technically possible to implement load balancing and resilience patterns in our web clients, this is logic that we should place on the back-end side. We implement it only once, and it's valid for any client. Also, we keep the front end's logic as simple as possible.
- With the current setup, if we add user authentication to our system, we would need to validate the security credentials in every back end's microservice. It seems more logical to put this logic *in the edge* of our back end, validate API calls there, and pass simple requests to other microservices. As long as we make sure the rest of the back-end services are not reachable from the outside, they don't need to take care of security concerns.

In this first section of the chapter, we'll introduce a Gateway microservice in our system to solve these problems. The gateway pattern centralizes the HTTP access and takes care of proxying requests to other underlying services. Usually, the gateway makes the decision for where to route a request based on some configured rules (a.k.a. predicates). Additionally, this routing service can modify requests and responses as they

pass through, with pieces of logic that are called *filters*. We'll soon put in practice rules and filters in our implementation to understand them better. See Figure 8-1 for a high-level overview of how the gateway fits into our system.

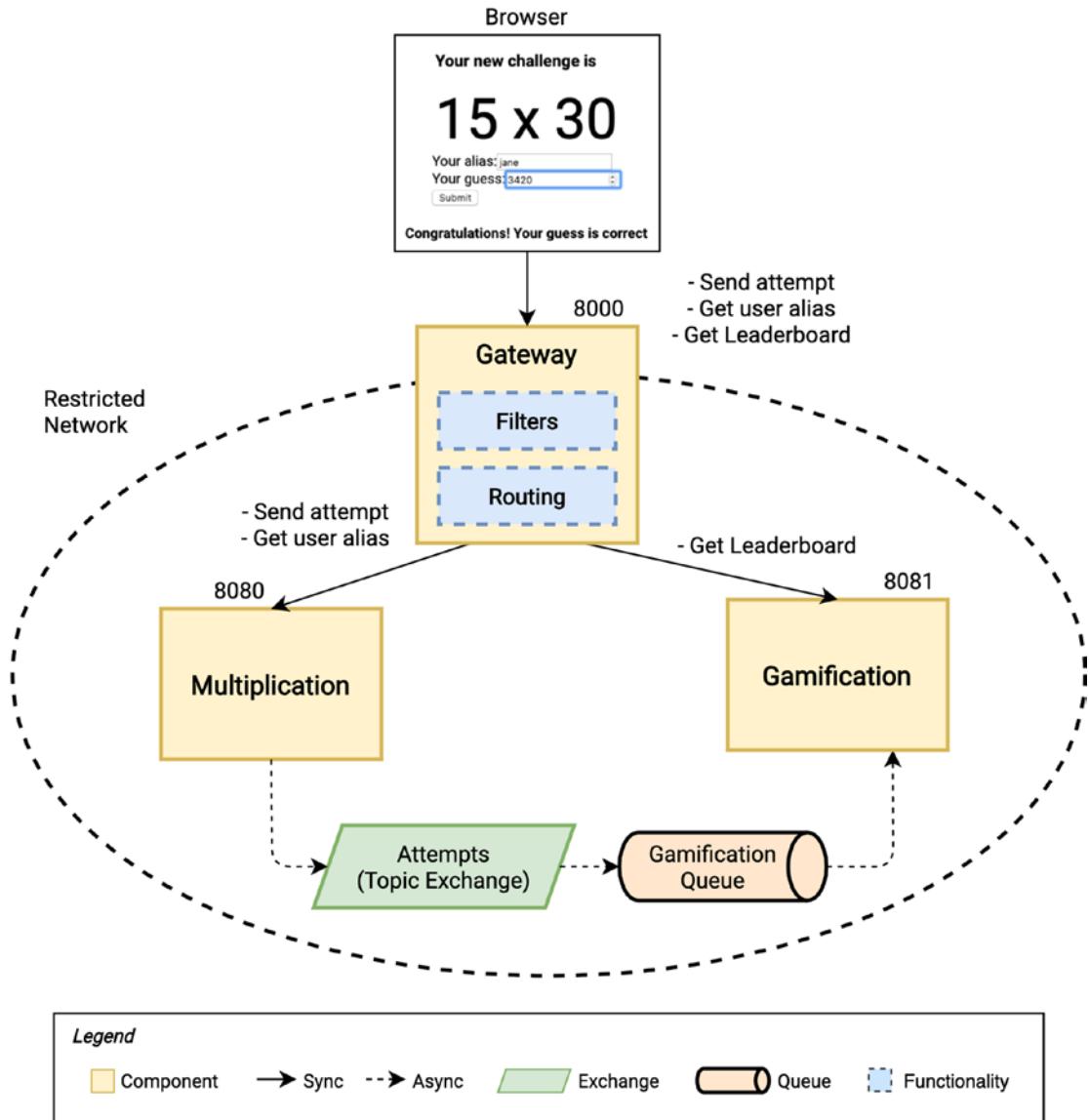


Figure 8-1. Gateway: high-level overview

Sometimes people refer to the gateway as an *edge service* because it's the way other systems have to access our back end, and it's routing external traffic to the corresponding internal microservices. As mentioned earlier, the introduction of a gateway normally comes with the restriction of access to other back-end services. For the first part of this chapter, we'll skip this restriction since we're running all services directly on our machine. We'll change this when we introduce *containerization*.

Spring Cloud Gateway

Spring Cloud is a separate group of projects within the Spring family that provide tools to quickly build common patterns required in distributed systems like our microservices case. These patterns are also called *cloud patterns*, although they are also applicable even if you deploy microservices in your own servers. We'll make use of several Spring Cloud projects within this chapter. If you want to check the complete list, check the overview page (<https://tpd.io/scloud>) in the reference documentation.

For the gateway pattern, Spring Cloud offers two options. The first alternative is to use an integration that Spring Cloud has been supporting for a long time: Spring Cloud Netflix. This is a project that includes several tools that Netflix developers have been publishing and maintaining as open source software (OSS) for many years. You can take a look at the Netflix OSS website (<https://tpd.io/noss>) if you want to know more about these tools. The component within Netflix OSS that implements the gateway pattern is Zuul, and its integration with Spring comes via the Spring Cloud Netflix Zuul module.

In this second edition of the book, we won't use Spring Cloud Netflix. The main reason is that Spring seems to be moving away from the Netflix OSS tool integrations and replacing them with other modules that integrate alternative tools, or even with their own implementations. A possible explanation for this change is that Netflix put some of its projects in maintenance mode, like Hystrix (circuit-breaking) and Ribbon (load balancing), so they are no longer in active development. This decision also affects other tools in the Netflix stack since they implement patterns that are often used together. An example is Eureka, the service discovery tool, which relies on Ribbon for load balancing.

We'll go for a newer alternative to implement the gateway pattern: the Spring Cloud Gateway. In this case, this replacement for Zuul is a stand-alone Spring project, so it doesn't depend on any external tool.

Knowing the Patterns, You Can Exchange Tools

Keep in mind that what's important to learn from this book are the microservice architecture patterns and the reasons why they're introduced from a practical perspective. You could use any other alternative in the market such as Nginx, HAProxy, Kong Gateway, etc.

The Spring Cloud Gateway project defines some core concepts (also shown in Figure 8-2).

- *Predicate*: A condition to be evaluated to decide where to route a request. Cloud Gateway provides a bunch of condition builders based on the request path, request headers, time, remote host, etc. You can even combine them as expressions. They're also known as *route predicates* since they always apply to a route.
- *Route*: It's the URI where the request will be proxied to if it matches the assigned predicate. For example, it could address an external request to an internal microservice endpoint, as we'll see later in practice.
- *Filter*: An optional processor that can be either attached to a route (route filters) or applied globally (global filters) to all the requests. Filters allow modifying requests (incoming filters) and responses (outgoing filters). There are a lot of built-in filters in Spring Cloud Gateway, so you can, for example, add or remove headers in the request, limit the number of requests from a given host, or transform a response from the proxied service before returning it to the requester.

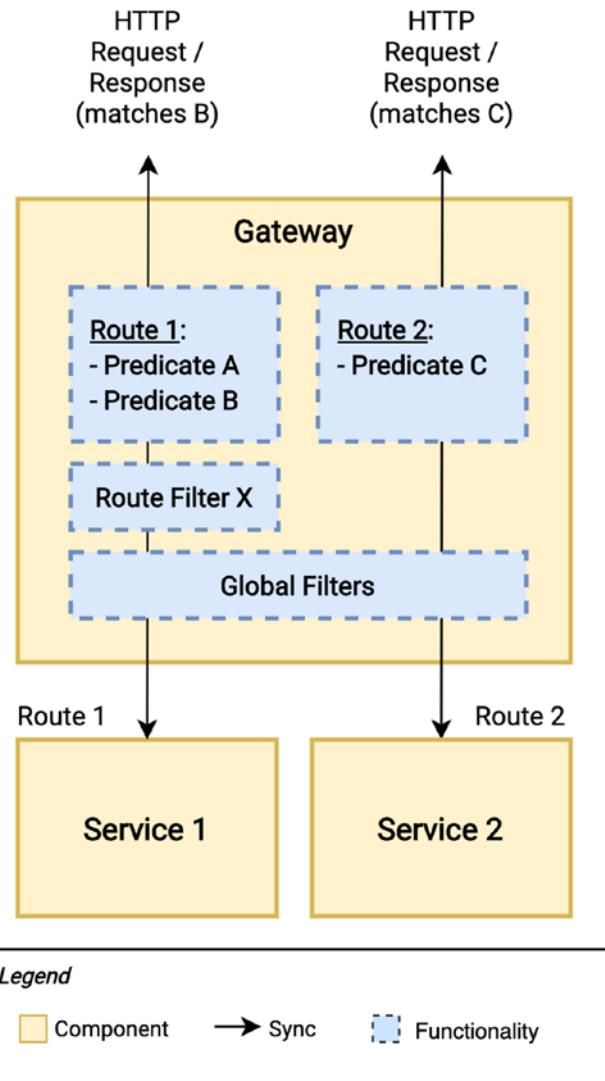


Figure 8-2. Gateway: routes, predicates, and filters

To define this configuration, we use Spring Boot's application properties. However, this time we'll use the YAML format because it's more readable for defining routes. The Cloud Gateway documentation defines a specific notation for predicates, routes, and filters. Additionally, we have two options when defining predicates and filters: *shortcut* and *fully expanded* configuration. They both work the same; the only difference is that you can use one-line expressions and avoid extra YAML with the shortcut version. If you want to see how they compare to each other, check the section "Shortcut Notation"

(<http://tpd.io/gw-notation>) in the docs. See Listing 8-1 for an example block of configuration that uses shortcut notation to define two routes. Keep reading for a detailed explanation of how they work.

Listing 8-1. An Example of Routing Configuration in Spring Cloud Gateway

```
spring:
  cloud:
    gateway:
      routes:
        - id: old-travel-conditions
          uri: http://oldhost/travel
          predicates:
            - Before=2021-01-01T10:00:00.000+01:00[Europe/Madrid]
            - Path=/travel-in-spain/**
        - id: change-travel-conditions
          uri: http://somehost/travel-new
          predicates:
            - After=2021-01-01T10:00:00.000+01:00[Europe/Madrid]
            - Path=/travel-in-spain/**
          filters:
            - AddResponseHeader=X-New-Conditions-Apply, 2021-Jan
```

Imagine that the gateway is externally accessible at `http://my.travel.gateway/`. This example configuration defines two routes that share a path route predicate (included in the gateway); see <https://tpd.io/pathpred>. Any request that starts with `http://my.travel.gateway/travel-in-spain/` is captured by that predicate definition. The additional condition in each route, defined by a Before (<https://tpd.io/befpred>) and an After route (<https://tpd.io/aftpred>) predicate, respectively, is what determines where to proxy the request.

- If the request happens before January 1, 2021, at 10 a.m. in Spain, it'll be proxied to `http://oldhost/travel-conditions/`. For example, the request `http://my.travel.gateway/travel-in-spain/tapas` gets proxied to `http://oldhost/travel/tapas`.
- Any request that happens after that time is captured by the change-travel-conditions route since it's using the counterpart predicate, After. In this case, the same request shown previously would be

proxied to `http://somehost/travel-new/tapas`. Additionally, the extra filter will add a response header, `X-New-Conditions-Apply`, with a value of `2021-Jan`.

Keep in mind that `http://oldhost` and `http://somehost` in this example don't need to be accessible from the outside; they're only visible to the gateway and other internal services in our back end.

The built-in predicates and filters allow us to fulfill a wide variety of requirements we may have for our gateway. In our application, we'll use mainly path route predicates to proxy external requests to the corresponding microservice, based on the API they're calling.

If you want to extend your knowledge about the Spring Cloud Gateway capabilities, check the reference docs (<https://tpd.io/gwdocs>).

The Gateway Microservice

Code Source

The code source in this chapter has been split into four parts. This way, you can better understand how the system evolves in smaller steps. The code source for this first part, including the Gateway implementation, are in the project `chapter08a`.

As you can imagine, Spring Boot provides a starter package for Spring Cloud Gateway. Only by adding this starter dependency to an empty Spring Boot application, we get a Gateway microservice that is ready to use. Actually, the Gateway project is built on top of Spring Boot, so it can work only within a Spring Boot application. For that reason, the autoconfiguration logic is located in this case within the core Spring Cloud Gateway artifact, and not in the Spring Boot's autoconfigure package. The class name is `GatewayAutoConfiguration` (see <https://tpd.io/gwautcfg>), and, among other tasks, it reads the `application.yml` configuration and builds the corresponding routing filters, predicates, etc.

We'll build this new microservice as usual, via Spring Initializr's website (see <https://tpd.io/spring-start>). Select the Gateway dependency and name the artifact `gateway`, as shown in Figure 8-3.

The screenshot shows the Spring Initializr web application interface. At the top, there's a logo and a search bar. Below the search bar, there are sections for Project type (Maven Project selected), Language (Java selected), and Spring Boot version (2.3.3 selected). A 'Dependencies' section includes an 'ADD ...' button and a 'Gateway' dependency listed under 'SPRING CLOUD ROUTING'. The 'Project Metadata' section contains fields for Group (microservices.book), Artifact (gateway), Name (gateway), Description (Gateway Microservice), Package name (microservices.book.gateway), Packaging (Jar selected), and Java version (14 selected). The Java version dropdown has options 14, 11, and 8.

Figure 8-3. Creating the Gateway microservice

After downloading the zip file, we copy its contents inside our main workspace folder, at the same level as the Multiplication and Gamification microservices. Load the project as an extra module in your workspace, and take a moment to explore the contents of the generated `pom.xml` file. When compared with the other projects, you'll see there a new `dependencyManagement` node and a new property for the Spring Cloud version to use (Hoxton). See Listing 8-2 for the main changes in the file. We require this additional Maven configuration since the Spring Cloud artifacts are not defined directly in the Spring Boot's parent project.

Listing 8-2. Spring Cloud Gateway Dependencies in Maven

```

<?xml version="1.0" encoding="UTF-8"?>
<project>
    <!-- ... -->
    <name>gateway</name>

    <properties>
        <spring-cloud.version>Hoxton.SR7</spring-cloud.version>
        <!-- ... -->
    </properties>

    <dependencies>
        <dependency>
            <groupId>org.springframework.cloud</groupId>
            <artifactId>spring-cloud-starter-gateway</artifactId>
        </dependency>
        <!-- ... -->
    </dependencies>

    <dependencyManagement>
        <dependencies>
            <dependency>
                <groupId>org.springframework.cloud</groupId>
                <artifactId>spring-cloud-dependencies</artifactId>
                <version>${spring-cloud.version}</version>
                <type>pom</type>
                <scope>import</scope>
            </dependency>
        </dependencies>
    </dependencyManagement>
    <!-- ... -->
</project>

```

The next step is to change the extension of the application.properties file to application.yml and add some configuration to proxy all the endpoints that belong to the Multiplication microservice to that application, and it's the same for Gamification's endpoints. We also change the server port of this new service to 8000 to avoid port conflicts when deploying locally. Additionally, we'll append some CORS configuration

for the UI to be allowed to make requests from its origin. Spring Cloud Gateway has a configuration-based style (<https://tpd.io/gwcors>) to accomplish this using the `globalcors` properties. See all these changes in Listing 8-3.

Listing 8-3. Gateway Configuration: First Approach

```
server:
  port: 8000

spring:
  cloud:
    gateway:
      routes:
        - id: multiplication
          uri: http://localhost:8080/
          predicates:
            - Path=/challenges/**,/attempts,/attempts/**,/users/**
        - id: gamification
          uri: http://localhost:8081/
          predicates:
            - Path=/leaders
      globalcors:
        cors-configurations:
          '//**':
            allowedOrigins: "http://localhost:3000"
            allowedHeaders:
              - "*"
            allowedMethods:
              - "GET"
              - "POST"
              - "OPTIONS"
```

The routes in that file will make the Gateway act as follows:

- Any request to or under `http://localhost:8000/attempts` will be proxied to the Multiplication microservice, deployed locally at `http://localhost:8080/`. The same will happen to other API contexts located within the same microservice, like challenges and users.

- Requests to `http://localhost:8000/leaders` will be translated to requests to the Gamification microservice, which uses the same host (`localhost`) but the port `8081`.

Alternatively, it would be possible to write a simpler configuration that wouldn't require an explicit list of endpoints routed to each microservice. We could do this by using another feature of the gateway that allows capturing path segments. If we get an API call such as `http://localhost:8000/multiplication/attempts`, we could extract `multiplication` as a value and use it to map to the corresponding service's host and port. However, this approach is valid only when each microservice contains only one API domain. In any other case, we'd be exposing our internal architecture to the client. In our case, we would require the client to call `http://localhost:8000/multiplication/users`, whereas we prefer them to point to `http://localhost:8000/users` and hide the fact that the Users domain still lives within the multiplication's deployable unit.

Changes in Other Projects

With the introduction of the Gateway microservice, we can keep all the configuration intended for external requests within the same service. This means we no longer need to add CORS configuration to the Multiplication and Gamification microservices. We can keep this configuration in the gateway since the other two services are placed behind this new proxy service. As a result, we can remove both `WebConfiguration` files from the existing project folders. Listing 8-4 shows the contents of the file in the Gamification microservice. Remember to delete not only this one but the equivalent class in the Multiplication microservice.

Listing 8-4. The `WebConfiguration` Class That We Can Remove

```
package microservices.book.gamification.configuration;

import org.springframework.context.annotation.Configuration;
import org.springframework.web.servlet.config.annotation.CorsRegistry;
import org.springframework.web.servlet.config.annotation.WebMvcConfigurer;

@Configuration
public class WebConfiguration implements WebMvcConfigurer {
```

```

/**
 * Enables Cross-Origin Resource Sharing (CORS)
 * More info: http://docs.spring.io/spring/docs/current/spring-framework-
 * reference/html/cors.html
 */
@Override
public void addCorsMappings(final CorsRegistry registry) {
    registry.addMapping("/**").allowedOrigins("http://localhost:3000");
}
}

```

We also need to change the React application to point to the same host/port for both services. See Listings 8-5 and 8-6. We could also refactor the GameApiClient and ChallengesApiClient classes as per our preferences for the UI structure: one single service to call all endpoints or a service per API context (challenges, users, etc.). We don't need two different server URLs anymore since the UI now treats the back end as a single host with multiple APIs.

Listing 8-5. Changing the Multiplication API URL to Point to the Gateway

```

class ChallengesApiClient {
    static SERVER_URL = 'http://localhost:8000';
    static GET_CHALLENGE = '/challenges/random';
    static POST_RESULT = '/attempts';
    // ...
}

```

Listing 8-6. Changing the Gamification API URL to Point to the Gateway

```

class GameApiClient {
    static SERVER_URL = 'http://localhost:8000';
    static GET_LEADERBOARD = '/leaders';
    // ...
}

```

Running the Gateway Microservice

To run our entire application, we have to add an extra step to the list. Remember that all Spring Boot apps can be executed from your IDE or from the command line, using the Maven wrapper.

1. Run the RabbitMQ server.
2. Start the Multiplication microservice.
3. Start the Gamification microservice.
4. Start the new Gateway microservice.
5. Run the front-end app with `npm start` from the `challenges-frontend` folder.

Within this chapter, this list will keep growing. It's important to highlight that you don't need to follow the order of the previous steps since you could even run all of these processes at the same time. During the start, the system might be unstable, but it'll eventually become ready. Spring Boot will retry to connect to RabbitMQ until it works.

When you access the UI, you won't notice any changes. The leaderboard gets loaded, and you can send attempts as usual. One way to verify that the requests are being proxied is to look at the Network tab in the browser's developer tools and select any request to the back end to see how the URL begins now with `http://localhost:8000`. A second option is to add some trace logging configuration to the gateway, so we see what's happening. See Listing 8-7 for the configuration you can add to the application. `yml` file in the Gateway project to enable these logs.

Listing 8-7. Adding Trace-Level Logs to the Gateway

```
# ... route config
logging:
level:
  org.springframework.cloud.gateway.handler.predicate: trace
```

If you restart the gateway with this new configuration, you'll see logs per request that the gateway handles. These logs are the result of going through all defined routes to see whether there is any that matches the request pattern. See Listing 8-8.

Listing 8-8. Gateway Logs with Pattern-Matching Messages

```
TRACE 48573 --- [ctor-http-nio-2] RoutePredicateFactory: Pattern "[/challenges/**, /attempts, /attempts/**, /users/**]" does not match against value "/leaders"
TRACE 48573 --- [ctor-http-nio-2] RoutePredicateFactory: Pattern "/leaders" matches against value "/leaders"
TRACE 48573 --- [ctor-http-nio-2] RoutePredicateFactory: Pattern "/users/**" matches against value "/users/72,49,60,101,96,107,1,45"
```

Now, let's remove again this logging configuration to avoid a too verbose output.

Next Steps

We got already a few advantages with this new setup.

- The front end remains unaware of the back end's structure.
- The common configuration for external requests remains in the same place. In our case, that's the CORS setup, but they could be also other common concerns such as user authentication, metrics, etc.

Next, we'll introduce load balancing in our gateway, so it can distribute the traffic across all available instances of each service. With that pattern, we'll add scalability and redundancy to our system. However, to make load balancing work properly, we need some prerequisites.

- The gateway needs to know the available instances for a given service. Our initial configuration points directly to a specific port because we assumed that there is only one instance. How would that look like with multiple replicas? We shouldn't include a hard-coded list in our routing configuration since the number of instances should be dynamic: we want to bring new ones up and down transparently.
- We need to implement the concept of healthiness of a back-end component. Only then will we know when an instance is not ready to handle traffic and switch to any other healthy instance.

To fulfill the first prerequisite, we need to introduce the service discovery pattern, with a common registry that the different distributed components can access to know the available services and where to find them.

For the second prerequisite, we'll use Spring Boot Actuator. Our microservices will expose an endpoint that indicates whether they're healthy or not, so other components will know. This is what we'll do next since it's also a requirement for service discovery.

Health

A system running on a production environment is never safe against errors. Network connections may fail, or a microservice instance may crash due to an out-of-memory problem caused by a bug in the code. We're determined to build a resilient system, so we want to be prepared against these errors with mechanisms like *redundancy* (multiple replicas of the same microservice) to minimize the impact of these incidents.

So, how do we know when a microservice is not working? If it's exposing an interface (like a REST API or RabbitMQ), we could interact with a sample probe and see if it reacts to it. But then, we should be careful when selecting that probe since we want to cover all possible scenarios that would make our microservice transition to an unhealthy state (not functional). Instead of leaking the logic to determine whether a service is working or not to the caller, it's much better to provide a standard, dumb probe interface that just tells whether the service is healthy. It's up to the service's logic to decide when to transition to an unhealthy state, based on the availability of the interfaces it uses, its own availability, and the criticality of the error. If the service can't even provide a response, the callers can also assume that it's unhealthy. See Figure 8-4 for a high-level conceptual view of this health interface.

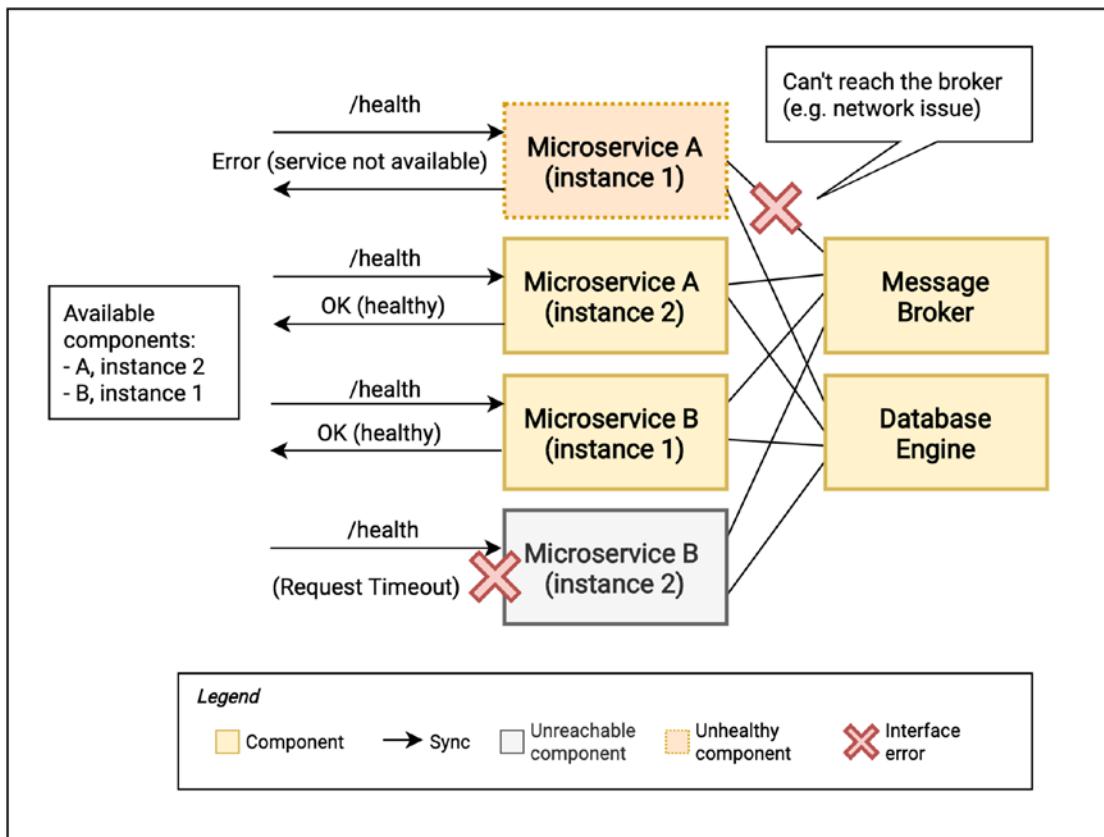


Figure 8-4. Health: high-level overview

This simple interface convention to determine the healthiness of services is required by many tools and frameworks (and not only for microservices). For example, load balancers can temporarily stop diverting traffic to an instance that doesn't respond to the health probe or responds with a nonready state. Service discovery tools may remove an instance from the registry if it's unhealthy. Container platforms like Kubernetes can decide to restart a service if it's not healthy for a configured period (we'll explain what a container platform is later in this chapter).

Spring Boot Actuator

Same as with other aspects of our application, Spring Boot provides an out-of-the-box solution to make our microservices report their health status: Spring Boot Actuator. Actually, that's not the only feature that Actuator contains; it can also expose other

endpoints to access different data about our application like the configured loggers, HTTP traces, audit events, etc. It can even open a management endpoint that allows you to shut down the application.

Actuator endpoints can be enabled or disabled independently, and they're available not only via web interface but also through Java Management eXtensions (JMX). We'll focus on the web interfaces that we'll use as REST API endpoints. The default configuration exposes only two endpoints: `info` and `health`. The first one is intended to provide general information about the application that you can enrich using contributors (<https://tpd.io/infocb>). The `health` endpoint is the one we're interested in for now. It outputs the status of our application, and to resolve it, it uses health indicators (<https://tpd.io/acthealth>).

There are multiple built-in health indicators that can contribute to the overall health status of the application. Many of these indicators are specific to certain tools, so they are available only when we use these tools in our application. That's controlled by Spring Boot's autoconfiguration, which we know already it can detect if we're using certain classes and inject some extra logic.

Let's use a practical example to see how that works: the `RabbitHealthIndicator` class included in the Spring Boot Actuator artifact. See Listing 8-9 for an overview of its source code (also available online at <http://tpd.io/rhi-source>). The health check implementation uses a `RabbitTemplate` object, Spring's way to interact with the RabbitMQ server. If this code can access the RabbitMQ server's version, the health check passes (it doesn't throw an exception).

Listing 8-9. The `RabbitHealthIndicator` Included in Spring Boot Actuator

```
public class RabbitHealthIndicator extends AbstractHealthIndicator {

    private final RabbitTemplate rabbitTemplate;

    public RabbitHealthIndicator(RabbitTemplate rabbitTemplate) {
        super("Rabbit health check failed");
        Assert.notNull(rabbitTemplate, "RabbitTemplate must not be null");
        this.rabbitTemplate = rabbitTemplate;
    }

    @Override
    protected void doHealthCheck(Health.Builder builder) throws Exception {
        builder.up().withDetail("version", getVersion());
    }
}
```

```

private String getVersion() {
    return this.rabbitTemplate
        .execute((channel) -> channel.getConnection()
            .getServerProperties().get("version").toString());
}
}

```

This indicator is automatically injected in the context if we use RabbitMQ. It contributes to the overall health status. The `RabbitHealthContributorAutoConfiguration` class, included in the artifact `spring-boot-actuator-autoconfigure` (part of Spring Boot Actuator dependency), takes care of that. See Listing 8-10 (also available at <http://tpd.io/rhc-autoconfig>). This configuration is conditional on the existence of a `RabbitTemplate` bean, which means we're using the RabbitMQ module. It creates a `HealthContributor` bean, in this case, a `RabbitHealthIndicator` that will be detected and aggregated by the overall health autoconfiguration.

Listing 8-10. How Spring Boot Autoconfigures the `RabbitHealthContributor`

```

@Configuration(proxyBeanMethods = false)
@ConditionalOnClass(RabbitTemplate.class)
@ConditionalOnBean(RabbitTemplate.class)
@ConditionalOnEnabledHealthIndicator("rabbit")
@AutoConfigureAfter(RabbitAutoConfiguration.class)
public class RabbitHealthContributorAutoConfiguration
    extends CompositeHealthContributorConfiguration<RabbitHealthIndicator,
    RabbitTemplate> {

    @Bean
    @ConditionalOnMissingBean(name = { "rabbitHealthIndicator",
        "rabbitHealthContributor" })
    public HealthContributor rabbitHealthContributor(Map<String, RabbitTemplate>
        rabbitTemplates) {
        return createContributor(rabbitTemplates);
    }
}

```

We'll see soon how this works in practice since we'll add Spring Boot Actuator to our microservices in the next section.

Keep in mind that you can configure multiple settings of the Actuator's endpoints, and you can create your own health indicators as well. For a complete list of features, check the official Spring Boot Actuator documentation (<https://tpd.io/sbactuator>).

Including Actuator in Our Microservices

Code Source

The code sources for the introduction of the health endpoints, service discovery, and load balancing are located in the repository `chapter08b`.

Adding the health endpoint to our applications is as easy as adding a dependency in the `pom.xml` file to our projects: `spring-boot-starter-actuator`. See Listing 8-11. We add this new artifact to all our Spring Boot applications: the Multiplication, Gamification, and Gateway microservices.

Listing 8-11. Adding Spring Boot Actuator to Our Microservices

```
<dependency>
  <groupId>org.springframework.boot</groupId>
  <artifactId>spring-boot-starter-actuator</artifactId>
</dependency>
```

The default configuration exposes the `health` and `info` web endpoints on the `/actuator` context. That's enough for us at this point, but this could be adjusted via properties if need it. Rebuild and restart the back-end applications to verify this new feature. We can use the command line, or our browser, to poll each service for its health status, just by switching the port number. Note that we're not exposing the `/health` endpoints via the gateway because it's not a feature we want to expose to the outside but just internal to our system. See Listing 8-12 for the request and response of the Multiplication microservice.

Listing 8-12. Testing the /health Endpoint for the First Time

```
$ http :8080/actuator/health
HTTP/1.1 200
Content-Type: application/vnd.spring-boot.actuator.v3+json

{
    "status": "UP"
}
```

If our system is running properly, we'll get the UP value and an HTTP status code 200. What we'll do next is to stop the RabbitMQ server and try this same request again. We already saw that the Actuator project contains a health indicator to check the RabbitMQ server, so this one should fail, causing the aggregated health status to switch to DOWN. This is indeed what we'll get if we make the request while the RabbitMQ server is stopped. See Listing 8-13.

Listing 8-13. The Status of Our App Switches to DOWN When RabbitMQ Is Unreachable

```
$ http :8080/actuator/health
HTTP/1.1 503
Content-Type: application/vnd.spring-boot.actuator.v3+json

{
    "status": "DOWN"
}
```

Note that the HTTP status code returned also changed to 503, Service Unavailable. Therefore, the caller doesn't even need to parse the response body; it can just check whether the response code is 200 to determine that the application is healthy. You can also see the logs of the failed attempt from RabbitHealthIndicator to retrieve the server's version in the Multiplication application's output. See Listing 8-14.

Listing 8-14. Rabbit Health Check Failing in the Multiplication Microservice

```
2020-08-30 10:20:04.019  INFO 59277 --- [io-8080-exec-10] o.s.a.r.c.CachingConnectionFactory      : Attempting to connect to: [localhost:5672]
2020-08-30 10:20:04.021  WARN 59277 --- [io-8080-exec-10] o.s.b.a.amqp.RabbitHealthIndicator      : Rabbit health check failed
```

```

org.springframework.amqp.AmqpConnectException: java.net.ConnectException:
Connection refused
    at org.springframework.amqp.rabbit.support.RabbitExceptionTranslator.convertRa-
    bbitAccessException(RabbitExceptionTranslator.java:61) ~[spring-rabbit-2.2.10.
    RELEASE.jar:2.2.10.RELEASE]
[...]
Caused by: java.net.ConnectException: Connection refused
    at java.base/sun.nio.ch.Net.pollConnect(Native Method) ~[na:na]
[...]

```

The Spring Boot application remains alive, and it can recover from that error. If we start the RabbitMQ server and check again for the health status, it'll switch to UP. The application keeps retrying to establish a connection to the server until it succeeds. This is exactly the behavior we want for a robust system: if the microservice has issues, it should flag it for other components to know; in the meantime, it should try to recover from the error and switch to a healthy state again when possible.

Service Discovery and Load Balancing

Now that we have the ability to know whether services are available, we can integrate service discovery and load balancing in our system.

The service discovery pattern consists of two main concepts.

- *The service registry:* A central place with a list of available services, the address where they're located, and some extra metadata like their name. It may contain entries for different services, but also multiple instances of the same service. In the last case, clients accessing the registry can obtain a list of available instances by querying a *service alias*. For example, various instances of the Multiplication microservice can register with the same alias, `multiplication`. Then, when querying that value, all instances are returned. That's the case of the example shown in Figure 8-5.
- *The registrar:* The logic in charge of registering the service instance at the registry. It can be an external running process observing the state of our microservice, or it can be embedded in the service itself as a library, like it'll be our case.

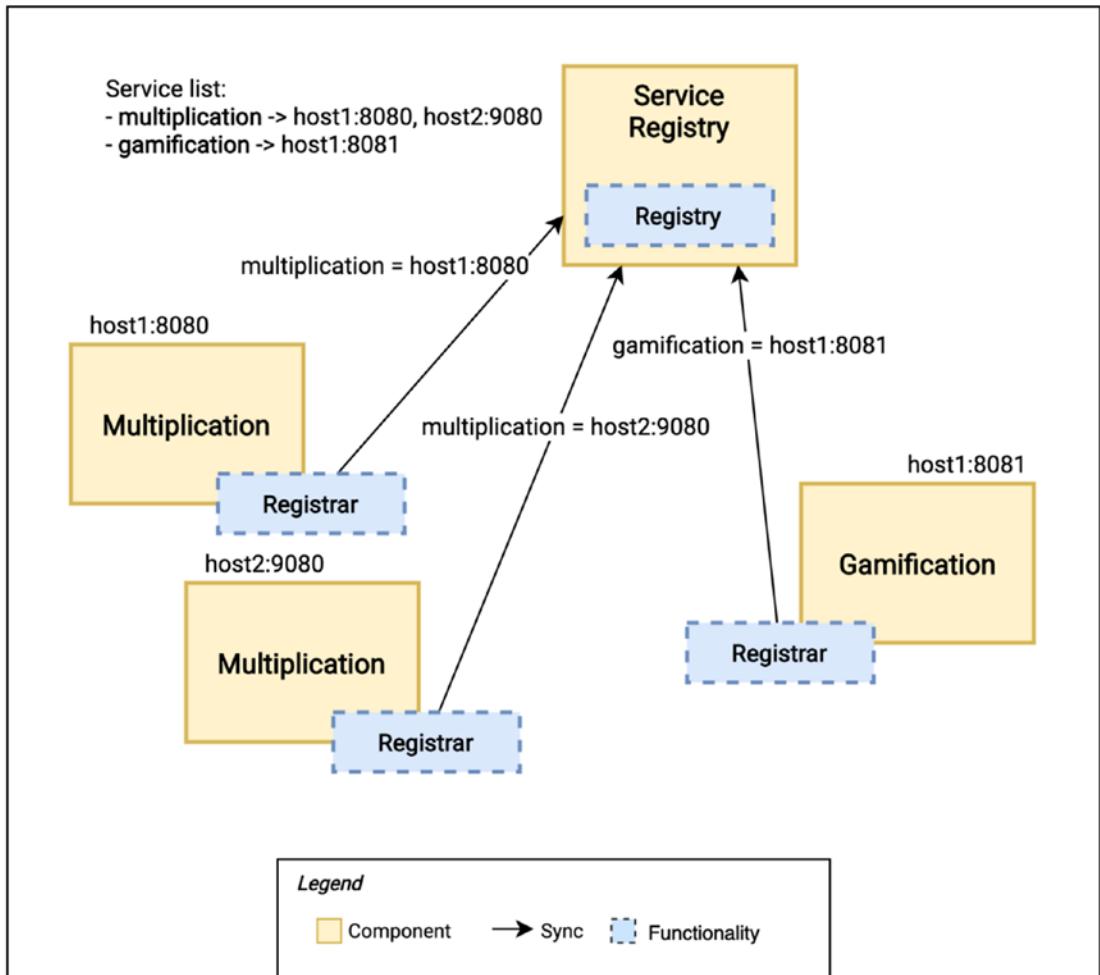


Figure 8-5. Service discovery: pattern overview

In Figure 8-5, we see an example of service registration with three services. The host1, a server's DNS address, has one instance of Multiplication at port 8080 and one instance of Gamification at port 8081. The host2, a different machine, has a second instance of Multiplication, located at port 9080. All of these instances know where they're located and send their corresponding URIs to the service registry using the registrar. Then, a Registry client can simply ask for the location of a service using its name (e.g., multiplication), and the registry returns the list of instances and their locations (e.g., host1:8080, host2:9080). We'll see this in practice soon.

The load balancing pattern is closely related to service discovery. If more than one service uses the same name upon registration, that implies there are multiple replicas available. We want to balance the traffic between them so we can increase the capacity of the system and make it more resilient in case of errors thanks to the added redundancy.

Other services may query a given service name from the registry, retrieve a list, and then decide which instance to call. This technique is known as *client-side discovery*, and it implies that clients are aware of the service registry and perform load balancing themselves. Note that, by client, we refer in this definition to an application, microservice, browser, etc., that wants to perform an HTTP call to another service. See Figure 8-6.

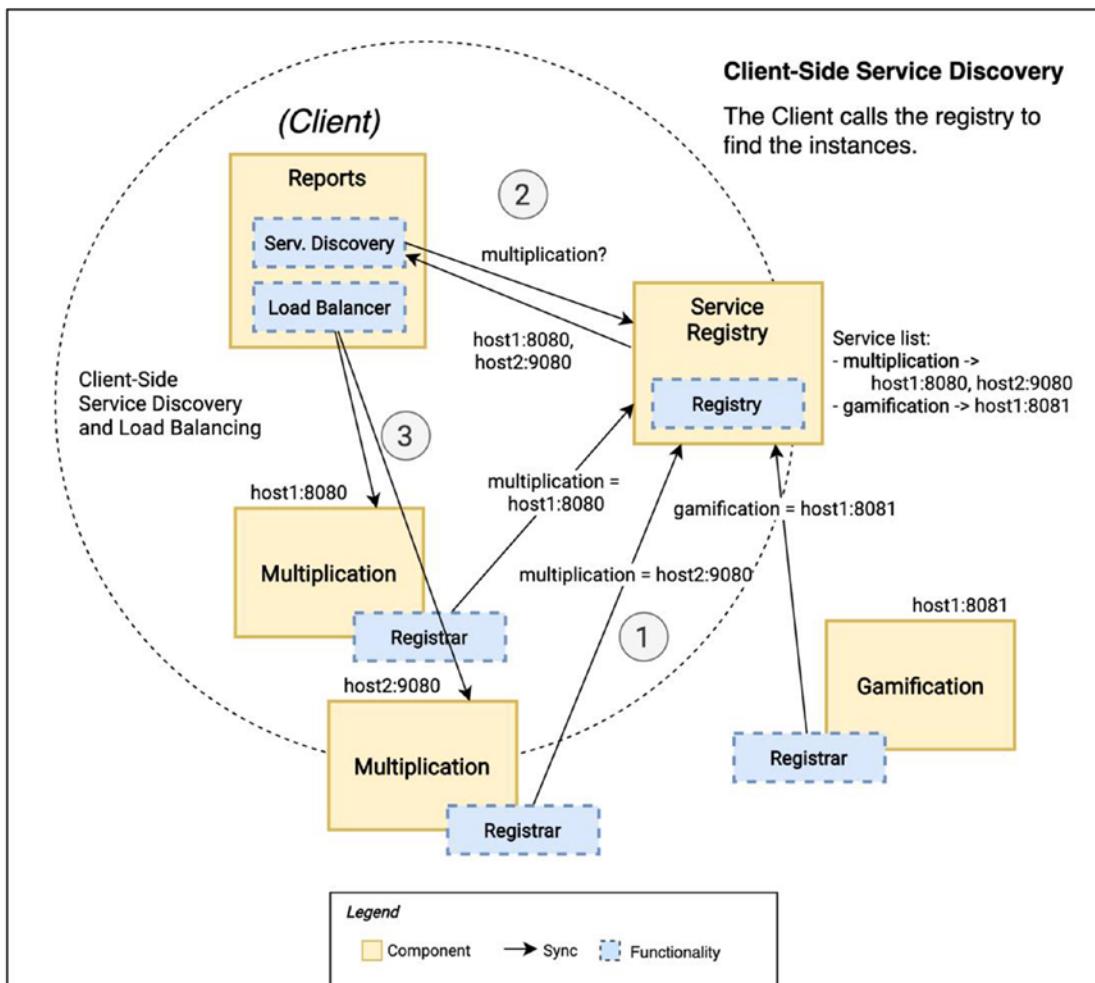


Figure 8-6. Client-side service discovery

On the other side, *server-side discovery* abstracts all this logic from the clients by providing a unique address, known in advance, where callers can find a given service. When they make the request, it's intercepted by a load balancer, which is aware of the registry. This balancer will proxy the request to one of the replicas. See Figure 8-7.

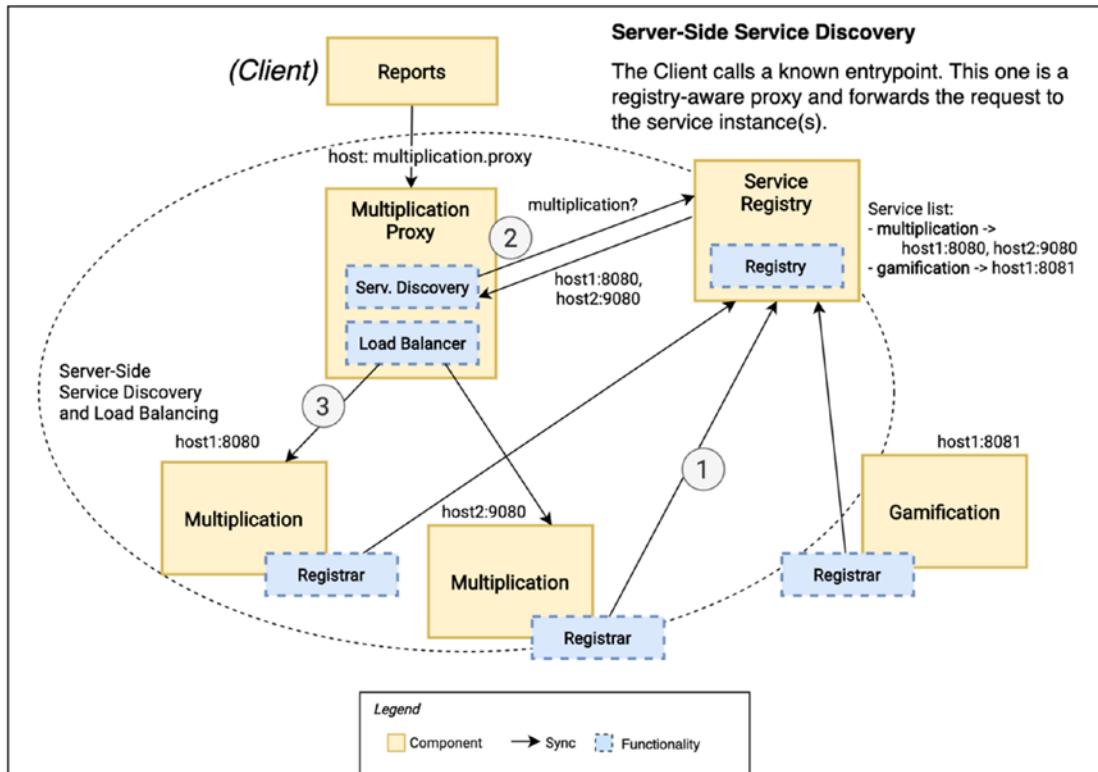


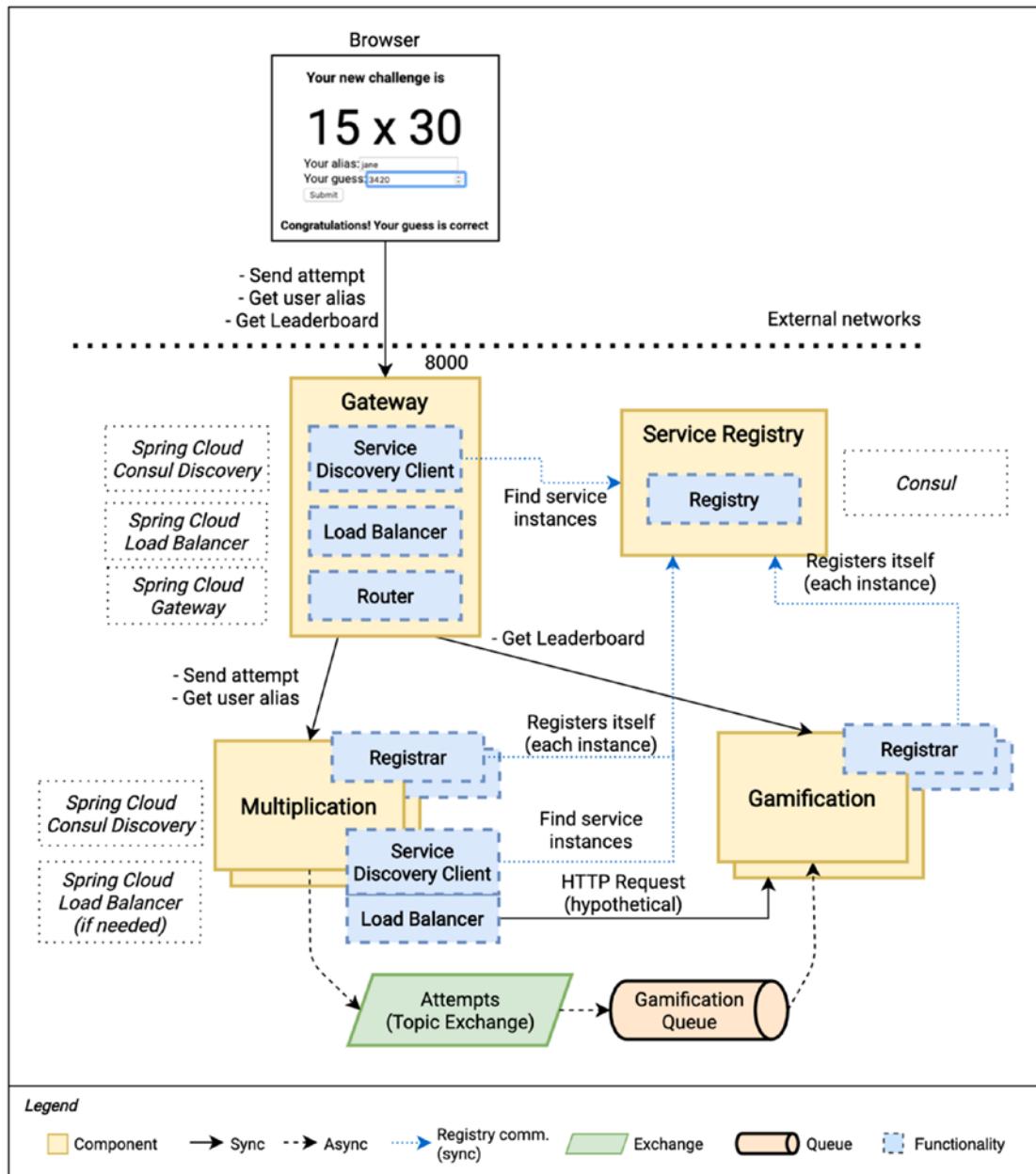
Figure 8-7. Server-side service discovery

Typically, in a microservices architecture, you'll see either both approaches combined or just server-side discovery. Client-side discovery doesn't work well when the API clients are outside our system, because we shouldn't require external clients to interact with a service registry and do load balancing themselves. Normally, the gateway takes this responsibility. Therefore, our API gateway will be connected to the service registry and will include a load balancer to distribute the load across the instances.

For any other service-to-service communication within our back end, we could either connect all of them to the registry for client-side discovery or abstract each cluster of services (all its instances) as a unique address with a load balancer. The latter is the technique chosen in some platforms like Kubernetes, where each service is assigned a unique address no matter how many replicas there are and in which node they're located (we'll come back to this later in this chapter).

Our microservices are no longer calling each other, but if that would be needed, it'd be straightforward to implement client-side discovery. Spring Boot has integrations to connect to the service registry and implement a load balancer (similar to what we'll do in the gateway).

As we mentioned already, any noninternal HTTP communication to our back end will use a server-side discovery approach. That means our gateway will not only route traffic but also take care of load balancing. See Figure 8-8, which also includes the solution we would choose in case of needing interservice communication. This diagram also introduces the name of the tool we'll choose to implement service discovery, Consul, and the Spring Cloud projects we'll add to our dependencies to integrate this tool and include a simple load balancer. We'll learn about them soon.

**Figure 8-8.** Gateway and service discovery integrations

Consul

Many tools implement the service discovery pattern: Consul, Eureka, Zookeeper, etc. There are also complete platforms that include this pattern as one of their features, as we'll describe later.

In the Spring ecosystem, Netflix's Eureka has been the most popular choice for a long time. However, for the reasons stated before (components in maintenance mode, new tools developed by Spring developers), that preference is no longer a sound option. We'll use Consul, a tool that provides service discovery among other features and has also good integration via a Spring Cloud module. Besides, we'll take advantage of one of the other Consul features later in this chapter to implement another pattern in microservice architectures, centralized configuration.

First, let's install the Consul Tools, which is available for multiple platforms at the Downloads page (<https://tpd.io/dlconsul>). Once you have installed it, you can run the Consul Agent in development mode with the command shown in Listing 8-15.

Listing 8-15. Starting the Consul Agent in Development Mode

```
$ consul agent -node=learnmicro -dev
==> Starting Consul agent...
      Version: 'v1.7.3'
      Node ID: '0a31db1f-edee-5b09-3fd2-bcc973867b65'
      Node name: 'learnmicro'
      Datacenter: 'dc1' (Segment: '<all>')
      Server: true (Bootstrap: false)
      Client Addr: [127.0.0.1] (HTTP: 8500, HTTPS: -1, gRPC: 8502, DNS: 8600)
      Cluster Addr: 127.0.0.1 (LAN: 8301, WAN: 8302)
      Encrypt: Gossip: false, TLS-Outgoing: false, TLS-Incoming: false,
      Auto-Encrypt-TLS: false
...
...
```

The logs should display some information about the server and some startup actions. We run the agent in *development mode* since we're using it locally, but a proper production setup of Consul would consist of a cluster with multiple datacenters. These datacenters may run one or more agents, where only one of them per server would act as a server agent. Agents use a protocol to communicate between them to sync information and elect a leader via consensus. All this setup ensures high availability. If a datacenter

becomes unreachable, agents would notice it and elect a new leader. If you want to know more about deploying Consul in production, check the Deployment Guide (<https://tpd.io/consulprod>). We'll stick to the development mode in this book, with a stand-alone agent.

As we see in the output, Consul runs an HTTP server on port 8500. It offers a RESTful API that we can use for service registration and discovery, among other features. Besides, it provides a UI that we can access if we navigate from a browser to `http://localhost:8500`. See Figure 8-9.

The screenshot shows the Consul UI interface. At the top, there is a navigation bar with tabs: Services (which is active and highlighted in dark blue), Nodes, Key/Value, ACL, Intentions, Documentation, and Settings. Below the navigation bar, the main content area has a title 'Services 1 total'. A search bar contains the query 'service:name tag:name status:critical search-term'. The main table lists one service: 'consul'. The table has columns: Service, Health Checks ⓘ, and Tags. The 'consul' entry shows 1 health check (indicated by a green checkmark icon) and no tags. At the bottom of the page, there is a footer with the HashiCorp logo, the text '© 2020 HashiCorp', 'Consul 1.7.3', and 'Documentation'.

Figure 8-9. Consul UI

The Services section displays a list of registered services. Since we didn't do anything yet, the only available service is the consul server. The other tabs show us the available Consul nodes, the key/value functionality we'll use later in this chapter, and some additional Consul features such as ACL and intentions that we won't use in this book.

You can also access the list of available services via the REST API. For example, using HTTPie, we can ask for a list of available services, which will output an empty response body for now. See Listing 8-16.

Listing 8-16. Requesting the List of Services from Consul

```
$ http -b :8500/v1/agent/services
{}
```

The Service API allows us to list services, query their information, know if they're healthy, register them, and deregister them. We won't use this API directly because the Spring Cloud Consul module does this for us, as we'll cover soon.

Consul includes functionality to verify the state of all services: the health checks feature. It offers multiple options that we can use to determine healthiness: HTTP, TCP, Scripts, etc. As you can imagine, our plan is to make Consul contact our microservices via the HTTP interface, more specifically on the /actuator/health endpoints. The health check location is configured at service registration time, and Consul triggers them on a periodic interval basis that can also be customized. If a service fails to respond or it does it with a non-OK status (other than 2XX), Consul will flag that service as unhealthy. We'll see a practical example soon. Read the Checks page (<https://tpd.io/consul-checks>) on the Consul documentation if you want to know more about how to configure them.

Spring Cloud Consul

We don't need to use the Consul API to register services, define health checks, or access the registry to find a service address. All these functionalities are abstracted by the Spring Cloud Consul project, so all we need is to include the corresponding starter in our Spring Boot applications and configure some settings if we choose not to use the default values.

The version of Spring Cloud Consul we'll use still comes with Netflix's Ribbon as an included dependency to implement the load balancer pattern. As we covered earlier, this tool is in maintenance mode, and the Spring documentation discourages its usage (see <https://tpd.io/no-ribbon>). We'll detail the alternative we'll use in the next section. For now, to keep our project clean, we're going to use Maven to exclude the transitive dependency on the Ribbon's starter. See Listing 8-17.

Listing 8-17. Adding the Spring Cloud Consul Discovery Dependency in Maven

```
<dependency>
```

```
  <groupId>org.springframework.cloud</groupId>
  <artifactId>spring-cloud-starter-consul-discovery</artifactId>
```

```

<exclusions>
  <exclusion>
    <groupId>org.springframework.cloud</groupId>
    <artifactId>spring-cloud-starter-netflix-ribbon</artifactId>
  </exclusion>
</exclusions>
</dependency>

```

We'll add this dependency to the Gateway project a bit later. For the other two microservices, we're adding a Spring Cloud dependency for the first time. Therefore, we need to add the dependencyManagement node to our pom.xml files and the Spring Cloud version. See Listing 8-18 for the required additions.

Listing 8-18. Adding Consul Discovery to Multiplication and Gamification

```

<project>
  <!-- ... -->
  <properties>
    <!-- ... -->
    <spring-cloud.version>Hoxton.SR7</spring-cloud.version>
  </properties>

  <dependencies>
    <!-- ... -->
    <dependency>
      <groupId>org.springframework.cloud</groupId>
      <artifactId>spring-cloud-starter-consul-discovery</artifactId>
      <exclusions>
        <exclusion>
          <groupId>org.springframework.cloud</groupId>
          <artifactId>spring-cloud-starter-netflix-ribbon</artifactId>
        </exclusion>
      </exclusions>
    </dependency>
  </dependencies>

```

```

<dependencyManagement>
  <dependencies>
    <dependency>
      <groupId>org.springframework.cloud</groupId>
      <artifactId>spring-cloud-dependencies</artifactId>
      <version>${spring-cloud.version}</version>
      <type>pom</type>
      <scope>import</scope>
    </dependency>
  </dependencies>
</dependencyManagement>
<!-- ... -->
</project>

```

The included Spring Boot autoconfiguration defaults for Consul are fine for us: the server is located at `http://localhost:8500`. See the source code (<https://tpd.io/consulprops>) of `ConsulProperties` if you want to check these default values. If we would need to change them, we could use these and other properties available under the `spring.cloud.consul` prefix. For a complete list of settings that you can override, check the Spring Cloud Consul's reference docs (<https://tpd.io/consulconfig>).

However, there is a new configuration property we need in our applications: the application name, specified by the `spring.application.name` property. We haven't needed it so far, but Spring Cloud Consul uses it to register the service with that value. This is the line we have to add in the `application.properties` file inside the Multiplication project:

```
spring.application.name=multiplication
```

Make sure to add this line to the Gamification microservice's configuration too, this time with the value `gamification`. In the Gateway project, we use YAML properties, but the change is similar.

```
spring:
  application:
    name: gateway
```

Now, let's start the Multiplication and Gamification microservices and see how they register themselves with their corresponding health checks. Remember to start the RabbitMQ server and the Consul agent as well. We still need some changes in the

Gateway service, so we don't need to start it yet. In the application logs, you should see a new line like the one shown in Listing 8-19 (note that it's only one line, but a very verbose one).

Listing 8-19. Gamification's Log Line Showing the Consul Registration Details

```
INFO 53587 --- [main] o.s.c.c.s.ConsulServiceRegistry: Registering service
with consul: NewService{id='gamification-8081', name='gamification',
tags=[secure=false], address='192.168.1.133', meta={}, port=8081,
enableTagOverride=null, check=Check{script='null', dockerContainerID='null',
shell='null', interval='10s', ttl='null', http='http://192.168.1.133:8081/
actuator/health', method='null', header={}, tcp='null', timeout='null', deregis-
terCriticalServiceAfter='null', tlsSkipVerify=null, status='null', grpc='null',
grpcUseTLS=null}, checks=null}
```

This line shows the service registration via Spring Cloud Consul when the application starts. We can see the request's contents: a unique ID composed of the service name and port, the service name that may group multiple instances, the local address, and a configured health check over HTTP to the address of the service's health endpoint exposed by Spring Boot Actuator. The interval in which Consul will verify this check is set to 10 seconds by default.

On the Consul server side (see Listing 8-20), the logs are set to DEBUG level by default in development mode, so we can see how Consul processes these requests and triggers the checks.

Listing 8-20. Consul Agent Logs

```
[DEBUG] agent.http: Request finished: method=PUT url=/v1/agent/service/
register?token=<hidden> from=127.0.0.1:54172 latency=2.424765ms
[DEBUG] agent: Node info in sync
[DEBUG] agent: Service in sync: service=gamification-8081
[DEBUG] agent: Check in sync: check=service:gamification-8081
```

Once we start both microservices with this new configuration, we can access the Consul's UI to see the updated status. See Figure 8-10.

Services 3 total

Service	Health Checks	Tags
consul	✓ 1	
gamification	✓ 2	secure=false
multiplication	✓ 2	secure=false

© 2020 HashiCorp Consul 1.7.3 Documentation

Figure 8-10. Services listed in Consul

Now navigate to Services, click “multiplication,” and click the only “multiplication” row displayed there. You’ll see the service’s health check. We can verify how Consul is getting an OK status (200) from the Spring Boot application. See Figure 8-11.

All Services < Service (multiplication) < Instance

multiplication

Service Name	Node Name
multiplication	learnmicro

Service Checks Node Checks Tags Meta Data

Service 'multiplication' check

ServiceName	CheckID	Type	Notes
multiplication	service:multiplication	http	-

Output

```
HTTP GET http://192.168.1.133:8080/actuator/health: 200 Output: {"status":"UP"} 
```

Figure 8-11. Service health check

We can also start a second instance of one of the microservices to see how this is managed by the registry. You can do that from your IDE if you override the port or directly from the command line. See Listing 8-21 for an example of how to start a second instance of the Multiplication microservice. As you can see, we can override the server port using the Spring Boot's Maven plugin (see <https://tpd.io/mvn-sb-props> for more details).

Listing 8-21. Running a Second Instance of the Multiplication Microservice from the Command Line

```
multiplication $ ./mvnw spring-boot:run -Dspring-boot.run.arguments="--server.port=9080"
[... logs ...]
```

In the Consul registry, there will be still a single multiplication service. If we click this service, we'll navigate to the instances tab. See Figure 8-12. There, we can see both instances, each of them with their corresponding health check. Note that the port is not used in the ID by Spring Boot when it's the default value: 8080.

The screenshot shows the Consul UI interface. At the top, there is a navigation bar with tabs: Services (selected), Nodes, Key/Value, ACL, Intentions, Documentation, and Settings. Below the navigation bar, there is a breadcrumb trail: < All Services. The main area displays a service named "multiplication". Under the "multiplication" service, there is a tab bar with Instances (selected), Routing, and Tags. To the right of the tab bar is a search bar with a magnifying glass icon. Below the search bar is a table showing two instances of the service:

ID	Node	Address	Node Checks	Service Checks
multiplication	learnmicro	192.168.1.133:8080	✓ 1	✓ 1
multiplication-9080	learnmicro	192.168.1.133:9080	✓ 1	✓ 1

At the bottom of the page, there is a footer with the HashiCorp logo, copyright information (© 2020 HashiCorp), Consul version (1.7.3), and links to Documentation and Settings.

Figure 8-12. Multiple instances in the Consul's registry

An API request to get the list of services from Consul retrieves now these two services, including information about both instances of the Multiplication application. See Listing 8-22 for a shortened version of the response.

Listing 8-22. Retrieving Registered Services Using the Consul API

```
$ http -b :8500/v1/agent/services
{
  "gamification-8081": {
    "Address": "192.168.1.133",
    "EnableTagOverride": false,
    "ID": "gamification-8081",
    "Meta": {},
    "Port": 8081,
    "Service": "gamification",
    ...
    "Weights": {
      "Passing": 1,
      "Warning": 1
    }
  },
  "multiplication": {
    "Address": "192.168.1.133",
    "EnableTagOverride": false,
    "ID": "multiplication",
    "Meta": {},
    "Port": 8080,
    "Service": "multiplication",
    ...
    "Weights": {
      "Passing": 1,
      "Warning": 1
    }
  },
  "multiplication-9080": {
    "Address": "192.168.1.133",
    "EnableTagOverride": false,
    "ID": "multiplication-9080",
```

```

    "Meta": {},
    "Port": 9080,
    "Service": "multiplication",
    ...
    "Weights": {
        "Passing": 1,
        "Warning": 1
    }
}
}
}

```

You can surely picture now how we would work with Consul as a client service if we wouldn't have the Spring abstraction. First, all services would need to know the HTTP host and port to reach the registry. Then, if the service wants to interact with the Gamification API, it would use Consul's Service API to get the list of available instances. The API also has an endpoint to retrieve information about the current health status for a given service identifier. Following a client-side discovery approach, the service would apply load balancing (e.g., round robin) and pick one healthy instance from the list. Then, knowing the address and the port to target the request, the client service can perform the request. We don't need to implement this logic since Spring Cloud Consul does that for us, including load balancing as we'll cover in the next section.

Given that Gateway is the only service in our system that is calling others, that's where we'll put into practice the Consul's service discovery logic. However, before doing that, we need to introduce the pattern we're still missing: the load balancer.

Spring Cloud Load Balancer

We'll implement a Client-side discovery approach where the back-end services query the registry and decide which instance to call if there is more than one available. This last part is a logic we could build ourselves, but it's even easier to rely on tools that do that for us. The Spring Cloud load balancer project is a component of Spring Cloud Commons that integrates with the service discovery integrations (both Consul and Eureka) to provide a simple load balancer implementation. By default, it autoconfigures a round-robin load balancer that goes through all instances iteratively.

As we mentioned earlier, Netflix's Ribbon used to be the preferred choice to implement the load balancer pattern. Since it's in maintenance mode, let's discard that option and choose the Spring's load balancer implementation. Both Ribbon and the Spring Cloud load balancer are included as dependencies within the Spring Cloud Consul starter, but we can switch between the two using configuration flags or explicitly excluding one of the dependencies (like we did when adding the Consul starter).

To make load-balanced calls between two applications, we can simply use the `@LoadBalanced` annotation when creating a `RestTemplate` object. Then, we use the service name as the hostname in our URLs when performing requests to that service. The Spring Cloud Consul and load balancer components will do the rest, querying the registry and selecting the next instance in order.

We used to have a call from the Multiplication service to the Gamification service before we moved to an event-driven approach, so let's use that one as an example. Listing 8-23 shows how we could have integrated service discovery and load balancing in the client, the Multiplication microservice. This was also illustrated in Figure 8-8. As you can see, we only need to declare a `RestTemplate` bean configured with the `@LoadBalanced` annotation and use the URL `http://gamification/attempts`. Note that you don't need to specify the port number, because it'll be included in the resolved instance URL after contacting the registry.

Listing 8-23. Example of How to Use a `RestTemplate` with Load Balancing Capabilities

```
@Configuration
public class RestConfiguration {

    @LoadBalanced
    @Bean
    RestTemplate restTemplate() {
        return new RestTemplate();
    }
}

@Slf4j
@Service
public class GamificationServiceClient {
```

```

private final RestTemplate restTemplate;

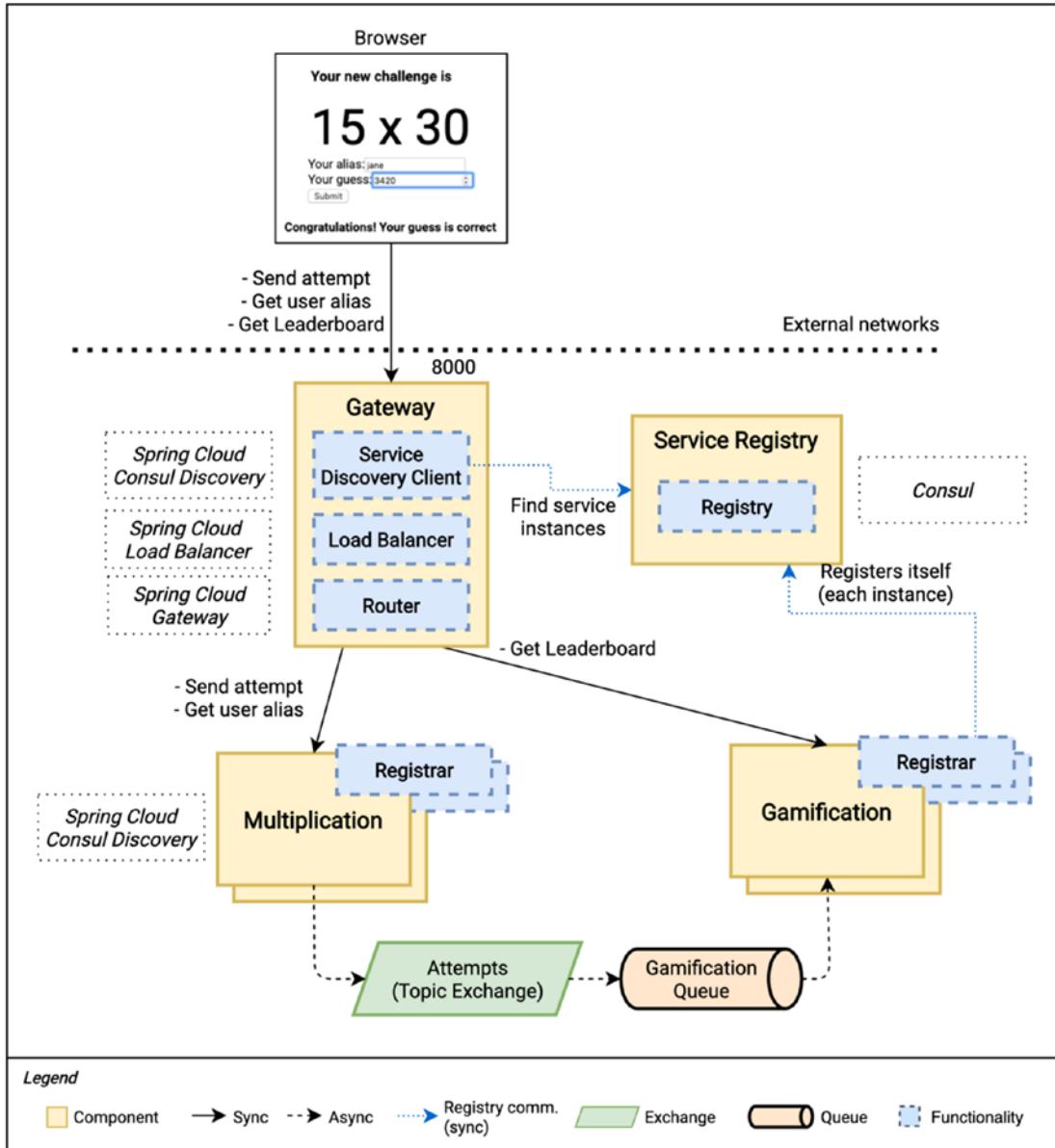
public GamificationServiceClient(final RestTemplate restTemplate) {
    this.restTemplate = restTemplate;
}

public boolean sendAttempt(final ChallengeAttempt attempt) {
    try {
        ChallengeSolvedDTO dto = new ChallengeSolvedDTO(attempt.getId(),
            attempt.isCorrect(), attempt.getFactorA(),
            attempt.getFactorB(), attempt.getUser().getId(),
            attempt.getUser().getAlias());
        ResponseEntity<String> r = restTemplate.postForEntity(
            "http://gamification/attempts", dto,
            String.class);
        log.info("Gamification service response: {}", r.getStatusCode());
        return r.getStatusCode().is2xxSuccessful();
    } catch (Exception e) {
        log.error("There was a problem sending the attempt.", e);
        return false;
    }
}
}

```

We won't follow this path since we already got rid of HTTP calls between our microservices, but this is a good approach for those scenarios where you need to have interservice HTTP interactions. With service discovery and load balancing, you reduce the risk of failure since you increase the chances that there is at least one instance available to handle these synchronous requests.

Our plan is to integrate service discovery and load balancing in the gateway. See Figure 8-13.

**Figure 8-13.** Gateway, service discovery, and load balancing in our system

Service Discovery and Load Balancing in the Gateway

After including the Spring Cloud Consul starter in our applications, they're contacting the registry to publish their information. However, we still have the gateway using explicit address/port combinations to proxy requests. It's time to integrate service discovery and load balancing there.

First, we add the Spring Cloud Consul dependency to the Gateway project. See Listing 8-24. Again, we exclude Ribbon because we'll use Spring Cloud load balancer.

Listing 8-24. Adding Spring Cloud Consul Discovery to the Gateway

```
<dependencies>
    <!-- ... -->
    <dependency>
        <groupId>org.springframework.cloud</groupId>
        <artifactId>spring-cloud-starter-consul-discovery</artifactId>
        <exclusions>
            <exclusion>
                <groupId>org.springframework.cloud</groupId>
                <artifactId>spring-cloud-starter-netflix-ribbon</artifactId>
            </exclusion>
        </exclusions>
    </dependency>
</dependencies>
```

All we need to do to take advantage of these new patterns is to add some configuration to the application.yml file. We can divide the changes into three groups.

- *Global settings:* We give a name to the application and make sure we use the Spring Cloud load balancer implementation. Besides, we'll add a configuration parameter to instruct the service discovery client to retrieve only the healthy services.
- *Routing configuration:* Instead of using explicit host and ports, we switch to service names with a URL pattern that also enables load balancing.
- *Resilience:* In case the gateway fails to proxy a request to a service, we want it to retry a few times. We'll elaborate on this topic.

See Listing 8-25 for the complete source code of our new Gateway configuration (`application.yml`) including these changes.

Listing 8-25. Gateway Configuration Including Load Balancing

```

server:
  port: 8000

spring:
  application:
    name: gateway
  cloud:
    loadbalancer:
      ribbon:
        # Not needed since we excluded the dependency, but
        # still good to add it here for better readability
        enabled: false
    consul:
      enabled: true
      discovery:
        # Get only services that are passing the health check
        query-passing: true
    gateway:
      routes:
        - id: multiplication
          uri: lb://multiplication/
          predicates:
            - Path=/challenges/**,/attempts,/attempts/**,/users/**
        - id: gamification
          uri: lb://gamification/
          predicates:
            - Path=/leaders
    globalcors:
      cors-configurations:
        '[/*]':
          allowedOrigins: "http://localhost:3000"
          allowedHeaders:
            - "*"

```

```

allowedMethods:
  - "GET"
  - "POST"
  - "OPTIONS"

default-filters:
  - name: Retry
    args:
      retries: 3
      methods: GET,POST

```

With the `query-passing` parameter set to true, the Spring implementation will use the Consul API with a filter to retrieve only those services that have a passing health check. We only want to proxy requests to healthy instances. A value `false` could make sense in cases where the service doesn't poll very often for the updated service list. In that case, it's good to get the complete list because we don't know their latest status and have mechanisms to deal with unhealthy instances (for example retries, as we'll learn soon).

The most relevant changes are those applied to the URLs. As you can see, now we use URLs like `lb://multiplication/`. Since we added the Consul client, the application will use the Service API to resolve the service name, `multiplication`, to the available instances. The special scheme `lb` tells Spring that should use the load balancer.

In addition to the basic configuration, we added a gateway filter that applies to all the requests because it's under the `default-filters` node: the `retry` `GatewayFilter` (see <https://tpd.io/gwretry> for details). This filter intercepts error responses and transparently retries the request again. When combined with a load balancer, this means the request will be proxied to the next instance, so we get a nice resilience pattern (retry) easily. We configure this filter to make three retries maximum for the HTTP methods we're using, which is more than enough to cover most failure situations. In case all the retries would fail, the gateway returns an error response to the client (service unavailable) because it can't proxy the request.

You might be wondering why it's necessary to include retries in the service discovery client, despite that we configured it to get only healthy instances. In theory, if they're all healthy, all calls should succeed. To understand this, we have to review how Consul (and typically any other service discovery tool) works. Each service registers itself with a configured health check to be polled every ten seconds (the default value, but we could also change it). The registry doesn't know in real time when services are not ready

to handle the traffic. It could be the case that Consul successfully checks the health of a given instance and that one goes down immediately after. The registry will list this instance as healthy for a few seconds (almost ten with our configuration) until it notices it's not available during the next check. Since we want to minimize request errors during that interval too, we can take advantage of the retry pattern to cover these situations. Once the registry gets updated, the Gateway won't get the unhealthy instance within the service list, so the retries won't be necessary anymore. Note that lowering the time between checks can reduce the number of errors, but it increases the network traffic.

Circuit Breakers

There might be cases where you don't want to keep trying future requests to a given service after you know it's failing. By doing that, you can save time wasted in response timeouts and alleviate potential congestion of the target service. This is especially useful for external service calls when there are no other resilience mechanisms in place like the service registry with health checks.

For these scenarios, you can use a circuit breaker. The circuit is *closed* when everything works fine. After a configurable number of request failures, the circuit becomes *open*. Then, the requests are not even tried, and the circuit breaker implementation returns a predefined response. Now and then, the circuit may switch to *half-open* to check again if the target service is working. In that case, the circuit will transition to *close*. If it's still failing, it goes back to the *open* state. Check <https://tpd.io/cbreak> for more information about this pattern.

After applying the new configuration, the Gateway microservice connects to Consul to find the available instances of other microservices and their network locations. Then, it balances load based on a simple round-robin algorithm included in Spring Cloud load balancer. Check again Figure 8-8 for a complete overview.

Given that we added the Consul starter, the Gateway service is also registering itself in Consul. That is not strictly necessary since other services won't call the gateway, but it's still useful for us to check its status. Alternatively, we could set the configuration parameter `spring.cloud.consul.discovery.register` to `false` to keep using the service discovery client features but disable the registration of the Gateway service.

In our setup, all the external HTTP traffic (not between microservices) go through the Gateway microservice via `localhost:8000`. In a production environment, we would typically expose this HTTP interface on port 80 (or 443 if we use HTTPS) and use a DNS address (e.g., `bookgame.tpd.io`) to point to the IP where our server lives. Nevertheless, there would be a single entry point for public access, and that makes this service a critical part of our system. It must be as highly available as we can. If the Gateway service goes down, our entire system goes down.

To reduce the risk, we could introduce *DNS load balancing* (a hostname that points to multiple IP addresses) to add redundancy to the gateway. However, it relies on the client (e.g., a browser) to manage the list of IP addresses and handle failover when one of the hosts doesn't respond (see <https://tpd.io/dnslbq> for an explanation). We could see this as an extra layer on top of the gateway, which adds client-side discovery (DNS resolution to a list of IP addresses), load balancing (choose an IP address from the list), and fault tolerance (try another IP after a timeout or error). This is not a typical approach.

Cloud providers such as Amazon, Microsoft, or Google offer the routing and load balancing patterns as managed services with high availability guarantees, so that's also an alternative to making sure the gateway remains operational at all times. Kubernetes, on the other hand, allows you to create a load balancer on top of your own gateway, so you can add redundancy to that layer too. We'll see more about platform implementations at the end of this chapter.

Playing with Service Discovery and Load Balancing

Let's put into practice the service discovery and load balancing features.

Before running our applications, we'll add a log line to `UserController` (`Multiplication`) and `LeaderBoardController` (`Gamification`) to quickly see in the logs the interactions with their APIs. See Listings 8-26 and 8-27.

Listing 8-26. Adding a Log Line to `UserController`

```
@Slf4j
@RequiredArgsConstructor
@RestController
@RequestMapping("/users")
public class UserController {
```

```

private final UserRepository userRepository;

@GetMapping("/{idList}")
public List<User> getUsersByIdList(@PathVariable final List<Long> idList) {
    log.info("Resolving aliases for users {}", idList);
    return userRepository.findAllByIdIn(idList);
}
}

```

Listing 8-27. Adding a Log Line to LeaderBoardController

```

@Slf4j
@RestController
@RequestMapping("/leaders")
@RequiredArgsConstructor
class LeaderBoardController {

    private final LeaderBoardService leaderBoardService;

    @GetMapping
    public List<LeaderBoardRow> getLeaderBoard() {
        log.info("Retrieving leaderboard");
        return leaderBoardService.getCurrentLeaderBoard();
    }
}

```

Now, let's run our complete system. The required steps are the same as before plus the new command to run the service registry:

1. Run the RabbitMQ server.
2. Run the Consul agent in development mode.
3. Start the Multiplication microservice.
4. Start the Gamification microservice.
5. Start the new Gateway microservice.
6. Run the front-end app.

Once we run the minimal setup, we add one extra instance of each of our services running business logic: Multiplication and Gamification. Remember that you need to override the `server.port` property. From a terminal, you could use the commands shown in Listing 8-28 in two separate tabs or windows (note that the folder where you run each command is different).

Listing 8-28. Running Two Additional Instances of Multiplication and Gamification

```
multiplication $ ./mvnw spring-boot:run -Dspring-boot.run.arguments="--server.  
port=9080"  
[... logs ...]  
gamification $ ./mvnw spring-boot:run -Dspring-boot.run.arguments="--server.  
port=9081"  
[... logs ...]
```

All the instances will publish their details in the registry (Consul). Besides, all of them can act as registry clients to retrieve the details of the different instances of a given service name and where they are located. In our system, this is done only from the gateway. See an overview of the services in the Consul UI (section Nodes/Services) after booting up the two extra instances in Figure 8-14.

The screenshot shows the Consul UI interface. At the top, there's a navigation bar with tabs: dc1, Services, Nodes (which is selected), Key/Value, ACL, Intentions, Documentation, and Settings. Below the navigation bar, there's a link to 'All Nodes'. The main title is 'learnmicro' with a '127.0.0.1' icon. Underneath, there are four tabs: Health Checks, Services (selected), Lock Sessions, and Meta Data. A search bar at the top right says 'Search by name/port' with a magnifying glass icon. The main content area displays a table of service instances:

Service	Port	Tags
consul	8300	
gamification (gamification-8081)	8081	secure=false
gamification (gamification-9081)	9081	secure=false
gateway (gateway-8000)	8000	secure=false
multiplication	8080	secure=false
multiplication (multiplication-9080)	9080	secure=false

At the bottom of the page, there's a footer with the HashiCorp logo, the text '© 2020 HashiCorp', 'Consul 1.7.3', and 'Documentation'.

Figure 8-14. Consul: multiple instances

Verifying that the gateway’s load balancer works is simple: check the logs of both Gamification service instances. With the newly added log lines, you can quickly see how both are getting alternating requests from the UI, related to leaderboard updates. If you refresh the browser’s page a few times to force new challenge requests, you’ll see similar behavior. Listings 8-29, 8-30, 8-31, and 8-32 show extracts for the same time window of logs for both Multiplication instances and both Gamification instances. As you can see, requests alternate between the available instances every five seconds.

Listing 8-29. Logs for Multiplication, First Instance (Port 8080)

```
2020-08-29 09:05:06.957 INFO 9999 --- [nio-8080-exec-6] m.b.multiplication.user.  
UserController : Resolving aliases for users [125, 49, 72, 60, 101, 1, 96, 107,  
3, 45, 6, 9, 14, 123]
```

```

2020-08-29 09:05:09.090 INFO 9999 --- [nio-8080-exec-7] m.b.multiplication.user.
UserController : Resolving aliases for users [125, 49, 72, 60, 101, 1, 96, 107,
                 3, 45, 6, 9, 14, 123]
2020-08-29 09:05:19.033 INFO 9999 --- [nio-8080-exec-9] m.b.multiplication.user.
UserController : Resolving aliases for users [125, 49, 72, 60, 101, 1, 96, 107,
                 3, 45, 6, 9, 14, 123]

```

Listing 8-30. Logs for Multiplication, Second Instance (Port 9080)

```

2020-08-29 09:05:09.009 INFO 10138 --- [nio-9080-exec-7] m.b.m.challenge.
ChallengeController : Generating a random challenge: Challenge(factorA=58,
                     factorB=96)
2020-08-29 09:05:14.040 INFO 10138 --- [nio-9080-exec-8] m.b.multiplication.user.
UserController : Resolving aliases for users [125, 49, 72, 60, 101, 1, 96, 107,
                 3, 45, 6, 9, 14, 123]
2020-08-29 09:05:24.042 INFO 10138 --- [io-9080-exec-10] m.b.multiplication.user.
UserController : Resolving aliases for users [125, 49, 72, 60, 101, 1, 96, 107,
                 3, 45, 6, 9, 14, 123]

```

Listing 8-31. Logs for Gamification, First Instance (Port 8081)

```

2020-08-29 09:05:03.208 INFO 9928 --- [nio-8081-exec-6] m.b.g.game.
LeaderBoardController : Retrieving leaderboard
2020-08-29 09:05:09.006 INFO 9928 --- [nio-8081-exec-8] m.b.g.game.
LeaderBoardController : Retrieving leaderboard
2020-08-29 09:05:19.014 INFO 9928 --- [io-8081-exec-10] m.b.g.game.
LeaderBoardController : Retrieving leaderboard

```

Listing 8-32. Logs for Gamification, Second Instance (Port 9081)

```

2020-08-29 09:04:58.107 INFO 10222 --- [nio-9081-exec-4] m.b.g.game.
LeaderBoardController : Retrieving leaderboard
2020-08-29 09:05:06.927 INFO 10222 --- [nio-9081-exec-6] m.b.g.game.
LeaderBoardController : Retrieving leaderboard
2020-08-29 09:05:14.010 INFO 10222 --- [nio-9081-exec-8] m.b.g.game.
LeaderBoardController : Retrieving leaderboard

```

That's a great achievement: we scaled up our system, and everything works as expected. The HTTP traffic is now balanced across all instances equally, in a similar way to how our RabbitMQ setup is distributing messages across consumers. We just smoothly doubled the capacity of our system.

Actually, we could start as many instances as we want of both microservices, and the load would be transparently distributed across all of them. Besides, with the gateway, we made our API clients unaware of our internal services, and we could easily implement there cross-cutting concerns such as user authentication or monitoring.

We should also check whether we're achieving the other nonfunctional requirements we aimed for: resilience, high availability, and fault tolerance. Let's start creating some chaos.

To check what happens when services become unexpectedly unavailable, we could stop them via our IDE or with a Ctrl-C signal in the terminal. However, that doesn't cover all potential incidents we could have in real life. When we do that, the Spring Boot application stops gracefully, so it has the opportunity to deregister itself from Consul. We want to simulate a *major incident* like a network issue or a service abruptly terminated. The best option we have to mimic that is to kill the Java process for a given instance. To know which process to kill, we can examine the logs. The default Spring Boot's Logback configuration prints the process ID after the log level (e.g., INFO) in every log line. As an example, this is the line that indicates we're running the Gamification microservice with a process ID of 97817:

```
2020-07-19 09:10:27.279  INFO 97817 --- [main] m.b.m.GamificationApplication      :  
Started GamificationApplication in 5.371 seconds (JVM running for 11.054)
```

In a Linux or Mac system, you can kill the process with the `kill` command, passing the argument `-9` to force an immediate termination.

```
$ kill -9 97817
```

If you're running Windows, you can use the `taskkill` command with the `/F` flag to force its termination.

```
> taskkill /PID 97817 /F
```

Now that you know how to create disruption, kill one of the Gamification microservice's instance processes. Make sure you have the UI opened in a browser so it keeps making requests to the back end. What you'll see is how after you kill one of the instances, the other one receives all requests and responds to them successfully. Users

don't even notice this. The leaderboard, which calls the API in the Gamification service, remains working. Users can also send new challenges; all attempts will end up in the only instance available. What happens here is that the retry filter in the gateway executes a second request transparently, which gets routed to the healthy instance thanks to the load balancer. See Figure 8-15.

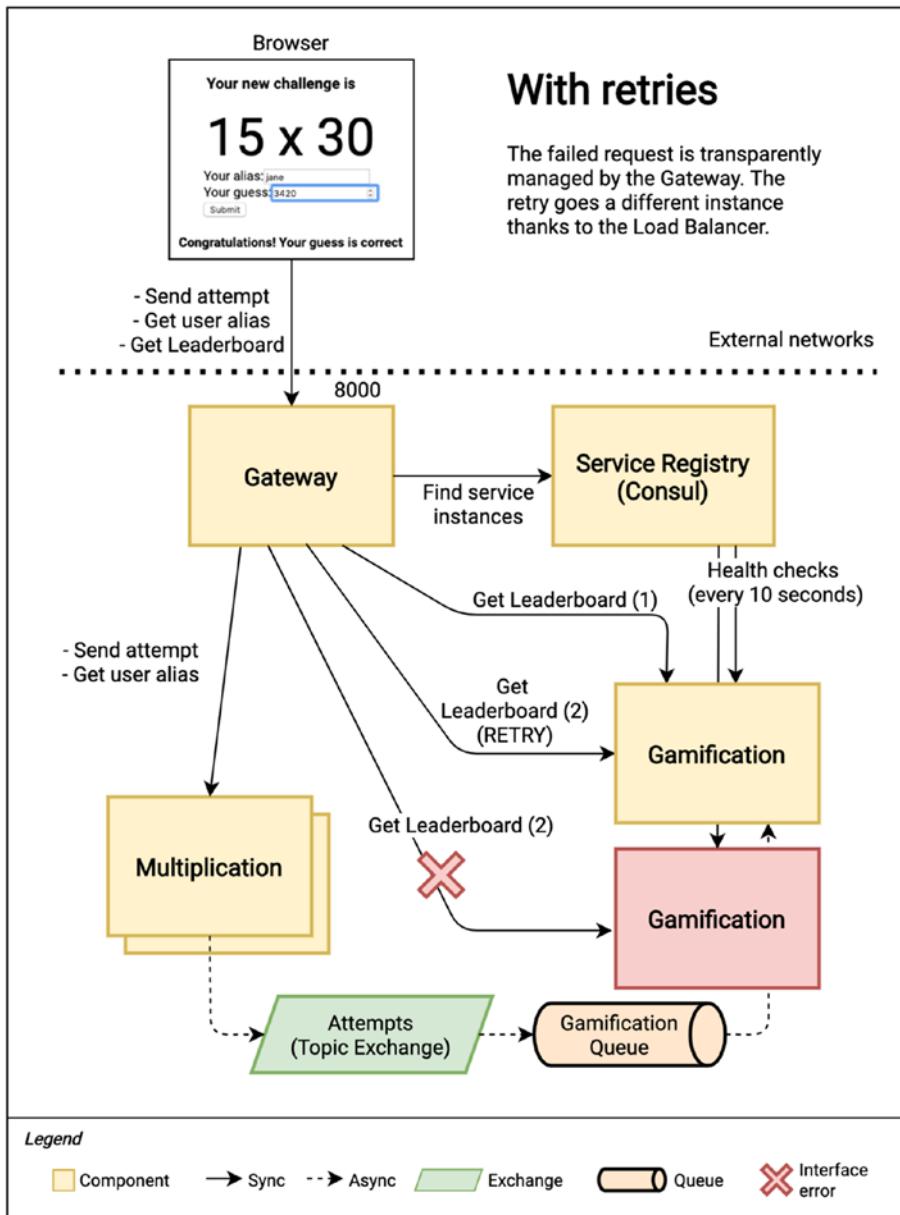


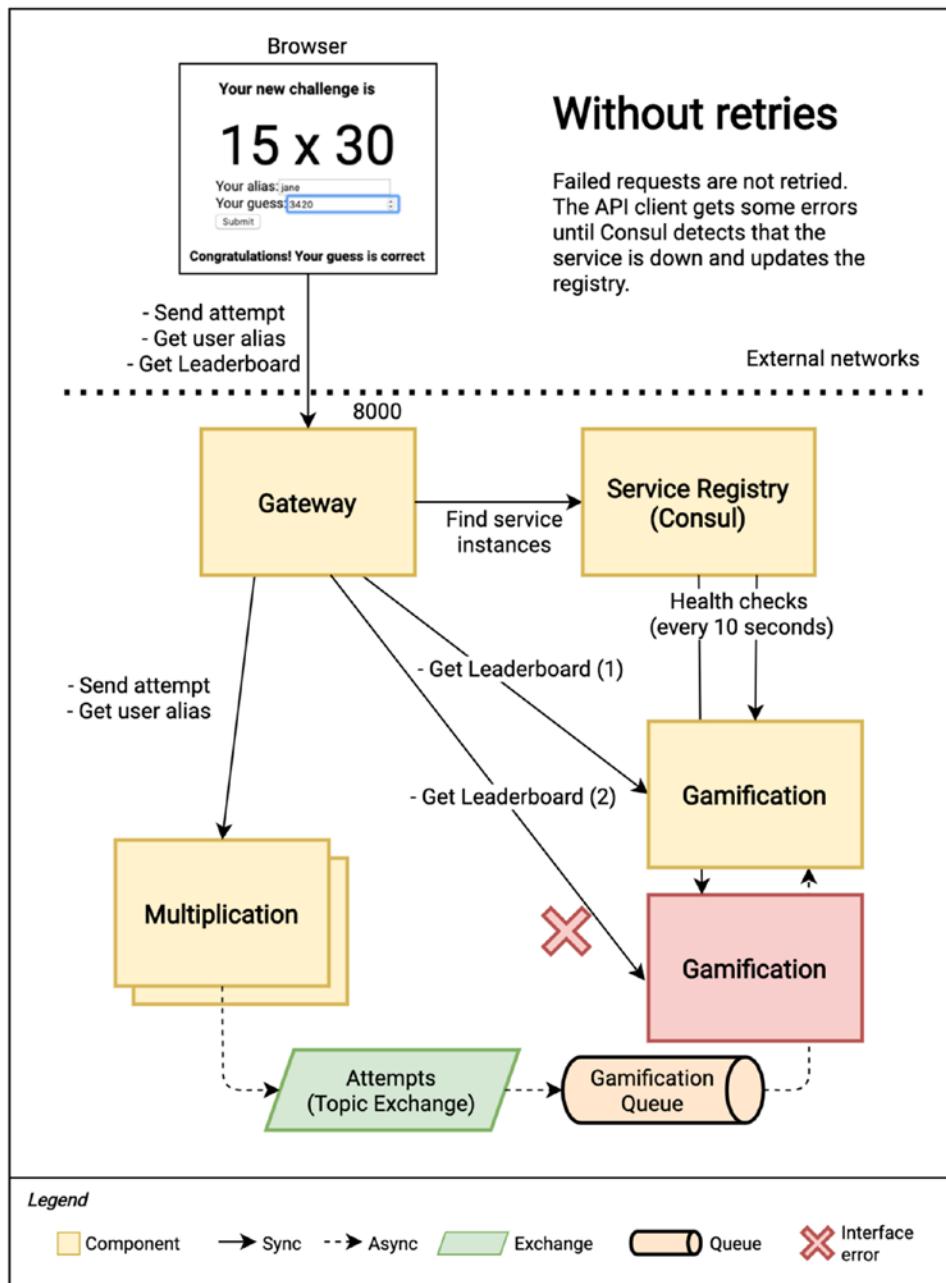
Figure 8-15. Resilience: retry pattern

We also want to verify how the patterns we introduced collaborate to achieve this successful result. To do that, let's remove temporarily the retry filter configuration in the gateway. See Listing 8-33.

Listing 8-33. Commenting a Block of Configuration in the Gateway

```
# We can comment this block of the configuration
# default-filters:
#   - name: Retry
#     args:
#       retries: 3
#       methods: GET,POST
```

Then, rebuild and restart the Gateway service (to apply the new configuration) and repeat a similar scenario. Make sure you boot up a second instance of Gamification again and give some time for the Gateway service to start routing traffic to it. Then, kill one of its instances while you look at the UI. What you'll see this time is that every other request fails to complete, causing an alternating error while displaying the leaderboard. This happens because Consul needs some time (the configured health check interval) to detect that the service is down. In the meantime, the gateway still gets both instances as healthy and proxies some requests to a dead server. See Figure 8-16. The retry mechanism we just removed handled this error transparently and made a second request to the next instance in the list, the one still working.

**Figure 8-16.** Resilience: no retry pattern

It Works on My Machine

Note that if you accidentally kill the process right before the health check, Consul will notice immediately the problem. In that case, you may not see the error in the UI. You can try again the same scenario, or you can configure a longer health check interval in Spring Cloud Consul (via application properties), so you have a higher chance to reproduce this error scenario.

The Consul registry UI also reflects the failing health checks if we navigate to Services. See Figure 8-17.

Service	Health Checks	Tags
consul	✓ 1	
gamification	✓ 3 ✗ 1	secure=false
gateway	✓ 2	secure=false
multiplication	✓ 4	secure=false

Figure 8-17. Consul UI: health check fails after killing the service

We completed an important milestone in our learning path: we implemented scalability in a microservice architecture. Besides, we achieved proper fault tolerance via a load balancer that uses a service discovery registry, which is aware of the health of the different components in our system. I hope that the practical approach helped you understand all these key concepts.

Configuration per Environment

As introduced in the second chapter, one of the main advantages of Spring Boot is the ability to configure profiles. A profile is a set of configuration properties that you can enable depending on your needs. For example, you could switch between connecting to a local RabbitMQ server while testing locally and the real RabbitMQ server running on production when you deploy it to that environment.

To introduce a new `rabbitprod` profile, we can create a file named `application-rabbitprod.properties`. Spring Boot uses the `application-{profile}` naming convention (for both properties and YAML formats) to allow us to define profiles in separate files. See Listing 8-34 for some example properties we could include. If we use this profile for the production environment, we may want to use different credentials, a cluster of nodes to connect to, a secure interface, etc.

Listing 8-34. Example of a Separate Properties File to Override Default Values in Production

```
spring.rabbitmq.addresses=rabbitserver1.tpd.network:5672,rabbitserver2.tpd.
network:5672
spring.rabbitmq.connection-timeout=20s
spring.rabbitmq.ssl.enabled=true
spring.rabbitmq.username=produser1
```

We have to make sure we enable this profile when we start the application in the target environment. To do that, we use the property `spring.profiles.active`. Spring Boot aggregates the base configuration (in `application.properties`) with the values in this file. In our case, all extra properties will be added to the resulting configuration. We can use a Spring Boot's Maven plugin command to enable this new profile for the multiplication microservice:

```
multiplication $ ./mvnw spring-boot:run -Dspring-boot.run.arguments="--spring.
profiles.active=rabbitprod"
```

As you can imagine, all our microservices could have a lot of common configuration values per environment. Not only the connection details for RabbitMQ are probably the same, but also some extra values we added, like the exchange name (`amqp.exchange.attempts`). The same applies to the common configuration for databases or in general to any other Spring Boot configuration that we want to apply to all our microservices.

We could keep these values in separate files per microservice, per environment, and per tool. For example, these four files could include different configurations for RabbitMQ and the H2 database in the staging and production environments.

- application-rabbitprod.properties
- application-databaseprod.properties
- application-rabbitstaging.properties
- application-databasestaging.properties

Then, we could copy them across microservices wherever they're needed. By grouping the configuration into separate profiles, we can reuse these values easily.

However, keeping all these copies still involves a lot of maintenance. In case we want to change one of the values in these common configuration blocks, we have to replace the corresponding file in each project's folder.

A better approach is to put this configuration in a common place in our system and make the applications sync its contents before they start. Then, we keep a centralized configuration per environment, so we need to adjust the values only once. See Figure 8-18. The good news is that this is a well-known pattern, known as *externalized* (or centralized) *configuration*, so there are out-of-the-box solutions to build a centralized configuration server.

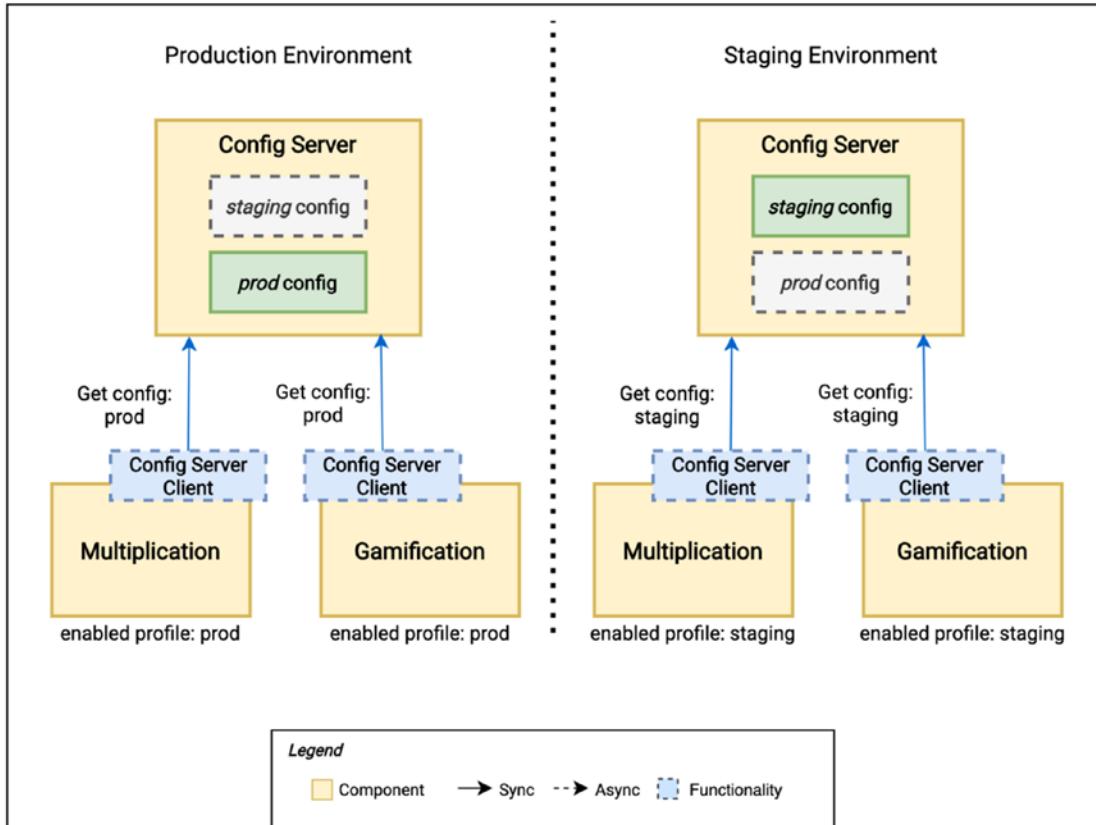


Figure 8-18. Centralized configuration: overview

The first solution that comes out from a simple web search when looking for a configuration server pattern for Spring is the Spring Cloud Config Server project. This is a native implementation included in the Spring Cloud family, which allows you to keep a set of configuration files distributed in folders and exposed via a REST API. On the client side, the projects using this dependency access the config server and request the corresponding configuration resources, depending on their active profiles. The only drawback of this solution for our system is that we need to create another microservice to act as the configuration server and expose the centralized files.

An alternative is to use Consul KV, a feature included in the default Consul package that we didn't explore yet. Spring Cloud also has an integration with this tool to implement a centralized configuration server. We'll choose this approach to reuse components and keep our system as simple as possible, with Consul combining service discovery, health checks, and centralized configuration.

Configuration in Consul

Consul KV is a key/value store installed with the Consul agent. Like with the service discovery feature, we can access this functionality via a REST API and a user interface. When we set up Consul as a cluster, this feature also benefits from replication, so there is less risk of a data loss or downtime due to services being unable to get their configuration.

This simple functionality can be also accessed from the browser since it's included in the consul agent we have already installed. With the agent running, navigate to `http://localhost:8500/ui/dc1/kv` (the Key/Value tab). Now, click Create. You'll see the editor to create a new key/value pair, as shown in Figure 8-19.

The screenshot shows the Consul UI Key/Value editor. At the top, there is a navigation bar with tabs: dc1, Services, Nodes, Key/Value (which is highlighted in blue), ACL, and Intentions. To the right of the tabs are links for Documentation and Settings. Below the navigation bar, there is a breadcrumb trail with a back arrow and the text 'Key / Values'. The main title is 'New Key / Value'. There are two input fields: 'Key or folder' and 'Value'. The 'Key or folder' field contains the value '1'. The 'Value' field is currently empty. To the right of the 'Value' field is a 'Code' toggle switch, which is turned off. At the bottom right of the editor area, there is a 'JSON' dropdown menu. At the very bottom of the editor area is a 'Save' button.

Figure 8-19. Consul: creating a key/value pair

We can use the toggle to switch between the code and the plain editor. The code editor has support for syntax coloring in a few notations, including YAML. Note that, as shown in Figure 8-19 under the text field, we can also create folders if we add a forward slash character at the end of the key name. We'll put this into practice soon.

The Consul KV REST API also allows us to create key/value pairs and folders via HTTP calls and retrieve them using their key names. Check <http://tpd.io/kv-api> if you're curious about how it works. Like with the service discovery feature, we won't need to interact with this API directly since we'll use a Spring abstraction that communicates with Consul KV: Spring Cloud Consul Config.

Spring Cloud Consul Config

The Spring Cloud project that implements centralized configuration with Consul KV is Spring Cloud Consul Config. To use this module, we need to add a new Spring Cloud dependency to our projects: `spring-cloud-starter-consul-config`. This artifact includes autoconfiguration classes that will try to find the Consul agent and read the corresponding KV values at an early stage while booting up our application, the special "bootstrap" phase. It uses this phase because we want Spring Boot to apply the centralized configuration values for the rest of its initialization (e.g., to connect to RabbitMQ).

Spring Cloud Consul Config expects each profile to map to a given key in the KV store. Its value should be a set of Spring Boot configuration values in either YAML or plain format (`.properties`).

We can configure a few settings that help our application find the corresponding keys in the server. These are the most relevant ones:

- *Prefix*: This is the root folder in Consul KV where all profiles are stored. The default value is `config`.
- *Format*: This specifies if the value (the Spring Boot configuration) is in YAML or properties syntax.
- *Default context*: This is the name of the folder used by all applications as common properties.
- *Profile separator*: Keys may combine multiple profiles. In that case, you can specify the character you want to use as a separator (e.g., with a comma `prod,extra-logging`).
- *Data key*: This is the name of the key which holds the properties or YAML content.

All the configuration values related to the setup of the config server must be placed in a separate file for each of our applications, with name `bootstrap.yml` or `bootstrap.properties` (depending on the format we choose). See Figure 8-20.

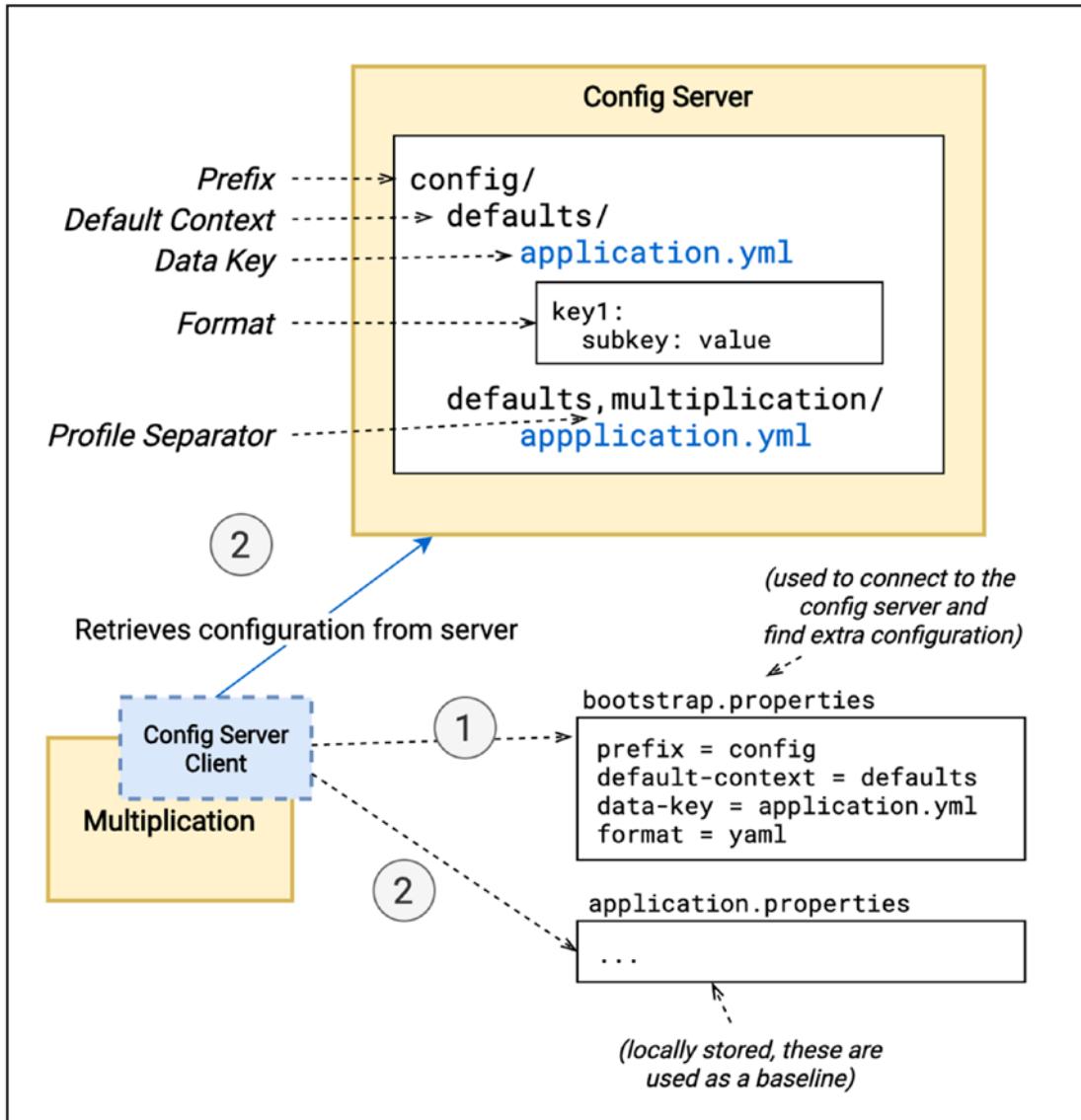


Figure 8-20. Config server properties: explanation

Keep in mind that, as shown in the previous figure, there is a difference between the application configuration to connect to the config server (in the bootstrap file) and the application configuration that results from merging the local properties (e.g., `application.properties`) with those downloaded from the config server. Since the first is a metaconfiguration, it can't be downloaded from the server, so we have to copy these values across our projects in the corresponding bootstrap configuration files.

Given that all these concepts are hard to understand without an example, let's use our system to explain how Consul Config works.

Implementing Centralized Configuration

Code Source

The code source with the integrated centralized configuration solution from Consul is located in the repository [chapter08c](#).

First, we need to add the new starter to the Multiplication, Gamification, and Gateway microservices. See Listing 8-35.

Listing 8-35. Adding the Spring Cloud Consul Config Dependency to Our Microservices

```
<dependencies>
    <!-- ... -->
    <dependency>
        <groupId>org.springframework.cloud</groupId>
        <artifactId>spring-cloud-starter-consul-config</artifactId>
    </dependency>
</dependencies>
```

By doing that, our apps will try to connect to Consul during the bootstrap phase and fetch the profile properties from the KV store using the defaults provided in Consul KV autoconfiguration. However, instead of using the defaults, we'll override some of these settings since that'll make the explanation clearer.

In the Multiplication and Gamification projects we use the properties format, so let's be consistent and create a separate file at the same level, named `bootstrap.properties`. In both applications, we'll set up the same settings. See Listing 8-36.

Listing 8-36. The new `bootstrap.properties` File in Multiplication and Gamification

```
spring.cloud.consul.config.prefix=config
spring.cloud.consul.config.format=yaml
spring.cloud.consul.config.default-context=defaults
spring.cloud.consul.config.data-key=application.yml
```

Note that we chose YAML as the format for the remote configuration, but our local file is in the `.properties` format. That's not a problem at all. Spring Cloud Consul Config can merge the values contained in the remote `application.yml` key with those stored locally in a different format.

Then, we create the equivalent settings in a `bootstrap.yml` file in the Gateway project, where we employed YAML for the application configuration. See Listing 8-37.

Listing 8-37. The New `bootstrap.yml` File in the Gateway Project

```
spring:
  cloud:
    consul:
      config:
        data-key: application.yml
        prefix: config
        format: yaml
        default-context: defaults
```

With these settings, our goal is to store all configuration within a root folder named `config` in Consul KV. Inside, we'll have a `defaults` folder that may contain a key named `application.yml` with the configuration that applies to all our microservices. We can have extra folders per application, or per combination of application and profiles that we want to use, and each of them may contain the `application.yml` key with the properties that should be added or overridden. To avoid mixing up formats in the configuration server, we'll stick to the YAML syntax. Review again the previous Figure 8-20 to understand better the overall structure of the configuration. What we've done until now is to add the

bootstrap files to Multiplication, Gamification, and Gateway, so they can connect to the config server and find the externalized configuration (if any). To enable this behavior, we added the Spring Cloud Consul Config starter dependency to all these projects too.

To use a more representative example, we could create the hierarchy shown in Listing 8-38 as folders and keys in Consul KV.

Listing 8-38. An Example Configuration Structure in the Configuration Server

```
+-- config
|   +- defaults
|   |   \- application.yml
|   +- defaults,production
|   |   \- application.yml
|   +- defaults,rabbitmq-production
|   |   \- application.yml
|   +- defaults,database-production
|   |   \- application.yml
|   +- multiplication,production
|   |   \- application.yml
|   +- gamification,production
|   |   \- application.yml
```

Then, if we run the Multiplication application with a list of active profiles equal to `production,rabbitmq-production,database-production`, the processing order would be the following (from lower to higher precedence):

1. The baseline values are those included in the local `application.properties` of the project that is accessing the configuration server, in this example Multiplication.
2. Then, Spring Boot merges and overrides the remote values included in the `application.yml` key inside the `defaults` folder, because it applies to all the services.
3. The next step is to merge the default values for all active profiles. That means all files that match the `defaults,{profile}` pattern: `defaults,production,defaults,rabbitmq-production,defaults,database-production`. Note that, if there are multiple profiles specified, the last one's values win.

- After that, it tries to find more specific settings for the corresponding application name and active profiles, following the pattern {application},{profile}. In our example, the key multiplication,production matches the pattern, so its configuration values will be merged. The precedence order is the same as before: the last profile in the enumeration wins.

See Figure 8-21 for a visual representation that surely will help you understand how all the configuration files gets applied.

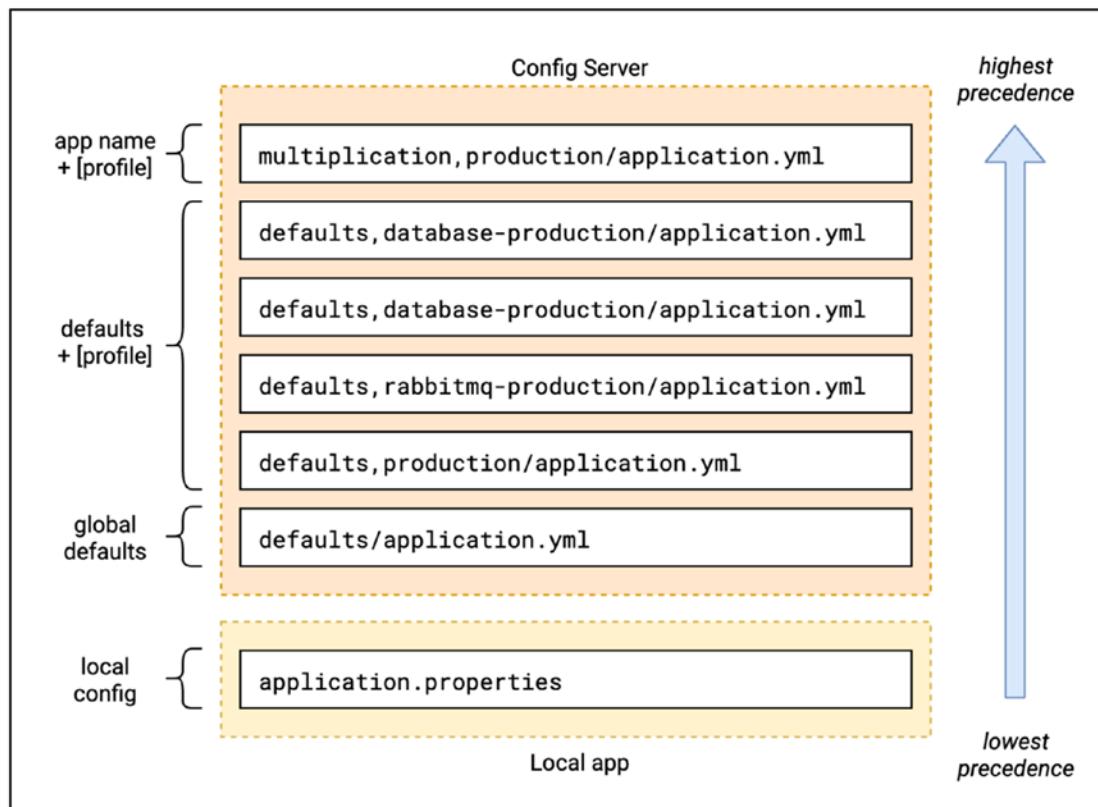


Figure 8-21. Configuration stack example

Therefore, a practical approach to structure configuration values could be as follows:

- Use defaults when you want to add global configuration to all the applications for all the environments such as when customizing JSON serialization.
- Use defaults,{profile} with a profile name representing a {tool}-{environment} pair to set up common values for a given tool per environment. For example, in our case, RabbitMQ connection values could be included in rabbitmq-production.
- Use {application},{profile} with a profile name equal to {environment} to set up specific settings for an application in a given environment. For example, we could reduce logging of the Multiplication microservice on production using properties inside multiplication,production.

Centralized Configuration in Practice

In the previous section, we added the new starter dependency to our projects and the additional bootstrap configuration properties to override some Consul Config defaults. If we start one of our services, it'll connect to Consul and try to retrieve configuration using the Consul's Key/Value API. For example, in the Multiplication application's logs, we'll see a new line with a list of *property sources* that Spring will try to find in the remote config server (Consul). See Listing 8-39.

Listing 8-39. Multiplication Logs, Indicating the Default Configuration Sources

```
INFO 54256 --- [main] b.c.PropertySourceBootstrapConfiguration : Located
property source: [BootstrapPropertySource {name='bootstrapProperties-config/
multiplication/'}, BootstrapPropertySource {name='bootstrapProperties-config/
defaults/'}]
```

This log line might be misleading since those property sources are not really located by the time this line is printed out. It's just a list of candidates. Their names match the patterns we described before. Given that we didn't enable any profile yet when starting the Multiplication application, it'll just try to find configuration under config/defaults and config/multiplication. As proven with this exercise, we don't need to create keys in Consul that match all possible candidates. Keys that don't exist will be just ignored.

Let's start creating some configuration in Consul. From the UI, on the Key/Value tab, click Create and enter config/ to create the root folder with the same name as we configured in our settings. Since we added the / character in the end, Consul knows it must create a folder. See Figure 8-22.

The screenshot shows the Consul UI interface. At the top, there is a navigation bar with tabs: dc1, Services, Nodes, Key/Value (which is highlighted in blue), ACL, and Intentions. To the right of the tabs are links for Documentation and Settings. Below the navigation bar, there is a breadcrumb trail labeled '< Key / Values'. The main content area has a heading 'config/'. Underneath the heading, there is a form field labeled 'Key or folder' containing the value 'config/'. A note below the field says 'To create a folder, end a key with /'. At the bottom of the form is a blue 'Save' button. At the very bottom of the page, there is a footer with the HashiCorp logo, the text '© 2020 HashiCorp', 'Consul 1.7.3', and a 'Documentation' link.

Figure 8-22. Consul: creating the config root folder

Now, navigate to the config folder by clicking the newly created item, and create a subfolder called defaults. See Figure 8-23.

The screenshot shows the Consul UI interface. At the top, there is a navigation bar with tabs: dc1, Services, Nodes, Key/Value (which is selected and highlighted in dark blue), ACL, Intentions, Documentation, and Settings. Below the navigation bar, the URL is shown as < Key / Values < config. The main content area has a heading 'defaults/'. Underneath it, there is a form field labeled 'Key or folder' containing 'defaults/'. A note below the field says 'To create a folder, end a key with /'. A blue 'Save' button is located at the bottom left of the form. At the very bottom of the page, there is a footer with the HashiCorp logo, the text '© 2020 HashiCorp', 'Consul 1.7.3', and 'Documentation'.

Figure 8-23. Consul: creating the defaults folder

Once more, navigate to the newly created folder by clicking it. You'll be seeing the contents of config/defaults, which are empty for now. Within this folder, we have to create a key named application.yml and put there the values we want to apply to all applications by default. Note that we decided to use a key name that looks like a file name to better distinguish folders from configuration contents. Let's add some logging configuration to enable the DEBUG level for a Spring package whose classes output some useful environment information. See Figure 8-24.

The screenshot shows the Consul UI interface. At the top, there's a navigation bar with icons for services, nodes, key/value, ACL, intentions, documentation, and settings. Below the navigation bar, the URL is shown as <Key / Values < config < defaults. The main area has a title "application.yml". Under "Key or folder", the value "application.yml" is entered. A note says "To create a folder, end a key with /". The "Value" section contains YAML code:

```

1 logging:
2   level:
3     org.springframework.core.env: DEBUG

```

A "Code" button is next to the "Value" label. At the bottom right of the editor area, there's a "YAML" dropdown and a "Save" button.

Figure 8-24. Consul: adding configuration to defaults

The Multiplication application should now pick up this new property. To verify it, we can restart it and check the logs, where we'll see now extra logging for the `org.springframework.core.env` package, in particular from the `PropertySourcesPropertyResolver` class:

```
DEBUG 61279 --- [main] o.s.c.e.PropertySourcesPropertyResolver : Found key
'spring.h2.console.enabled' in PropertySource 'configurationProperties' with value
of type String
```

This proves that the service reached the centralized configuration server (Consul) and applied the settings included in the existing expected keys, in this case `config/defaults`.

To make it more interesting, let's enable some profiles for the Multiplication application. From the command line, you can execute the following:

```
multiplication $ ./mvnw spring-boot:run -Dspring-boot.run.arguments="--spring.profiles.active=production,rabbitmq-production"
```

With that command, we're running the application with the production and rabbitmq-production profiles. The logs show the resulting candidate keys to look for. See Listing 8-40.

Listing 8-40. Multiplication Logs, Indicating All Candidate Property Sources After Enabling Extra Profiles

```
INFO 52274 --- [main] b.c.PropertySourceBootstrapConfiguration : Located property source: [BootstrapPropertySource {name='bootstrapProperties-config/multiplication,rabbitmq-production/'}, BootstrapPropertySource {name='bootstrapProperties-config/multiplication,production/'}, BootstrapPropertySource {name='bootstrapProperties-config/multiplication/'}, BootstrapPropertySource {name='bootstrapProperties-config/defaults,rabbitmq-production/'}, BootstrapPropertySource {name='bootstrapProperties-config/defaults,production/'}, BootstrapPropertySource {name='bootstrapProperties-config/defaults/'}]
```

Let's extract the property source names as a list for a better visualization. The list follows the same order as in the logs, from highest to lowest precedence.

1. config/multiplication,rabbitmq-production/
2. config/multiplication,production/
3. config/multiplication/
4. config/defaults,rabbitmq-production/
5. config/defaults,production/
6. config/defaults/

As we described in the previous section, Spring looks for keys that result from combining defaults with each profile, and then it looks for the combination of the application name with each profile. So far, we only added a key config/defaults, so that's the only one the service picks up.

In real life, we probably don't want to have the logs we added to all the applications on the production environment. To achieve this, we can configure the production profile to revert what we did earlier. Since this configuration has higher precedence, it'll override the previous value. Go to Consul UI and create a key within the config folder, with name `defaults,production`. Inside, you have to create a key `application.yml`, and its value should be the YAML configuration to set the log level of the package back to INFO. See Figure 8-25.

The screenshot shows the Consul UI interface. At the top, there's a navigation bar with tabs for Services, Nodes, Key/Value (which is selected), ACL, Intentions, Documentation, and Settings. Below the navigation bar, the URL path is shown as < Key / Values < config < defaults,production. The main content area is titled "application.yml". It contains a code editor with the following YAML configuration:

```

Value
1 logging:
2   level:
3     org.springframework.core.env: INFO

```

The code editor has a "Code" button (which is currently off) and a "YAML" button at the bottom right. At the bottom of the screen, there are three buttons: "Save" (blue), "Cancel changes" (gray), and "Delete" (red).

Figure 8-25. Consul: adding configuration to `defaults,production`

When we restart the application using the same last command (which enables the production profile), we'll see how the debug logging for that package is gone.

Remember that, same as we did with this simple logging example, we could add YAML values to tune any other configuration parameters in our Spring Boot applications to adapt them to the production environment. Also, note how we can play with the scope of the configuration using any of the six possible combinations listed

earlier, which we got by adding two active profiles. For example, we could add values that apply only to RabbitMQ on production inside a key named `defaults,rabbitmq-production`. The most specific combinations are `multiplication,rabbitmq-production` and `multiplication,production`. Check again Figure 8-21 for some visual help if you need it.

To demonstrate that configuration is not limited to logging, let's imagine that we want to run the Multiplication microservice on a different port (e.g., 10080) when deployed to production. To get this working, we only need to add an `application.yml` key inside the `multiplication,production` key in Consul and change the `server.port` property. See Figure 8-26.

The screenshot shows the Consul UI interface. At the top, there is a navigation bar with icons for dc1, Services, Nodes, Key/Value (which is selected and highlighted in dark blue), ACL, Intentions, Documentation, and Settings. Below the navigation bar, the URL is shown as < Key / Values < config < multiplication,production. The main area is titled "application.yml". On the left, there is a "Value" section containing the following YAML code:

```

1 server:
2   port: 10080

```

On the right, there is a "Code" toggle switch (set to "Code") and a "YAML" dropdown menu. At the bottom, there are three buttons: "Save" (blue background), "Cancel changes" (white background), and "Delete" (red border).

Figure 8-26. Consul: adding configuration to multiplication,production

Next time we start the Multiplication app with the production profile active, we'll see how it's running on this newly specified port:

```
INFO 29019 --- [main] o.s.b.w.embedded.tomcat.TomcatWebServer : Tomcat started on port(s): 10080 (http) with context path ''
```

With this exercise, we completed our overview of the centralized configuration pattern. Now, we know how to minimize the maintenance of common configuration and how to adapt applications to the environment they're running. See Figure 8-27 for an updated architecture view of our system including the new configuration server.

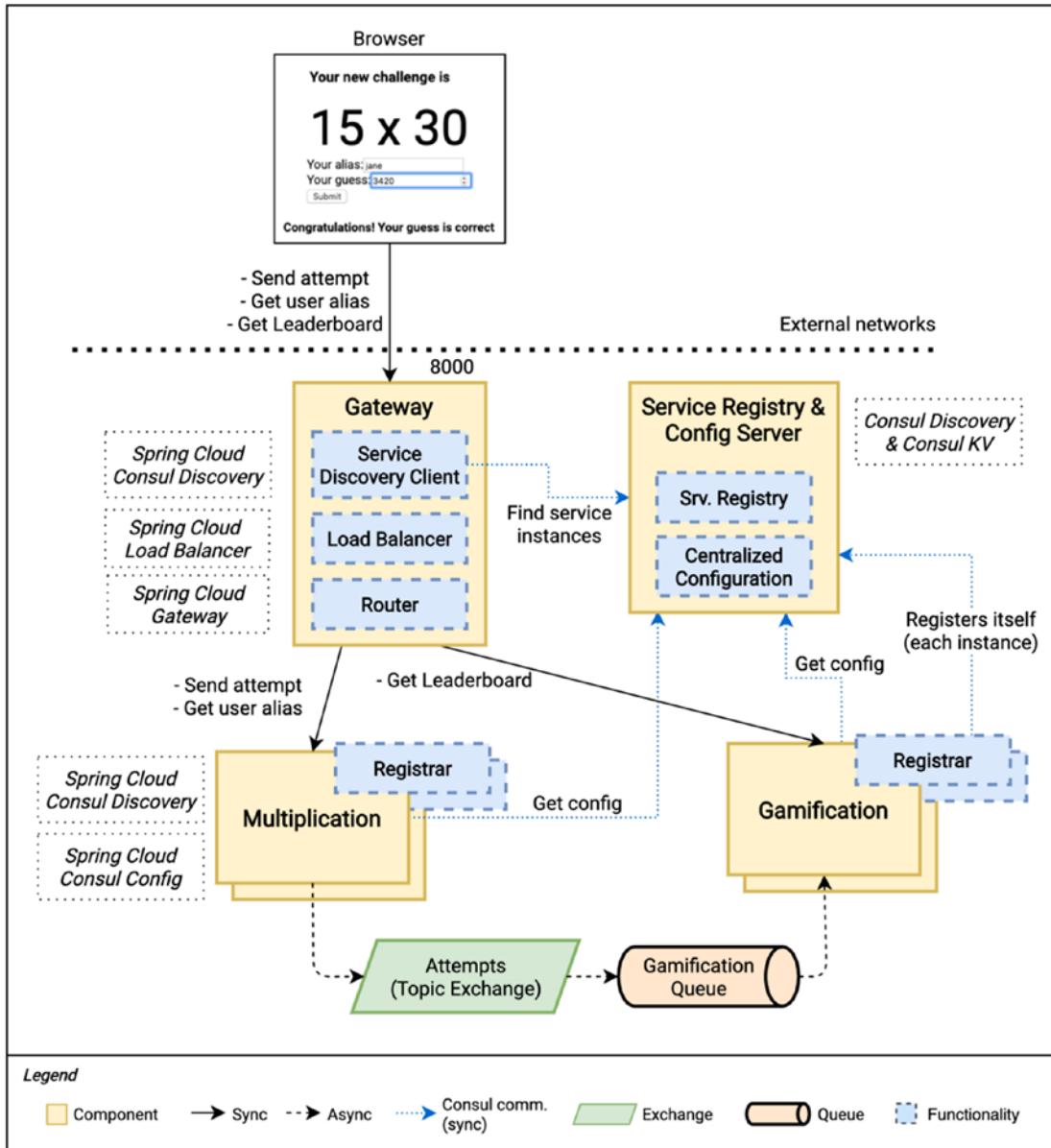


Figure 8-27. High-level overview: configuration server

Note that the applications have now a dependency with the configuration server at starting time. Luckily, we can configure Consul to be highly available in production, as we mentioned when covering the service discovery pattern (check <https://tpd.io/consulprod>). Additionally, Spring Cloud Consul counts with a retry mechanism by default, so our applications will keep retrying the connection to Consul when it's not available. This dependency is only at starting time; if Consul goes down while your applications are running, they keep working with the configuration loaded initially.

Note: Consul Configuration and Tests

By default, the integration tests in our projects will use the same application configuration. That means our controller tests and the default `@SpringBootTest` created by the Initializr will fail if Consul is not running because they keep waiting for the configuration server to be available. You can also disable Consul Config for tests easily; check <https://github.com/Book-Microservices-v2/chapter08c> if you're curious.

Centralized Logs

We already have multiple components in our system that produce logs (Multiplication, Gamification, Gateway, Consul, and RabbitMQ), and some of them might be running multiple instances. That's a lot of log outputs running independently, which makes it hard to get an overall view of the system activity. If a user reports an error, it would be hard to find out which component or instance failed. Arranging multiple log windows on a single screen would help for a while, but that's not a viable solution when your microservice instances grow in number.

To properly maintain a distributed system like our microservice architecture, we need a central place where we can access all the aggregated logs and search across them.

Log Aggregation Pattern

Basically, the idea is to send all the log outputs from our applications to another component in our system, which will consume them and put them all together. Besides, we want to persist these logs for some time, so this component should have a data storage. Ideally, we should be able to navigate through these logs, search, and filter out messages per microservice, instance, class, etc. To do this, many of these tools offer a user interface that connects to the aggregated logs storage. See Figure 8-28.

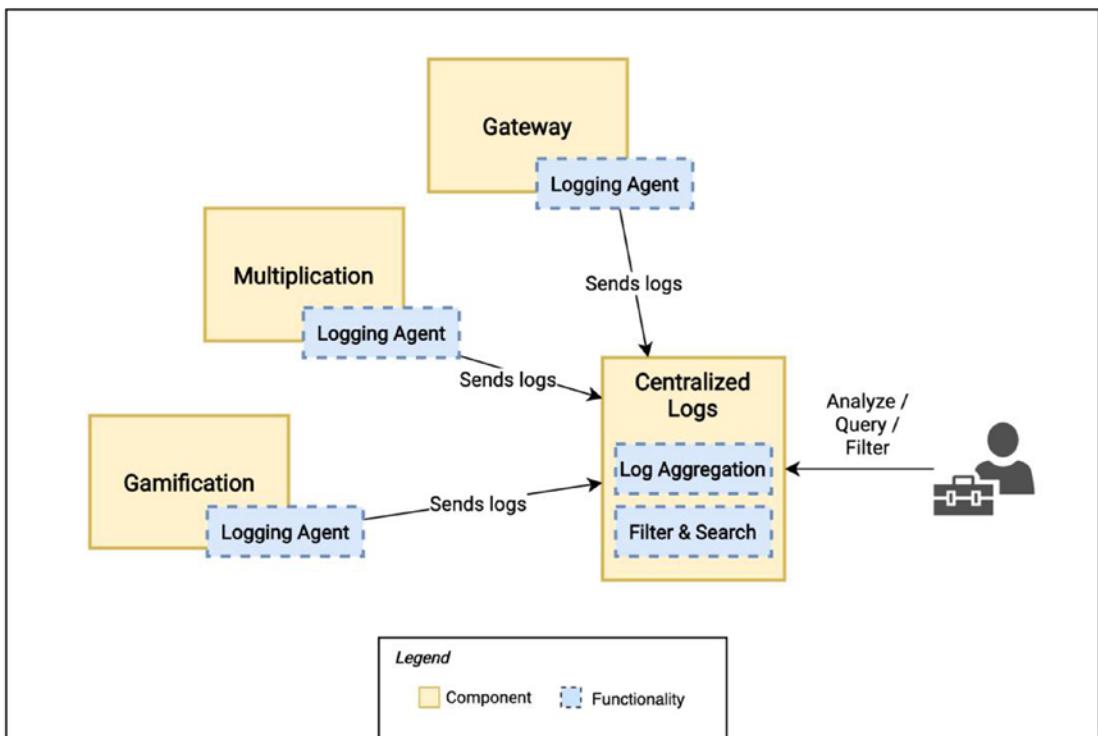


Figure 8-28. Log aggregation: overview

A common best practice when implementing the centralized logging approach is to keep the application logic unaware of this pattern. The services should just output messages using a common interface (e.g., a Logger in Java). The logging agent that channels these logs to the central aggregator works independently, capturing the output that the application produces.

There are multiple implementations of this pattern available in the market, both free and paid solutions. Among the most popular ones is the ELK stack, an alias for a combination of products from Elastic (<https://tpd.io/elasticsearch>): Elasticsearch (the storage system with powerful text search features), Logstash (the agent to channel logs to Elasticsearch from multiple sources), and Kibana (the UI tool to manage and query logs).

Even though setting up an ELK stack is becoming easier over time, it's still not a simple task. For that reason, we won't use an ELK implementation in this book since it could easily extend to cover a full chapter. In any case, I recommend you to check the ELK docs (<https://tpd.io/elk>) after reading this book, so you learn how to set up a production-ready logging system.

A simple Solution for Log Centralization

Code Source

The rest of the code sources for this chapter are in the repository `chapter08d`. It includes the changes for adding centralized logs, distributed tracing, and containerization.

What we'll do is to set up a new microservice to aggregate logs from all our Spring Boot applications. To keep it simple, it won't have a data layer to persist logs; it'll just receive log lines from other services and print them together to the standard output. This basic solution will serve us to demonstrate this pattern and the next one, distributed tracing.

To channel the log outputs, we'll use a tool we already have in our system and is perfect for that purpose: RabbitMQ. To capture each logged line in the applications and send them as RabbitMQ messages, we'll benefit from Logback, the logger implementation we've been using within Spring Boot. Given that this tool is driven by an external configuration file, we don't need to modify the code in our applications.

In Logback, the piece of logic that writes a log line to the specific destination is called an *appender*. This logging library includes some built-in appenders to print messages to the console (`ConsoleAppender`) or files (`FileAppender` and `RollingFileAppender`). We didn't need to configure them because Spring Boot includes some default Logback configuration within its dependencies and also sets up the printed message patterns.

The good news for us is that Spring AMQP provides a Logback AMQP logging appender that does exactly what we need: it takes each log line and produces a message to a given exchange in RabbitMQ, with a format and some extra options that we can customize.

First, let's prepare the Logback configuration we need to add to our applications. Spring Boot allows us to extend the defaults by creating a file named `logback-spring.xml` in the application resources folder (`src/main/resources`), which will be picked up automatically upon application initialization. See Listing 8-41. In this file, we import the existing default values and create and set a new appender for all messages that have level INFO or higher. The AMQP appender documentation (<https://tpd.io/amqp-appender>) lists all parameters and their meaning; let's detail the ones we need.

- `applicationId`: We set it to the application name so we can distinguish the source when we aggregate logs.
- `host`: This is the host where RabbitMQ is running. Since it can be different per environment, we'll connect this value to the Spring property `spring.rabbitmq.host`. Spring allows us to do this via the tag `springProperty`. We give this Logback property a name, `rabbitMQHost`, and we use the syntax `${rabbitMQHost:-localhost}` to either use the property value if it's set or use the default `localhost` (defaults are set with the `:`- separator).
- `routingKeyPattern`: This is the routing key per message, which we set to a concatenation of the `applicationId` and `level` (notated with `%p`) for more flexibility if we want to filter on the consumer side.
- `exchangeName`: We specify the name of the exchange in RabbitMQ to publish messages. It'll be a topic exchange by default, so we can call it `logs.topic`.
- `declareExchange`: We set it to `true` to create the exchange if it's not there yet.
- `durable`: Also set this to `true` so the exchange survives server restarts.
- `deliveryMode`: We make it `PERSISTENT` so log messages are stored until they're consumed by the aggregator.

- `generateId`: We set it to true so each message will have a unique identifier.
- `charset`: It's a good practice to set it to UTF-8 to make sure all parties use the same encoding.

Listing 8-41 shows the full contents of the `logback-spring.xml` file in the Gamification project. Note how we're adding a layout with a custom pattern to our new appender. This way, we can encode our messages including not only the message (`%msg`) but also some extra information like the time (`%d{HH:mm:ss.SSS}`), the thread name (`[%t]`), and the logger class (`%logger{36}`). If you're curious about the pattern notation, check the Logback's reference docs (<https://tpd.io/logback-layout>). The last part of the file configures the root logger (the default one) to use both the CONSOLE appender, defined in one of the included files, and the newly defined AMQP appender.

Listing 8-41. New logback-spring.xml File in the Gamification Project

<configuration>

```

<include resource="org/springframework/boot/logging/logback/defaults.xml" />
<include resource="org/springframework/boot/logging/logback/console-appender.
xml" />

<springProperty scope="context" name="rabbitMQHost" source="spring.rabbitmq.
host"/>

<appender name="AMQP"
          class="org.springframework.amqp.rabbit.logback.AmqpAppender">
    <layout>
        <pattern>%d{HH:mm:ss.SSS} [%t] %logger{36} - %msg</pattern>
    </layout>
    <applicationId>gamification</applicationId>
    <host>${rabbitMQHost:-localhost}</host>
    <routingKeyPattern>%property{applicationId}.%p</routingKeyPattern>
    <exchangeName>logs.topic</exchangeName>
    <declareExchange>true</declareExchange>
    <durable>true</durable>
    <deliveryMode>PERSISTENT</deliveryMode>
    <generateId>true</generateId>

```

```

< charset>UTF-8</ charset>
</ appender >

< root level="INFO" >
    < appender-ref ref="CONSOLE" />
    < appender-ref ref="AMQP" />
</ root >
</ configuration >
```

Now, we have to make sure we add this file to the three Spring Boot projects we have: Multiplication, Gamification, and Gateway. In each one of them, we must change the applicationId value accordingly.

In addition to this basic setup of log producers, we can adjust the log level for the class that the appender uses to connect to RabbitMQ as `WARN`. This is an optional step, but it avoids hundreds of logs when the RabbitMQ server is not available (e.g., while starting up our system). Since the appender is configured during the bootstrap phase, we'll add this configuration setting to the corresponding `bootstrap.properties` and `bootstrap.yml` files, depending on the project. See Listings 8-42 and 8-43.

Listing 8-42. Reducing RabbitMQ Logging Level in Multiplication and Gamification

```
logging.level.org.springframework.amqp.rabbit.connection.CachingConnectionFactory
= WARN
```

Listing 8-43. Reducing RabbitMQ Logging Level in the Gateway

```
logging:
  level:
    org.springframework.amqp.rabbit.connection.CachingConnectionFactory: WARN
```

The next time we start our applications, all logs will be output not only to the console but also as messages produced to the `logs.topic` exchange in RabbitMQ. You can verify that by accessing the RabbitMQ Web UI at `localhost:15672`. See Figure 8-29.

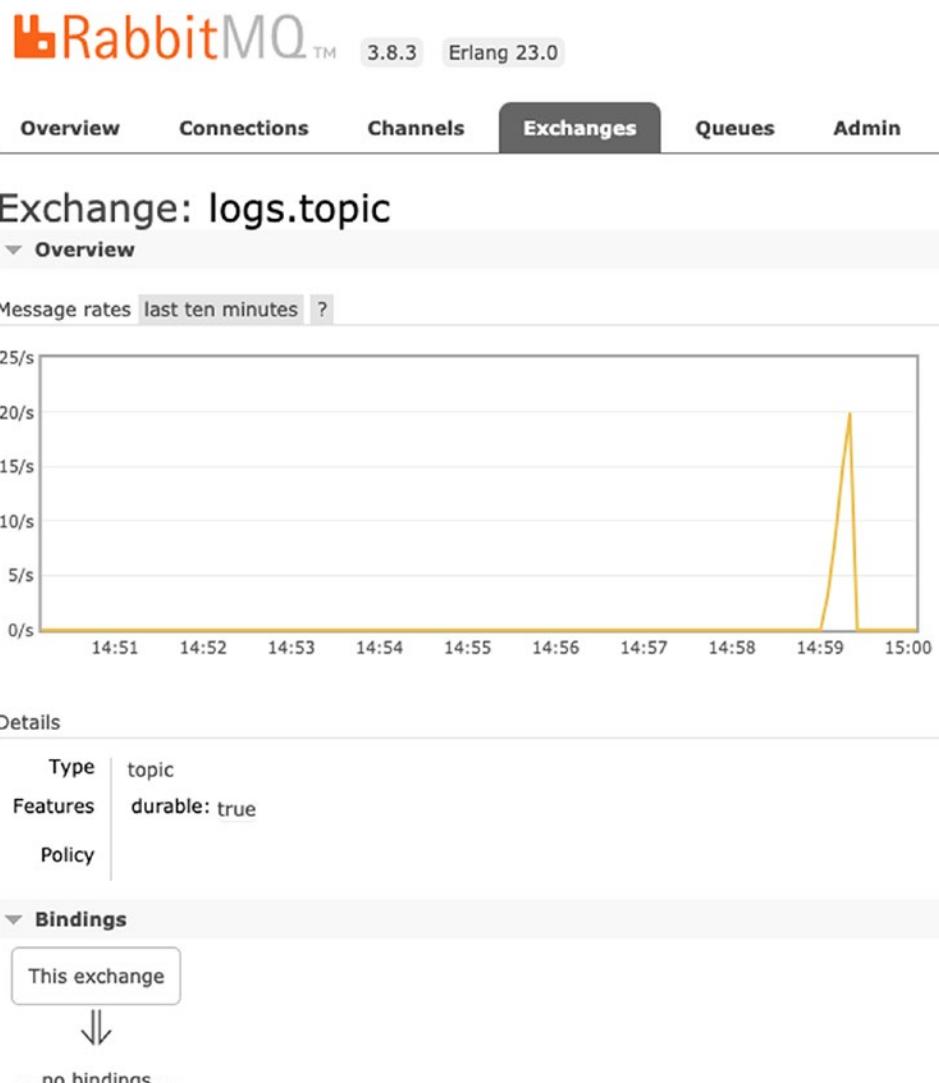


Figure 8-29. RabbitMQ UI: logs exchange

Consuming Logs and Printing Them

Now that we have all logs together published to an exchange, we'll build the consumer side: a new microservice that consumes all these messages and outputs them together.

CHAPTER 8 COMMON PATTERNS IN MICROSERVICE ARCHITECTURES

First, navigate to the Spring Initializr site start.spring.io/ (<https://start.spring.io/>) and create a logs project using the same setup as we chose for other applications: Maven and JDK 14. In the list of dependencies, we add Spring for RabbitMQ, Spring Web, Validation, Spring Boot Actuator, Lombok, and Consul Configuration. Note that we don't need to make this service discoverable, so we're not adding Consul Discovery. See Figure 8-30.

The screenshot shows the Spring Initializr interface for creating a 'logs' microservice. The 'Project' section is set to 'Maven Project'. The 'Language' section shows 'Java' selected. Under 'Spring Boot', version '2.4.0 (SNAPSHOT)' is selected. The 'Dependencies' section includes 'Lombok' (selected), 'Spring Web' (selected), 'Validation' (selected), 'Spring for RabbitMQ' (selected), and 'Spring Boot Actuator' (selected). The 'Project Metadata' section shows the group as 'microservices.book', artifact as 'logs', name as 'logs', and package name as 'microservices.book.logs'. The Java version is set to 14. At the bottom, there are 'GENERATE' and 'EXPLORE' buttons, along with a 'SHARE...' button.

Figure 8-30. Creating the Logs microservice

Once we import this project into our workspace, we add some configuration to make it possible to connect to the configuration server. We're not going to add any specific configuration for now, but it's good to do this to make it consistent with the rest of the

microservices. In the `main/src/resources` folder, copy the contents of the `bootstrap.properties` file we included in other projects. Besides, let's set the application name and a dedicated port in the `application.properties` file. See Listing 8-44.

Listing 8-44. Adding Content to the `application.properties` File in the New Logs Application

```
spring.application.name=logs
server.port=8580
```

We need a Spring Boot configuration class to declare the exchange, the queue where we want to consume the messages from, and the binding object to attach the queue to the topic exchange with a binding key pattern to consume all of them, `#`. See Listing 8-45. Remember that since we added the logging level to the routing keys, we could also adjust this value to get only errors, for example. Anyway, in our case, we subscribe to all messages (`#`).

Listing 8-45. AMQPConfiguration Class in the Logs Application

```
package microservices.book.logs;

import org.springframework.amqp.core.*;
import org.springframework.context.annotation.Bean;
import org.springframework.context.annotation.Configuration;

@Configuration
public class AMQPConfiguration {

    @Bean
    public TopicExchange logsExchange() {
        return ExchangeBuilder.topicExchange("logs.topic")
            .durable(true)
            .build();
    }

    @Bean
    public Queue logsQueue() {
        return QueueBuilder.durable("logs.queue").build();
    }
}
```

```

@Bean
public Binding logsBinding(final Queue logsQueue,
                           final TopicExchange logsExchange) {
    return BindingBuilder.bind(logsQueue)
        .to(logsExchange).with("#");
}
}

```

The next step is to create a simple service with the `@RabbitListener` that maps the logging level of the received messages, passed as a RabbitMQ message header, to a logging level in the Logs microservice, using the corresponding `log.info()`, `log.error()`, or `log.warn()`. Note that we use here the `@Header` annotation to extract AMQP headers as method arguments. We also use a logging Marker to add the application name (`appId`) to the log line without needing to concatenate it as part of the message. This is a flexible way in the SLF4J standard to add contextual values to logs. See Listing 8-46.

Listing 8-46. The Consumer Class That Receives All Log Messages via RabbitMQ

```

package microservices.book.logs;

import org.springframework.amqp.rabbit.annotation.RabbitListener;
import org.springframework.messaging.handler.annotation.Header;
import org.springframework.stereotype.Service;

import org.slf4j.Marker;
import org.slf4j.MarkerFactory;

import lombok.extern.slf4j.Slf4j;

@Slf4j
@Service
public class LogsConsumer {

    @RabbitListener(queues = "logs.queue")
    public void log(final String msg,
                   @Header("level") String level,
                   @Header("amqp_appId") String appId) {
        Marker marker = MarkerFactory.getMarker(appId);

```

```

switch (level) {
    case "INFO" -> log.info(marker, msg);
    case "ERROR" -> log.error(marker, msg);
    case "WARN" -> log.warn(marker, msg);
}
}
}
}

```

Finally, we customize the log output produced by this new microservice. Since it'll aggregate multiple logs from different services, the most relevant property is the application name. We override the Spring Boot defaults this time and define a simple format in a `logback-spring.xml` file for the CONSOLE appender that outputs the marker, the level, and the message. See Listing 8-47.

Listing 8-47. The LogBack Configuration for the Logs Application

```

<configurationappender name="CONSOLE" class="ch.qos.logback.core.ConsoleAppender">
    <layout class="ch.qos.logback.classic.PatternLayout">
      <PatternPattern>
    </layout>
  </appender>

  <root level="INFO">
    <appender-ref ref="CONSOLE" />
  </root>
</configuration>

```

That's all the code we need in this new project. Now, we can build the sources and start this new microservice with the rest of the components in our system.

1. Run the RabbitMQ server.
2. Run the Consul agent in development mode.
3. Start the Multiplication microservice.
4. Start the Gamification microservice.

5. Start the Gateway microservice.
6. Start the Logs microservice.
7. Run the front-end app.

Once we start this new microservice, it'll consume all log messages produced by the other applications. To see that in practice, you can solve a challenge. You'll see the log lines shown in Listing 8-48 in the console of the Logs microservice.

Listing 8-48. Centralized Logs in the New Logs Application

```
[multiplication ] INFO 15:14:20.203 [http-nio-8080-exec-1] m.b.m.c.ChallengeAttemptController - Received new attempt from test1
[gamification ] INFO 15:14:20.357 [org.springframework.amqp.rabbit.RabbitListenerEndpointContainer#0-1] m.b.g.game.GameEventHandler - Challenge Solved Event received: 122
[gamification ] INFO 15:14:20.390 [org.springframework.amqp.rabbit.RabbitListenerEndpointContainer#0-1] m.b.g.game.GameServiceImpl - User test1 scored 10 points for attempt id 122
```

This simple log aggregator didn't take us much time, and now we can search for logs within the same source and see a near-real-time output stream from all our services together. See Figure 8-31 for an updated version of our high-level architecture diagram including this new component.

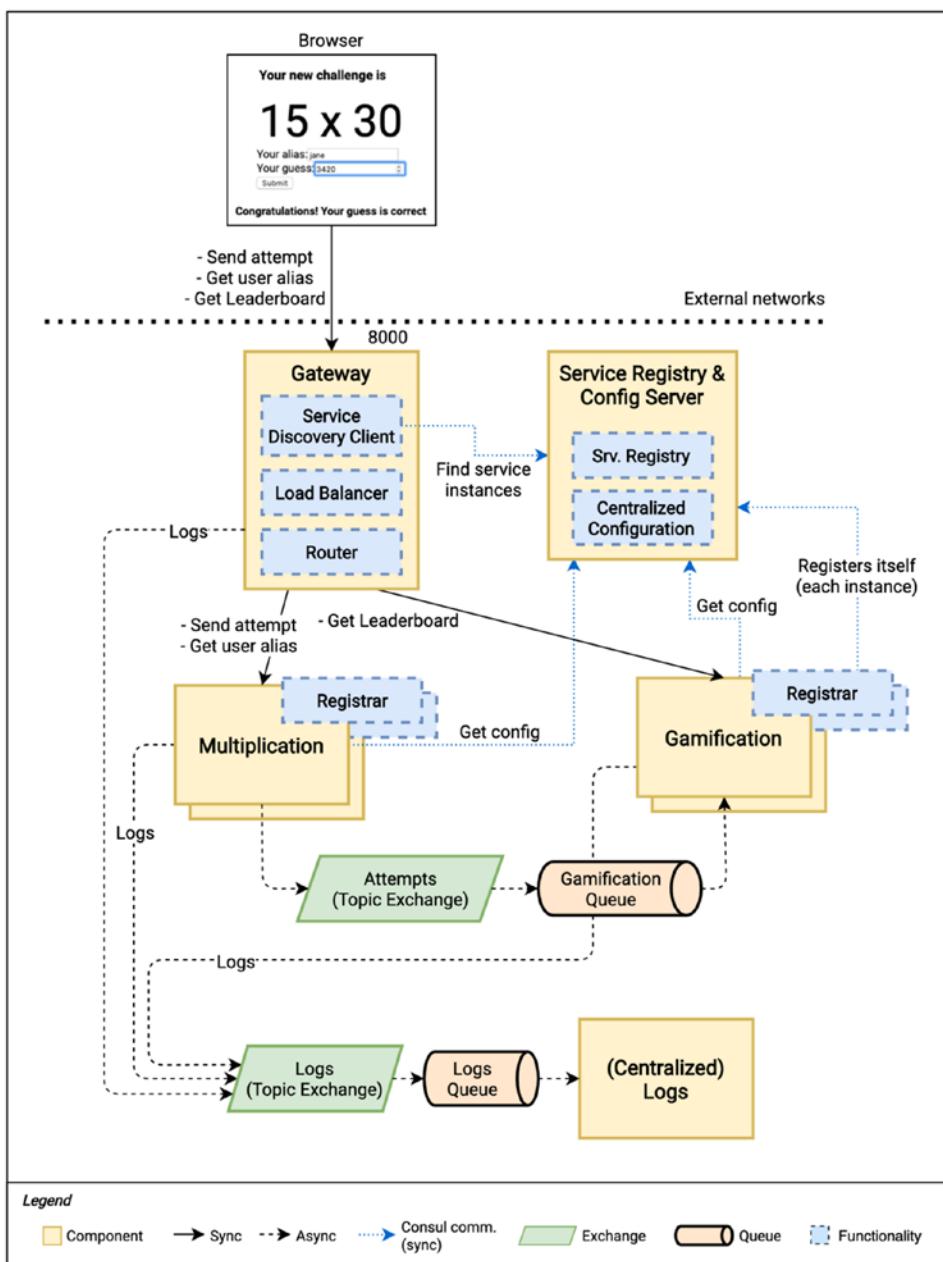


Figure 8-31. High-level overview: centralized logs

If we would choose an existing solution for our log aggregation, the overall steps would be similar. Many of these tools like the ELK stack can integrate with Logback to get your logs, via custom appenders. Then, in the case of non-cloud-based log aggregators, we would also need to deploy the log server in our system, as we did with the basic microservice we created.

Distributed Tracing

Having all the logs in one place is a great achievement that improves *observability*, but we don't have proper *traceability* yet. In the previous chapter, we described how a mature event-driven system may have processes that span different microservices. Knowing what's going on with many concurrent users and multiple event chains might become an impossible task, especially when these chains have branches with multiple event types triggering the same action.

To solve this problem, we need to correlate all actions and events within the same process chain. A simple way of doing this is to inject the same identifier across all HTTP calls, RabbitMQ messages, and Java threads handling the different actions. Then, we can print this identifier in all the related logs.

In our system, we work with user identifiers. If we think that all our future functionality will be built around user actions, we could propagate a `userId` field in every event and call. Then, we can log it in the different services, so we could correlate logs to specific users. That would definitively improve traceability. However, we may also have multiple actions from the same user happening in a short period, for example, two attempts to solve a multiplication within one second (a fast user, but you get the idea), spread across multiple instances. In that case, it would be hard to distinguish between individual flows across our microservices. Ideally, we should have a unique identifier per action, which is generated at the origin of the chain. Furthermore, it'd be better if we can propagate it transparently, without having to model this traceability concern explicitly in all our services.

As it happens many times in software development, we're not the first ones dealing with this challenge. That's again good news since it means there are solutions available for us to use with minimum effort. In this case, the tool that implements distributed tracing in Spring is called Sleuth.

Spring Cloud Sleuth

Sleuth is part of the Spring Cloud family, and it uses the Brave library (<https://tpd.io/brave>) to implement distributed tracing. It builds traces across different components by correlating units of work called *spans*. For example, in our system, one span is checking the attempt in the Multiplication microservice, and a different one is adding score and badges based on the RabbitMQ event. Each span has a different unique identifier, but both spans are part of the same trace, so they have the same trace identifier. Besides, each span links to its parent, except the root one because it's the original action. See Figure 8-32.

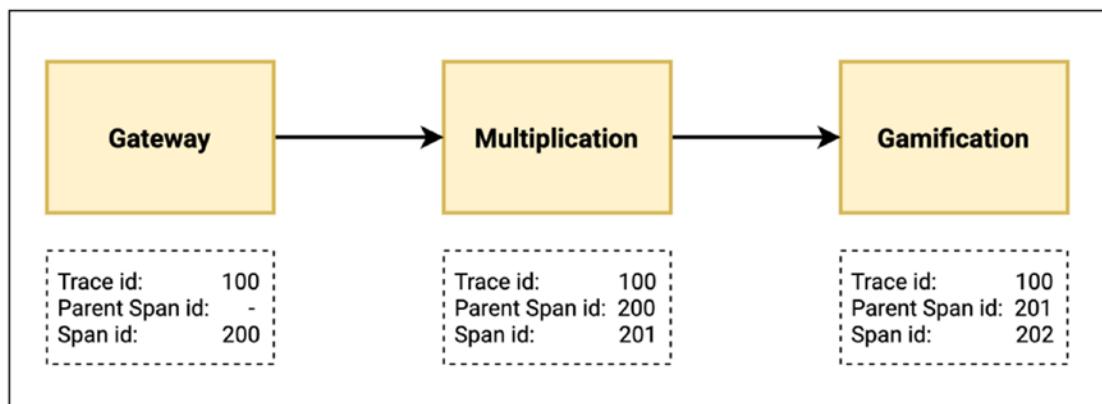


Figure 8-32. Distributed tracing: simple example

In more evolved systems, there might be complex trace structures where multiple spans have the same parent. See Figure 8-33.

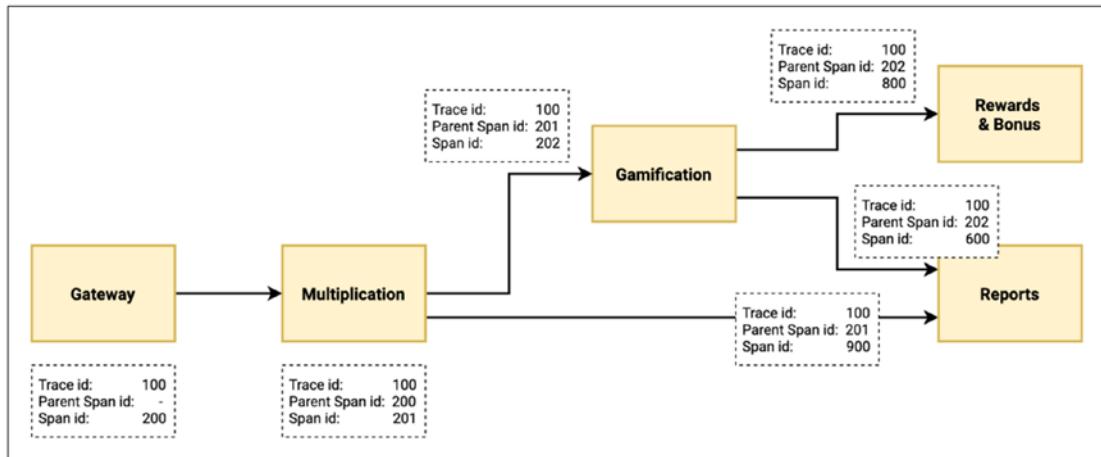


Figure 8-33. Distributed tracing: tree example

To inject these values transparently, Sleuth uses the SLF4J's Mapped Diagnostic Context (MDC) object, a logging context whose lifecycle is limited to the current thread. The project also allows us to inject our own fields in this context, so we could propagate them and use these values in logs.

Spring Boot autoconfigures some built-in interceptors in Sleuth to automatically inspect and modify HTTP calls and RabbitMQ messages. It also has integration with Kafka, gRPC, and other communication interfaces. These interceptors all work in a similar way: for incoming communications, they check if there are tracing headers added to the calls or messages and put them into the MDC; when doing calls as a client or publishing data, these interceptors take these fields from the MDC and add headers to the requests or messages.

Sleuth is sometimes combined with Zipkin, a tool that uses trace sampling to measure the time spent in each span and therefore in the complete chain. These samples can be sent to a Zipkin server, which exposes a UI that you can use to see the trace hierarchy and the time it takes for each service to do its part. We won't use Zipkin in this book since that doesn't add much value on top of a centralized logging system with trace and span identifiers, where you can also know the time spent in each service if you check the logged timestamps. Anyway, you can easily integrate Zipkin in our example project by following the instructions in the reference docs (<http://tpd.io/spans-zipkin>).

Implementing Distributed Tracing

As mentioned, Spring Cloud Sleuth provides interceptors for REST APIs and RabbitMQ messages, and Spring Boot autoconfigures them for us. It's not hard to have distributed tracing in our system.

First, let's add the corresponding Spring Cloud starter to our Gateway, Multiplication, Gamification, and Logs microservices. See Listing 8-49 for the dependency we have to add to the `pom.xml` files.

Listing 8-49. Adding Spring Cloud Sleuth to All Our Spring Boot Projects

```
<dependency>
  <groupId>org.springframework.cloud</groupId>
  <artifactId>spring-cloud-starter-sleuth</artifactId>
</dependency>
```

Only by adding this dependency will Sleuth inject the trace and span identifiers to every supported communication channel and to the MDC object. The default Spring Boot logging patterns are also automatically adapted to print the trace and span values in the logs.

To make our logs more verbose and see the trace identifiers in action, let's add a log line to `ChallengeAttemptController` to print a message each time a user sends an attempt. See the change in Listing 8-50.

Listing 8-50. Adding a Log Line to `ChallengeAttemptController`

```
@PostMapping
 ResponseEntity<ChallengeAttempt> postResult(
     @RequestBody @Valid ChallengeAttemptDTO challengeAttemptDTO) {
    log.info("Received new attempt from {}", challengeAttemptDTO.getUserAlias());
    return ResponseEntity.ok(challengeService.verifyAttempt(challengeAttemptDTO));
}
```

Besides, we want to have the trace and parent identifiers in our centralized logs as well. To do that, let's add manually the properties `X-B3-TraceId` and `X-B3-SpanId` from the MDC context (injected by Sleuth using Brave) to our pattern in the `logback-spring.xml` file in the Logs project. These headers are part of the OpenZipkin's B3 Propagation

specification (see <http://tpd.io/b3-headers> for more details), and they are included in the MDC by the Sleuth's interceptors. We need to do this manually for our Logs microservice since we're not using the Spring Boot defaults in this logging configuration file. See Listing 8-51.

Listing 8-51. Adding Trace Fields to Each Log Line Printed by the Logs Application

<configuration>

```

<appender name="CONSOLE" class="ch.qos.logback.core.ConsoleAppender">
    <layout class="ch.qos.logback.classic.PatternLayout">
        <Pattern>
            [%-15marker] [%X{X-B3-TraceId:-},%X{X-B3-SpanId:-}] %highlight
            (%-5level) %msg%n
        </Pattern>
    </layout>
</appender>

<root level="INFO">
    <appender-ref ref="CONSOLE" />
</root>
</configuration>

```

Once we restart all back-end services, Sleuth will do its part. Let's use the terminal to send a correct attempt directly to the back end.

```
$ http POST :8000/attempts factorA=15 factorB=20 userAlias=test-user-tracing
guess=300
```

Then, we check the output of the Logs service. We'll see the two fields showing the common trace identifier across the Multiplication and Gamification microservices, fa114ad129920dc7. Each line also has its own span ID. See Listing 8-52.

Listing 8-52. Centralized Logs with Trace Identifiers

```
[multiplication] [fa114ad129920dc7,4cdc6ab33116ce2d] INFO 10:16:01.813 [http-nio-8080-exec-8] m.b.m.c.ChallengeAttemptController - Received new attempt from test-user-tracing
```

```
[multiplication ] [fa114ad129920dc7,f70ea1f6a1ff6cac] INFO 10:16:01.814 [http-nio-8080-exec-8] m.b.m.challenge.ChallengeServiceImpl - Creating new user with alias test-user-tracing
[gamification ] [fa114ad129920dc7,861cbac20a1f3b2c] INFO 10:16:01.818 [org.springframework.amqp.rabbit.RabbitListenerEndpointContainer#0-1] m.b.g.game.GameEventHandler - Challenge Solved Event received: 126
[gamification ] [fa114ad129920dc7,78ae53a82e49b770] INFO 10:16:01.819 [org.springframework.amqp.rabbit.RabbitListenerEndpointContainer#0-1] m.b.g.game.GameServiceImpl - User test-user-tracing scored 10 points for attempt id 126
```

As you can see, with very little effort, we got a powerful feature that allows us to discern separate processes in our distributed system. As you can imagine, this works even better when we output all logs with their traces and spans to a more sophisticated centralized logs tooling like ELK, where we could use these identifiers to perform filtered text searches.

Containerization

Until now, we have been executing locally all our Java microservices, the React front end, RabbitMQ, and Consul. To make that work, we needed to install the JDK to compile the sources and run the JAR packages, Node.js to build and run the UI, the RabbitMQ server (Erlang included), and Consul's agent. As our architecture evolves, we might need to introduce other tools and services, and they surely have their own installation processes that may differ depending on the operating system and its version.

As an overall goal, we want to be able to run our back-end system in multiple environments, no matter what OS version they're running. Ideally, we'd like to benefit from a “build once, deploy anywhere” strategy, and avoid repeating all the configuration and installation steps in every environment we want to deploy our system. Besides, the deployment process should be as straightforward as possible.

In the past, a common approach to package complete systems to run them anywhere was to create a virtual machine (VM). There are a few solutions to create and run VMs, and they're called *hypervisors*. An advantage of hypervisors is that a physical machine can run multiple VMs at the same time, and they all share the hardware resources. Every VM requires its own operating system, which is then connected via the hypervisor to the host's CPU, RAM, hard disks, etc.

In our case, we could create a VM starting with a Linux distribution and set up and install there all the tools and services that are needed to run our system: Consul, RabbitMQ, a Java Runtime, the JAR applications, etc. Once we know the virtual machine works, we could transfer it to any other computer running the Hypervisor. Since the package contains everything it's needed, it should work the same in a different host. See Figure 8-34.

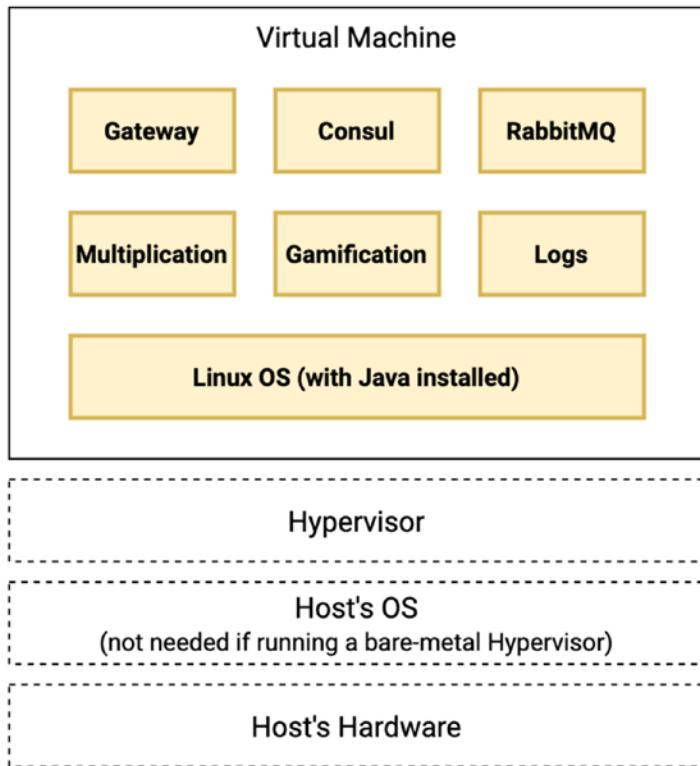


Figure 8-34. Virtual machine deployment: single

However, putting everything together in the same VM is not very flexible. If we want to scale up our system, we have to go inside the virtual machine, add new instances, and make sure we allocate more CPU, memory, etc. We need to know how everything works, so the deployment process is not that easy anymore.

A more dynamic approach would be to have separate virtual machines per service and tooling. Then, we add some network configuration to make sure they can connect to each other. Since we use service discovery and dynamic scaling, we could add more instances of virtual machines running microservices (e.g., Multiplication-VM), and

they will be used transparently. These new instances just need to register themselves in Consul using their address (within the VM network). See Figure 8-35. That's better than the single VM, but it's a huge waste of resources given that each virtual machine requires its own operating system. Also, it would bring a lot of challenges in terms of VM orchestration: monitoring them, creating new instances, configuring networks, storage, etc.

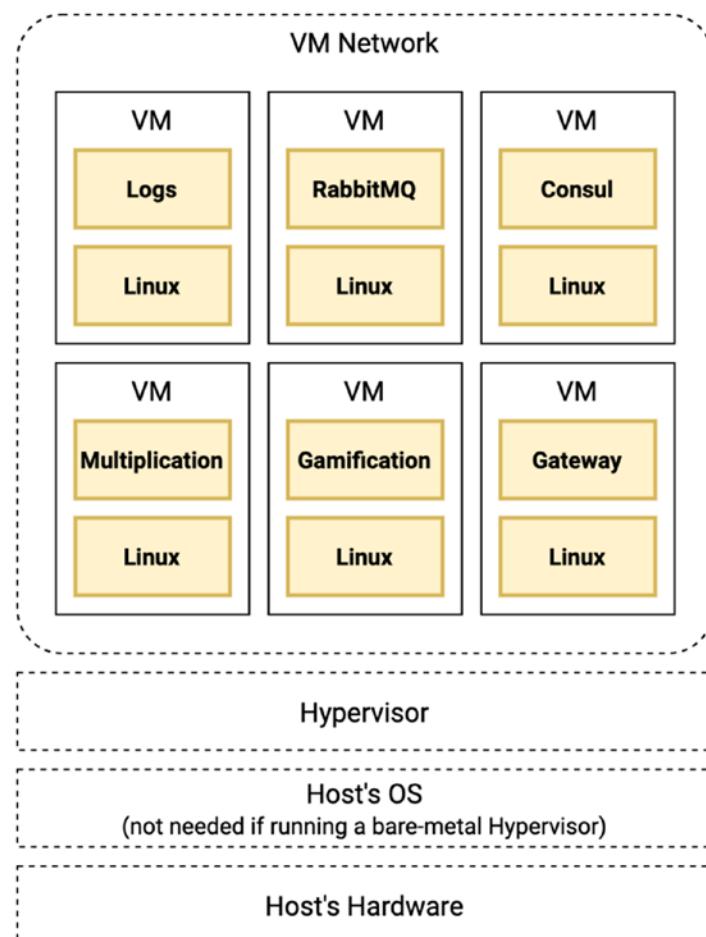


Figure 8-35. Virtual machine deployment: multiple

With the evolution of containerization technologies in the early 2010s, VMs have fallen into disuse, and containers have emerged as the most popular way of application virtualization. Containers are much smaller because they don't need the operating system; they run on top of a host's Linux operating system.

On the other hand, the introduction of containerization platforms like Docker has facilitated cloud and on-premise deployments drastically, with easy-to-use tools to package applications, run them as containers, and share them in a common registry. Let's explore this platform's features in more detail.

Docker

It would be impossible to cover in this book all the concepts of the Docker platform, but let's try to give an overview that is good enough to understand how it facilitates the deployment of a distributed system. The Get Started page (<https://tpd.io/docker-start>) on the official website is a good place to continue learning from here.

In Docker, we can package our application and any supporting component it may need as images. These can be based on other existing images that you can pull from a Docker registry, so we can reuse them and save a lot of time. The official public registry of images is the Docker Hub (<https://tpd.io/docker-hub>).

As an example, a Docker image for the Multiplication microservice could be based on the existing JDK 14 image. Then, we can add on top of it the JAR file packaged by Spring Boot. To create images, we need a Dockerfile, with a set of instructions for the Docker CLI tool. Listing 8-53 shows how an example Dockerfile could look like for the Multiplication microservice. This file should be placed in the root folder of the project.

Listing 8-53. A Basic Dockerfile to Create a Docker Image for the Multiplication Microservice

```
FROM openjdk:14
COPY ./target/multiplication-0.0.1-SNAPSHOT.jar /usr/src/multiplication/
WORKDIR /usr/src/multiplication
EXPOSE 8080
CMD ["java", "-jar", "multiplication-0.0.1-SNAPSHOT.jar"]
```

These instructions tell Docker to use version 14 of the official openjdk image in the public registry (Docker Hub, <https://tpd.io/docker-jdk>) as a base (FROM). Then, it copies the distributable .jar file from the current project to the /usr/src/multiplication/ folder in the image (COPY). The third instruction, WORKDIR, changes the working directory of the image to this newly created folder. The command EXPOSE informs Docker that this image exposes a port, 8080, where we serve the REST API. Finally, we define the command to

execute when running this image with CMD. That's just the classic Java command to run a .jar file, split into three parts to comply with the expected syntax. There are many other instructions that you can use in a Dockerfile, as you can see in the reference docs (<https://tpd.io/dockerfile-ref>).

To build images, we have to download and install the *Docker CLI* tool, which comes with the standard Docker installation package. Follow the instructions on the Docker website (<https://tpd.io/getdocker>) to get the proper package for your OS. Once downloaded and started, the Docker daemon should be running as a background service. Then, you can build and deploy images using Docker commands from a terminal. For example, the command shown in Listing 8-54 builds the multiplication image based on the Dockerfile we created earlier. Note that, as a prerequisite, you have to make sure you build and package the app in a .jar file, for example by running `./mvnw clean` package from the project's root folder.

Listing 8-54. Building a Docker Image Manually

```
multiplication$ docker build -t multiplication:1.0.0 .

Sending build context to Docker daemon 59.31MB
Step 1/5 : FROM openjdk:14
--> 4fba8120f640
Step 2/5 : COPY ./target/multiplication-0.0.1-SNAPSHOT.jar /usr/src/
multiplication/
--> 2e48612d3e40
Step 3/5 : WORKDIR /usr/src/multiplication
--> Running in c58cde6bda82
Removing intermediate container c58cde6bda82
--> 8d5457683f2c
Step 4/5 : EXPOSE 8080
--> Running in 7696319884c7
Removing intermediate container 7696319884c7
--> abc3a60b73b2
Step 5/5 : CMD ["java", "-jar", "multiplication-0.0.1-SNAPSHOT.jar"]
--> Running in 176cd53fe750
Removing intermediate container 176cd53fe750
--> a42cc81bab51
Successfully built a42cc81bab51
Successfully tagged multiplication:1.0.0
```

As you see in the output, Docker processes every line in the file and creates an image named `multiplication:1.0.0`. This image is available only locally, but we could *push* it to a remote location if we would like others to use it, as we'll explain later.

Once you have built a Docker image, you can run it as a container, which is a running instance of an image. As an example, this command would run a Docker container in our machine:

```
$ docker run -it -p 18080:8080 multiplication:1.0.0
```

The `run` command in Docker pulls the image if it's not available locally and executes it on the Docker platform as a container. The `-it` flags are used to attach to the container's terminal, so we can see the output from our command line and also stop the container with a `Ctrl+C` signal. The `-p` option is to expose the internal port 8080 so it's accessible from the host in port 18080. These are just a few options that we can use when running containers; you can see all of them by using `docker run --help` from the command line.

When we start this container, it'll run on top of the Docker platform. If you're running a Linux OS, the containers will use the host's native virtualization capabilities. When running on Windows or Mac, the Docker platform sets up a Linux virtualization layer in between, which may use the native support from these operating systems if they're available.

Unfortunately, our container doesn't work properly. It can't connect to RabbitMQ or Consul, even if we have them up and running in the *Docker's host machine* (our computer). Listing 8-55 shows an extract of these errors from the container logs. Remember that, by default, Spring Boot tries to find the RabbitMQ host on `localhost`, same as for Consul. In a container, `localhost` refers to the own container, and there is nothing else there but the Spring Boot app. Moreover, containers are isolated units running on a Docker platform network, so they should not connect to services running on the host anyway.

Listing 8-55. The Multiplication Container Can't Reach Consul at Localhost

```
2020-08-29 10:03:44.565 ERROR [,,,] 1 --- [           main] o.s.c.c.c.ConsulProper
tySourceLocator      : Fail fast is set and there was an error reading configuration
from consul.

2020-08-29 10:03:45.572 ERROR [,,,] 1 --- [           main] o.s.c.c.c.ConsulProper
tySourceLocator      : Fail fast is set and there was an error reading configuration
from consul.
```

```
2020-08-29 10:03:46.675 ERROR [,,,] 1 --- [           main] o.s.c.c.c.ConsulProper
tySourceLocator      : Fail fast is set and there was an error reading configuration
from consul.
[...]
```

To properly set up our back-end system to run in Docker, we have to deploy RabbitMQ and Consul as containers and connect all these different instances between them using Docker networking. See Figure 8-36.

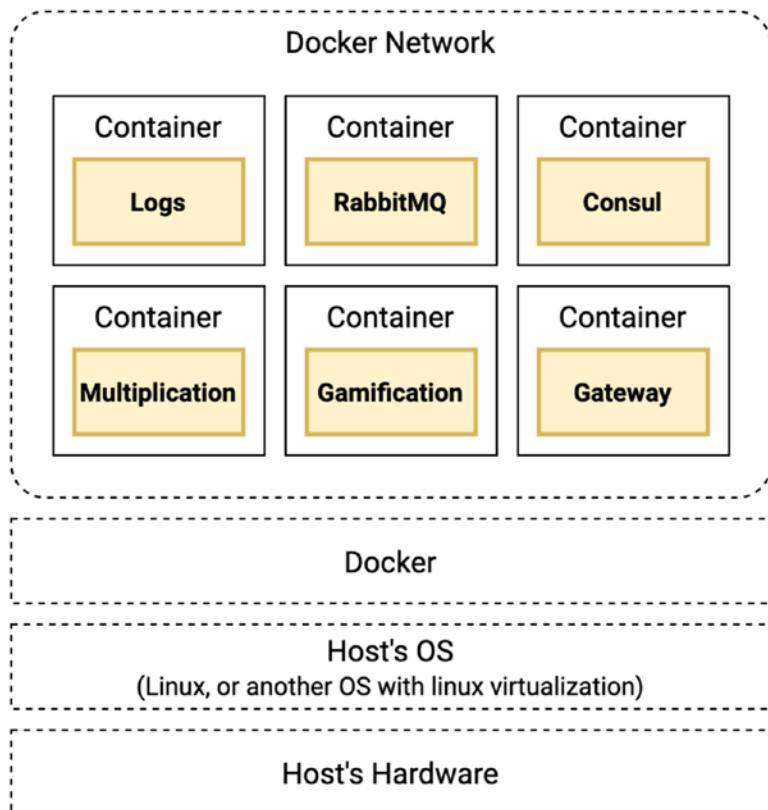


Figure 8-36. Our back end in Docker containers

Before learning how to accomplish that, let's explore how Spring Boot can help us build Docker images, so we don't need to prepare the Dockerfile ourselves.

Spring Boot and Buildpacks

Since version 2.3.0, Spring Boot's Maven and Gradle plugins have the option to build Open Container Initiative (OCI) images using Cloud Native Buildpacks (<https://tpd.io/buildpacks>), a project that aims to help package your applications for deploying them to any cloud provider. You can run the resulting images in Docker and also in other container platforms.

A nice feature of the Buildpacks plugin is that it prepares a plan based on your project's Maven configuration, and then it packages a Docker image that is ready to be deployed. Besides, it structures the image in *layers* in a way so they can be reused by future versions of your application and even by other microservice images built with this same tool (e.g., layers with all the Spring Boot core libraries). That contributes to faster testing and deployment.

We can see Buildpacks in action if we run the `build-image` goal from the command line, for example from the Gamification's project folder:

```
gamification $ ./mvnw spring-boot:build-image
```

You should see some extra logs from the Maven plugin, which is now downloading some required images and building the application image. If everything goes well, you should see this line by the end:

```
[INFO] Successfully built image 'docker.io/library/gamification:0.0.1-SNAPSHOT'
```

The Docker tag is set to the Maven artifact's name and version we specified in the `pom.xml` file: `gamification:0.0.1-SNAPSHOT`. The prefix `docker.io/library/` is the default for all public Docker images. There are multiple options we could customize for this plugin, you can check the reference docs (<https://tpd.io/buildpack-doc>) for all the details.

The same as we ran a container before for an image we built ourselves, we could do that now for this new image generated by the Spring Boot's Maven plugin:

```
$ docker run -it -p 18081:8081 gamification:0.0.1-SNAPSHOT
```

Unsurprisingly, the container will also output the same errors. Remember that the application can't connect to RabbitMQ and Consul, and it requires both services to start properly. We'll fix that soon.

For your own projects, you should consider the pros and cons of using Cloud Native Buildpacks versus maintaining your own Docker files. If you plan to use these standard OCI images to deploy to a public cloud that supports them, it might be a good idea

since you can save a lot of time. Buildpacks also takes care of organizing your images in reusable layers, so you can avoid doing that yourself. Besides, you can customize the base builder image used by the plugin, so you have some flexibility to customize the process. However, in case you want to be fully in control of what you're building and the tools and files you want to include in your images, it might be better to define the Dockerfile instructions yourself. As we saw before, it's not hard for a basic setup.

Running Our System in Docker

Let's build or find a Docker image for each component in our system, so we can deploy it as a group of containers.

- *Multiplication, Gamification, Gateway, and Logs microservices:* We'll use the Spring Boot Maven plugin with Buildpacks to generate these Docker images.
- *RabbitMQ:* We can run a container using the official RabbitMQ image version that contains the management plugin (UI): `rabbitmq:3-management` (see [Docker Hub](#)).
- *Consul:* There is an official Docker image too. We'll use the tag `consul:1.7.2` from Docker Hub (<https://tpd.io/consul-docker>). Additionally, we'll run a second container to load some configuration as key/value pairs for centralized configuration. More details will be given in its specific section.
- *Frontend:* If we want to deploy the complete system in Docker, we'll also need a web server to host the generated HTML/JavaScript files from the React build. We can use a lightweight static server like Nginx, with its official Docker image `nginx:1.19` (see Docker Hub, <https://tpd.io/nginx-docker>). In this case, we'll build our own image with `nginx` as base since we need to copy the generated files too.

Therefore, our plan requires building six different Docker images and using two public ones. See Figure 8-37.

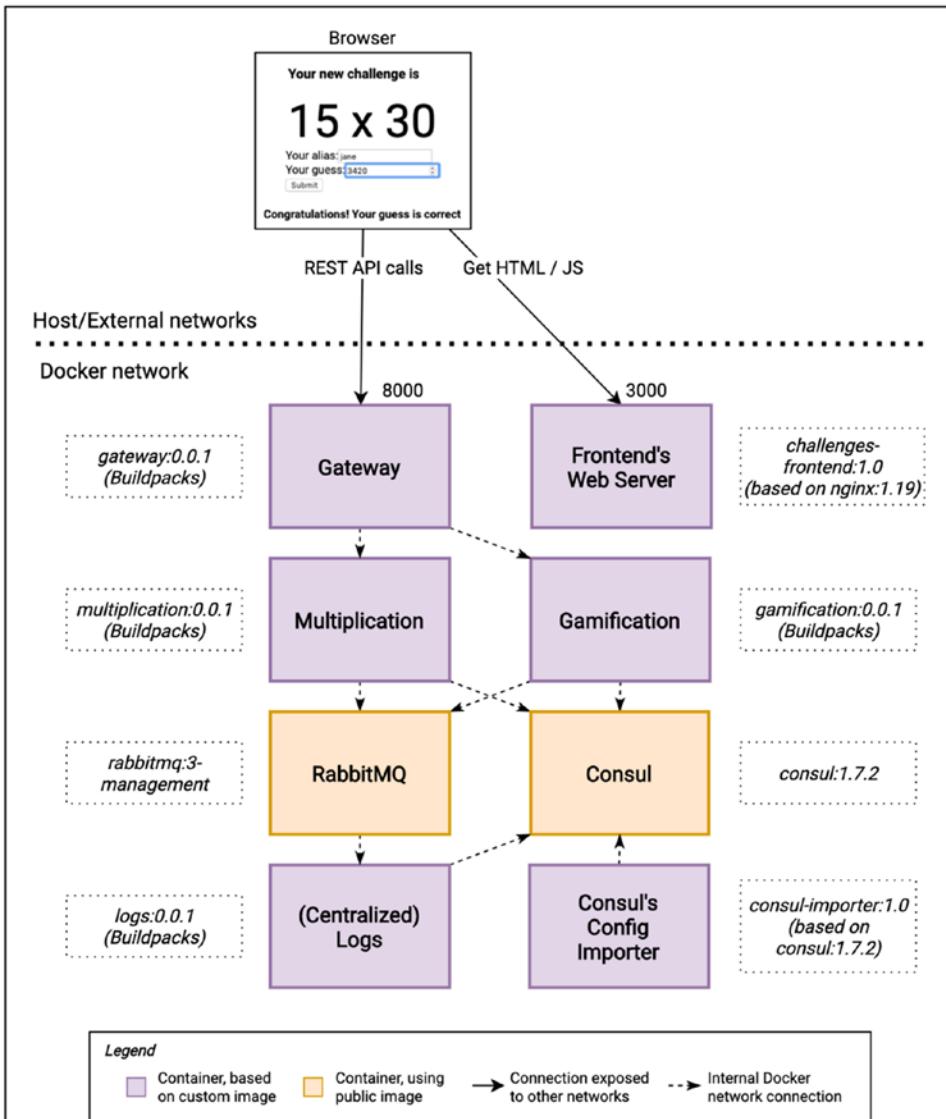


Figure 8-37. High-level overview: containerized system

Dockerizing Microservices

First, let's build all the images for the Spring Boot applications. From each of the project folders, we need to run this command:

```
$ ./mvnw spring-boot:build-image
```

Note that Docker has to be running locally, same as Consul and RabbitMQ for the tests to pass. Once you generate all images, you can verify they're all available in Docker by running the `docker images` command. See Listing 8-56.

Listing 8-56. Listing Docker Images Generated with Cloud Native Buildpacks

```
$ docker images
REPOSITORY      TAG          IMAGE ID   CREATED        SIZE
logs            0.0.1-SNAPSHOT 2fae1d82cd5d 40 years ago  311MB
gamification   0.0.1-SNAPSHOT 5552940b9bfd 40 years ago  333MB
multiplication  0.0.1-SNAPSHOT 05a4d852fa2d 40 years ago  333MB
gateway         0.0.1-SNAPSHOT d50be5ba137a 40 years ago  313MB
```

As you can see, images are generated with an old date. This is a feature from Buildpacks to make builds reproducible: every time you build this image, they get the same creation date, and they're also located at the end of the list.

Dockerizing the UI

The next step is to create a Dockerfile in the `challenges-frontend` folder, the root directory of our React application. The only two instructions we need are the base image (Nginx) and a `COPY` command to put all HTML/JavaScript files inside the image. We copy them in the folder that the Nginx web server uses by default to serve contents. See Listing 8-57.

Listing 8-57. A Simple Dockerfile to Create an Image for the front End's Web Server

```
FROM nginx:1.19
COPY build /usr/share/nginx/html/
```

Before creating the Docker image, let's make sure we generate the latest artifacts for the front end. To compile the React project, we have to execute the following:

```
challenges-frontend $ npm run build
```

Once the build folder has been generated, we can create our Docker image. We'll assign a name and a tag with the `-t` flag, and we use `.` to indicate that the Dockerfile is located in the current folder.

```
challenges-frontend $ docker build -t challenges-frontend:1.0 .
```

Dockerizing the Configuration Importer

Now, let's also prepare a Docker image to load some predefined centralized configuration. We'll have a Consul container running the server, which can use the official image directly. Our plan is to run an extra container to execute the Consul CLI to load some KV data: a docker profile. This way, we can use this preloaded profile configuration when running our microservices in Docker, since they require a different RabbitMQ host parameter, as an example.

To get the configuration we want to load in a file format, we can create it in our local Consul server and export it via CLI command. We use the UI to create the config root, and a subfolder named `defaults,docker`. Inside, we create a key named `application.yml` with the configuration shown in Listing 8-58. This configuration does the following:

- Sets the host for RabbitMQ to `rabbitmq`, which overrides the default `localhost`. Later, we'll make sure the message broker's container is available at that address.
- Overrides the instance identifier assigned to the running service to be used at the service registry. The default Spring Consul configuration concatenates the application name with the port number, but that approach will be no longer valid with containers. When we run multiple instances of the same service in Docker (as containers), they all use the same internal port, so they would end up with the same identifier. To solve this issue, we can use as a suffix a random integer. Spring Boot has support for that via the special `random` property notation (see the documentation at <https://tpd.io/random-properties> for more details).

Listing 8-58. The YAML Configuration to Connect Our Apps in Docker Containers to RabbitMQ

```
spring:  
  rabbitmq:  
    host: rabbitmq  
  cloud:  
    consul:  
      discovery:  
        instance-id: ${spring.application.name}-${random.int(1000)}
```

Figure 8-38 shows this content added via Consul UI.

The screenshot shows the Consul UI interface. At the top, there is a navigation bar with tabs: dc1, Services, Nodes, Key/Value (which is selected and highlighted in dark blue), ACL, Intentions, Documentation, and Settings. Below the navigation bar, the URL is shown as < Key / Values < config < defaults,docker. The main content area has a title "application.yml". Underneath the title, there is a "Value" section containing the YAML configuration from Listing 8-58. To the right of the "Value" section is a "Code" button with a toggle switch. At the bottom of the "Value" section, there is a "YAML" label and a dropdown arrow. At the very bottom of the screen, there are three buttons: "Save" (blue background), "Cancel changes" (white background), and "Delete" (red border).

Figure 8-38. Consul UI: prepare configuration for export

The next step is to use a different terminal to export the configuration to a file. To do that, execute the following:

```
$ consul kv export config/ > consul-kv-docker.json
```

Now, let's create a new folder named docker in the root of the workspace to place all our Docker configuration. Inside, we create a subfolder called consul. The JSON file generated with the previous command should be copied there. Then, add a new Dockerfile with the instructions in Listing 8-59.

Listing 8-59. Dockerfile Contents for the Consul Configuration Loader

```
FROM consul:1.7.2
COPY ./consul-kv-docker.json /usr/src/consul/
WORKDIR /usr/src/consul
ENV CONSUL_HTTP_ADDR=consul:8500
ENTRYPOINT until consul kv import @consul-kv-docker.json; do echo "Waiting for Consul"; sleep 2; done
```

See Listing 8-60 with the resulting file structure for the docker folder.

Listing 8-60. Creating the Consul Docker Configuration in a Separate Folder

```
+-- docker
|   +- consul
|   |   \- consul-kv-docker.json
|   \- Dockerfile
```

The Dockerfile steps in Listing 8-59 use Consul as the base image, so the CLI tool is available. The JSON file is copied inside the image, and the working directory is set to the same location as the file. Then, the ENV instruction sets a new environment variable for Consul CLI to use a remote host to reach the server, in this case located at consul:8500. That'll be the Consul server container (we'll see soon how the host gets the consul name). Finally, the ENTRYPOINT for this container (the command to run when it starts) is an inline shell script following the pattern `until [command]; do ...; sleep 2; done`. This script runs a command until it succeeds, with a period in between retries of two seconds. The main command is `consul kv import @consul-kv-docker.json`, which imports the contents of the file to the KV storage. We need to execute this in a loop because the Consul server might not have started yet when this Consul Configuration Importer runs.

To have the importer image available in the registry, we have to build it and give it a name.

```
docker/consul$ docker build -t consul-importer:1.0 .
```

We'll detail soon how to run this importer in Docker to load the configuration into Consul.

Docker Compose

Once we have built all images, we need to run our system as a set of containers, so it's time to learn how we can start all these containers together and communicate them.

We could use individual Docker commands to start all the required containers and set up the network for them to connect to each other. However, if we want to tell other people how to start our system, we need to pass them a script or documentation with all these commands and instructions. Luckily, there is a better way in Docker to group container configuration and deployment instructions: Docker Compose.

With Compose, we use a YAML file to define applications that are based on multiple containers. Then, we use the command `docker-compose` to run all services. Docker Compose is installed by default with the Windows and Mac versions of Docker Desktop. If you're running Linux or you can't find it in your Docker distribution, follow the instructions on the Docker website (<https://tpd.io/compose-install>) to install it.

As a first example, see Listing 8-61 for the YAML definition of the RabbitMQ and Consul services we need to run as containers in our system. This YAML definition has to be added to a new file `docker-compose.yml` that we can create inside the existent `docker` folder. We'll use version 3 of the Compose syntax; the full reference is available at <https://tpd.io/compose3>. Keep reading for the high-level details on how this syntax works.

Listing 8-61. A First Version of the `docker-compose.yml` File with RabbitMQ and Consul

```
version: "3"

services:
  consul-dev:
    image: consul:1.7.2
    container_name: consul
```

```

# The UDP port 8600 is used by Consul nodes to talk to each other, so
# it's good to add it here even though we're using a single-node setup.
ports:
  - '8500:8500'
  - '8600:8600/udp'
command: 'agent -dev -node=learnmicro -client=0.0.0.0 -log-level=INFO'
networks:
  - microservices

rabbitmq-dev:
image: rabbitmq:3-management
container_name: rabbitmq
ports:
  - '5672:5672'
  - '15672:15672'
networks:
  - microservices

networks:
  microservices:
    driver: bridge

```

Within the services section, we define two of them: `consul-dev` and `rabbitmq-dev`. We can use any name for our services, so we're adding the `-dev` suffix to these two to indicate that we're running them both in development mode (stand-alone nodes without a cluster). These two services use Docker images that we didn't create, but they are available in the Docker Hub as public images. They'll be pulled the first time we run the containers. If we don't specify the `command` to start the container, the default one from the image will be used. The default command can be specified in the Dockerfile used to build the image. That's what happens with the RabbitMQ service, which starts the server by default. For the Consul image, we define our own command, which is similar to the one we've been using so far. The differences are that it also includes a flag to reduce the log level, and the `client` parameter, which is required for the agent to work in the Docker network. These instructions are available on the Docker image's documentation (<https://tpd.io/consul-docker>).

Both services define a `container_name` parameter. This is useful because it sets the DNS name of the container, so other containers can find it by this alias. In our case, that means the applications can connect to the RabbitMQ server using the address `rabbitmq:5672` instead of the default one, `localhost:5672` (which now points to the

same container as we saw in previous sections). The `ports` parameter in every service allows us to expose ports to the host system in the format `host-port:container-port`. We include here the standard ones that both servers use, so we can still access them from our desktop (e.g., to use their UI tools on ports 8500 and 15672, respectively). Note that we're mapping to the same ports in the host, which means we can't have the RabbitMQ and Consul server processes running locally (as we've been doing until now) at the same time because that would cause a port conflict.

In this file, we also define a network of type `bridge` with name `microservices`. This driver type is the default one, used to connect stand-alone containers. Then, we use the parameter `networks` in each service definition to set the `microservices` network as the one they have access to. In practice, this means these services can connect to each other because they belong to the same network. The Docker network is isolated from the host network, so we can't access any service except the ones we exposed explicitly with the `ports` parameter. This is great since it's one of the good practices we were missing when we introduced the gateway pattern.

Now, we can use this new `docker-compose.yml` file to run both Consul and RabbitMQ Docker containers. We just need to execute the `docker-compose` command from a terminal:

```
docker $ docker-compose up
```

Docker Compose takes `docker-compose.yml` automatically without specifying the name because that's the default file name that it expects. The output for all containers is *attached* to the current terminal and containers. If we would want to run them in the background as *daemon* processes, we just need to add the `-d` flag to the command. In our case, we'll see all the logs together in the terminal output, for both `consul` and `rabbitmq` containers. See Listing 8-62 for an example.

Listing 8-62. Docker Compose Logs, Showing the Initialization of Both Containers

```
Creating network "docker_microservices" with driver "bridge"
Creating consul    ... done
Creating rabbitmq ... done
Attaching to consul, rabbitmq
consul      | ==> Starting Consul agent...
consul      |          Version: 'v1.7.2'
consul      |          Node ID: 'a69c4c04-d1e7-6bdc-5903-c63934f01f6e'
```

CHAPTER 8 COMMON PATTERNS IN MICROSERVICE ARCHITECTURES

```
consul |           Node name: 'learnmicro'  
consul |           Datacenter: 'dc1' (Segment: '<all>')  
consul |               Server: true (Bootstrap: false)  
consul |               Client Addr: [0.0.0.0] (HTTP: 8500, HTTPS: -1, gRPC:  
|                   8502, DNS: 8600)  
consul |               Cluster Addr: 127.0.0.1 (LAN: 8301, WAN: 8302)  
consul |               Encrypt: Gossip: false, TLS-Outgoing: false, TLS-  
|                   Incoming: false, Auto-Encrypt-TLS: false  
consul |  
consul | ==> Log data will now stream in as it occurs:  
[...]  
rabbitmq | 2020-07-30 05:36:28.785 [info] <0.8.0> Feature flags: list of  
|       feature flags found:  
rabbitmq | 2020-07-30 05:36:28.785 [info] <0.8.0> Feature flags: [ ]  
|       drop_unroutable_metric  
rabbitmq | 2020-07-30 05:36:28.785 [info] <0.8.0> Feature flags: [ ]  
|       empty_basic_get_metric  
rabbitmq | 2020-07-30 05:36:28.785 [info] <0.8.0> Feature flags: [ ]  
|       implicit_default_bindings  
rabbitmq | 2020-07-30 05:36:28.785 [info] <0.8.0> Feature flags: [ ]  
|       quorum_queue  
rabbitmq | 2020-07-30 05:36:28.786 [info] <0.8.0> Feature flags: [ ]  
|       virtual_host_metadata  
rabbitmq | 2020-07-30 05:36:28.786 [info] <0.8.0> Feature flags: feature  
|       flag states written to disk: yes  
rabbitmq | 2020-07-30 05:36:28.830 [info] <0.268.0> ra: meta data store  
|       initialised. 0 record(s) recovered  
rabbitmq | 2020-07-30 05:36:28.831 [info] <0.273.0> WAL: recovering []  
rabbitmq | 2020-07-30 05:36:28.833 [info] <0.277.0>  
rabbitmq |   Starting RabbitMQ 3.8.2 on Erlang 22.2.8  
[...]
```

We can also verify how we can access the Consul UI from your browser at `localhost:8500`. This time, the website is served from the container. It works exactly the same because we exposed the port to the same host's port, and it's being redirected by Docker.

To stop these containers, we can press Ctrl+C, but that could make Docker keep some state in between executions. To properly shut them down and remove any potential data they created in Docker *volumes* (the units that containers define to store data), we can run the command in Listing 8-63 from a different terminal.

Listing 8-63. Stopping Docker Containers and Removing Volumes with Docker Compose

```
docker $ docker-compose down -v
Stopping consul    ... done
Stopping rabbitmq ... done
Removing consul    ... done
Removing rabbitmq ... done
Removing network docker_default
WARNING: Network docker_default not found.
Removing network docker_microservices
```

Our next step is to add the image we created for loading configuration into Consul KV, the `consul-importer`, to the Docker Compose file. See Listing 8-64.

Listing 8-64. Adding the Consul Importer Image to the `docker-compose.yml` File

```
version: "3"

services:
  consul-importer:
    image: consul-importer:1.0
    depends_on:
      - consul-dev
    networks:
      - microservices
  consul-dev:
    # ... same as before
  rabbitmq-dev:
    # ... same as before

networks:
  microservices:
    driver: bridge
```

This time, the image `consul-importer:1.0` is not a public one; it's not available in the Docker Hub. However, it's available in the local Docker registry because we built it earlier, so Docker can find it by its name and tag we defined earlier.

We can establish dependencies in the compose file with the parameter `depends_on`. Here, we use it to make this container start after the `consul-dev` container, which runs the Consul server. Anyway, that doesn't guarantee that the server is ready by the time the `consul-importer` runs. The reason is that Docker knows only when the container has started but doesn't know when the consul server has booted up and is ready to accept requests. That's the reason why we added a script to our importer image, which retries the import until it succeeds (see Listing 8-59).

When you run `docker-compose up` again with this new configuration, you'll see the output from this new container too. Eventually, you should see the lines that load the configuration, and then Docker will inform that this container exited successfully (with code 0). See Listing 8-65.

Listing 8-65. Running docker-compose for the Second Time to See the Importer's Logs

```
docker $ docker-compose up
[...]
consul-importer_1 | Imported: config/
consul-importer_1 | Imported: config/defaults,docker/
consul-importer_1 | Imported: config/defaults,docker/application.yml
docker_consul-importer_1 exited with code 0
[...]
```

The new container runs as a function, not as a continuously running service. This is because we replaced the default command in the `consul` image, which is defined in their internal Dockerfile to run the server as a process, by a command that simply loads configuration and then finishes (it's not an indefinitely running process). Docker knows that the container has nothing more to do because the command exited, so there is no need to keep the container alive.

We can also know what are the running containers for a `docker-compose` configuration. To get this list, we can execute `docker-compose ps` from a different terminal. See Listing 8-66.

Listing 8-66. Running docker-compose ps to See the Status of the Containers

```
docker $ docker-compose ps
```

Name	Command	State	Ports
<hr/>			
consul	docker-e[...]	Up	8300/tcp, [...]
docker_consul-importer_1	/bin/sh [...]	Exit 0	
rabbitmq	docker-e[...]	Up	15671/tcp, [...]

The output (trimmed for better readability) also details the command used by the container, its state, and the exposed ports.

If we navigate with our browser to the Consul UI at `http://localhost:8500/ui`, we'll see how the configuration has been properly loaded, and we have a config entry with the nested defaults,docker subfolder and the corresponding application.yml key. See Figure 8-39. Our importer works perfectly.

The screenshot shows the Consul UI interface. At the top, there's a navigation bar with tabs for dc1, Services, Nodes, Key/Value (which is selected), ACL, Intentions, Documentation, and Settings. Below the navigation bar, the URL is shown as < Key / Values < config < defaults,docker. The main content area is titled "application.yml". It displays a YAML configuration snippet:

```

Value
1 spring:
2   rabbitmq:
3     host: rabbitmq
4   cloud:
5     consul:
6       discovery:
7         instance-id: ${spring.application.name}-${random.int(1000)}

```

On the right side of the code editor, there are "Value" and "Code" buttons, with "Code" being selected. At the bottom right, there's a "YAML" button and a small downward arrow icon.

Figure 8-39. Docker configuration inside the Consul container

Let's continue with the front-end definition in Docker Compose. This one is easy; we just need to add the image we built based on Nginx and expose port 3000 with a redirection to the internal one, which is 80 by default in the base image (as per <https://tpd.io/nginx-docker>). See Listing 8-67. You can change the exposed port but then remember to adjust the CORS configuration in the Gateway accordingly (or refactor it so it can be configured via external properties).

Listing 8-67. Adding the Web Server to the docker-compose.yml File

```
version: "3"

services:
  frontend:
    image: challenges-frontend:1.0
    ports:
      - '3000:80'
  consul-importer:
    # ... same as before
  consul-dev:
    # ... same as before
  rabbitmq-dev:
    # ... same as before

networks:
  microservices:
    driver: bridge
```

To make the complete system work, we need to add the Spring Boot microservices to the Docker Compose file. We'll configure them to use the same network we created. Each of these containers will need to reach the `consul` and `rabbitmq` containers to work properly. We'll use two different strategies for that.

- For the Consul setup, the centralized configuration feature in Spring requires that the service knows where to find the server at the bootstrap phase. We need to override the property `spring.cloud.consul.host` used in the local `bootstrap.yml` and point it to the `consul` container. We'll do this via environment variables. In Spring Boot, if you set an environment variable that matches an existing property or

that follows a certain naming convention (like SPRING_CLOUD_CONSUL_HOST), its value overrides the local configuration. For more details, see <https://tpd.io/binding-vars> in Spring Boot docs.

- For RabbitMQ configuration, we'll use the docker profile. Given that the microservices will connect to Consul and the configuration server has a preloaded entry for defaults, docker, all of them will use the properties in there. Remember that we changed the RabbitMQ host in that profile to rabbitmq, the DNS name of the container. To activate the docker profile in each microservice, we'll use the Spring Boot property to enable profiles, passed via an environment variable: SPRING_PROFILES_ACTIVE=docker.

Besides, these are some extra considerations for the configuration of the Spring Boot containers in Compose:

- We don't want to expose the back-end services directly to the host except for the Gateway service, on localhost:8000. Therefore, we won't add the ports section to Multiplication, Gamification, and Logs.
- Additionally, we'll use the depends_on parameter for the back-end containers to wait until the consul-importer runs, so the Consul configuration for the docker profile will be available by the time the Spring Boot apps start.
- We'll also include rabbitmq as a dependency for these services, but remember that this doesn't guarantee that the RabbitMQ server is ready before our applications boot up. Docker only verifies that the container has started. Luckily, Spring Boot retries to connect by default to the server as a resilience technique, so eventually, the system will become stable.

See Listing 8-68 for the complete Docker Compose configuration required to start our system.

Listing 8-68. docker-compose.yml File with Everything Needed to Run Our Complete System

```
version: "3"

services:
  frontend:
    image: challenges-frontend:1.0
    ports:
      - '3000:80'
  multiplication:
    image: multiplication:0.0.1-SNAPSHOT
    environment:
      - SPRING_PROFILES_ACTIVE=docker
      - SPRING_CLOUD_CONSUL_HOST=consul
  depends_on:
    - rabbitmq-dev
    - consul-importer
  networks:
    - microservices
  gamification:
    image: gamification:0.0.1-SNAPSHOT
    environment:
      - SPRING_PROFILES_ACTIVE=docker
      - SPRING_CLOUD_CONSUL_HOST=consul
  depends_on:
    - rabbitmq-dev
    - consul-importer
  networks:
    - microservices
  gateway:
    image: gateway:0.0.1-SNAPSHOT
    ports:
      - '8000:8000'
    environment:
      - SPRING_PROFILES_ACTIVE=docker
      - SPRING_CLOUD_CONSUL_HOST=consul
  depends_on:
```

```
- rabbitmq-dev
- consul-importer
networks:
- microservices
logs:
image: logs:0.0.1-SNAPSHOT
environment:
- SPRING_PROFILES_ACTIVE=docker
- SPRING_CLOUD_CONSUL_HOST=consul
depends_on:
- rabbitmq-dev
- consul-importer
networks:
- microservices
consul-importer:
image: consul-importer:1.0
depends_on:
- consul-dev
networks:
- microservices
consul-dev:
image: consul:1.7.2
container_name: consul
ports:
- '8500:8500'
- '8600:8600/udp'
command: 'agent -dev -node=learnmicro -client=0.0.0.0 -log-level=INFO'
networks:
- microservices
rabbitmq-dev:
image: rabbitmq:3-management
container_name: rabbitmq
ports:
- '5672:5672'
- '15672:15672'
networks:
- microservices
```

networks:

```
microservices:
  driver: bridge
```

It's time to test our complete system running as Docker containers. Like before, we run the `docker-compose up` command. We'll see many logs in the output, produced by multiple services that are starting simultaneously, or right after the ones defined as dependencies.

The first thing you might notice is that some back-end services throw exceptions when trying to connect to RabbitMQ. This is an expected situation. As mentioned, RabbitMQ may take longer to start than the microservice applications. This should fix by itself after the `rabbitmq` container becomes ready.

You could also experience errors caused by not having enough memory or CPU in your system to run all the containers together. This is not exceptional since each microservice container can take up to 1GB of RAM. If you can't run all these containers, I hope the book explanations still help you understand how everything works together.

To know the status of the system, we can use the aggregated logs provided by Docker (the attached output) or the output from the logs container. To try the second option, we can use another Docker command from a different terminal, `docker-compose logs [container_name]`. See Listing 8-69. Note that our service name is `logs`, which explains the word repetition.

Listing 8-69. Checking the Logs of the Logs Container

```
docker $ docker-compose logs logs
[...]
logs_1           | [gamification    ] [aadd7c03a8b161da,34c00bc3e3197ff2]
INFO 07:24:52.386 [main] o.s.d.r.c.DeferredRepositoryInitializationListener -
Triggering deferred initialization of Spring Data repositories?
logs_1           | [multiplication ] [33284735df2b2be1,bc998f237af7bebb]
INFO 07:24:52.396 [main] o.s.d.r.c.DeferredRepositoryInitializationListener -
Triggering deferred initialization of Spring Data repositories?
logs_1           | [multiplication ] [b87fc916703f6b56,fd729db4060c1c74]
INFO 07:24:52.723 [main] o.s.d.r.c.DeferredRepositoryInitializationListener -
Spring Data repositories initialized!
logs_1           | [multiplication ] [97f86da754679510,9fa61b768e26a
eb5] INFO 07:24:52.760 [main] m.b.m.MultiplicationApplication - Started
MultiplicationApplication in 44.974 seconds (JVM running for 47.993)
```

```

logs_1           | [gamification ] [5ec42be452ce0e04,03dfa6fc3656b7fe]
INFO 07:24:53.017 [main] o.s.d.r.c.DeferredRepositoryInitializationListener -
Spring Data repositories initialized!
logs_1           | [gamification ] [f90c5542963e7eea,a9f52df128ac5
c7d] INFO 07:24:53.053 [main] m.b.g.GamificationApplication - Started
GamificationApplication in 45.368 seconds (JVM running for 48.286)
logs_1           | [gateway       ] [59c9f14c24b84b32,36219539a1a0d01b]
WARN 07:24:53.762 [boundedElastic-1] o.s.c.l.core.RoundRobinLoadBalancer - No
servers available for service: gamification

```

Additionally, you can also monitor the service statuses by checking the Consul UI's service list, available at localhost:8500. There, you'll see if the health checks are passing, which means the services are already serving and connected to RabbitMQ. See Figure 8-40.

The screenshot shows the Consul UI interface. At the top, there is a navigation bar with tabs for Services (which is selected), Nodes, Key/Value, ACL, Intentions, and links to Documentation and Settings. Below the navigation bar, the title "Services" is displayed with a note "4 total". A search bar contains the query "service:name tag:name status:critical search-term". The main table lists four services:

Service	Health Checks	Tags
consul	✓ 1	
gamification	✓ 2	secure=false
gateway	✓ 2	secure=false
multiplication	✓ 2	secure=false

At the bottom of the page, there is a footer with the HashiCorp logo, copyright information (© 2020 HashiCorp), the version (Consul 1.7.2), and a link to Documentation.

Figure 8-40. Consul UI: checking containers health

If you click one of the services (e.g., gamification), you'll see how the host address is now the container's address within the docker network. See Figure 8-41. This is an alternative to the container name for services to connect to each other. Actually, this dynamic host address registration in Consul allows us to have multiple instances of a given service. If we would use a `container_name` parameter, we couldn't start more than one instance since their addresses would conflict.

The applications use in this case the Docker's host address because Spring Cloud detects when an application is running on a Docker container. Then, the Consul Discovery libraries use this value at registration time.

ID	Node	Address	Node Checks	Service Checks
gamification-496	learnmicro	c5910ebfd96a:8081	✓ 1	✓ 1

Figure 8-41. Consul UI: Docker container address

Once containers are green, we can navigate with our browser to `localhost:3000` and start playing with our application. It works the same as before. When we solve a challenge, we'll see in the logs how the event is consumed by gamification, which adds the score and badges. The front end is accessing via the gateway, the only microservice exposed to the host. See Figure 8-42.

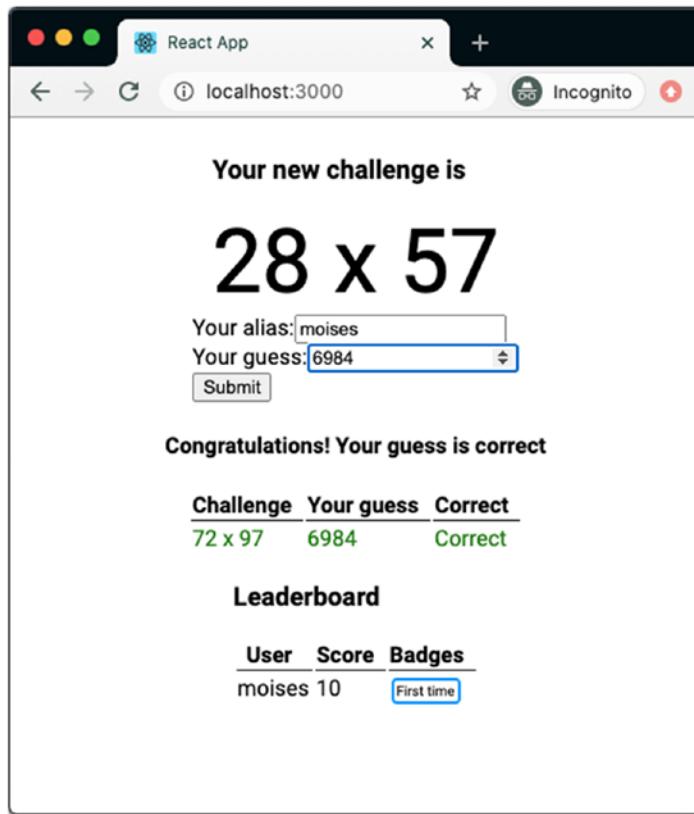


Figure 8-42. Application running on Docker

We didn't add any persistence, so all the data will be gone when we bring the containers down. If you want to extend your knowledge of Docker and Docker Compose, consider adding *volumes* to store the DB files (see <https://tpd.io/compose-volumes>). Also, don't forget to remove the `-v` flag when executing `docker-compose down`, so the volumes are kept between executions.

Scaling Up the System with Docker

With Docker Compose, we can also scale services up and down with a single command. First, let's boot up the system as it was before. If you brought it down already, execute the following:

```
docker$ docker-compose up
```

Then, from a different terminal, we run again the command with the scale parameter, indicating the service name and the number of instances we want to get. We can use the parameter multiple times within a single command.

```
docker$ docker-compose up --scale multiplication=2 --scale gamification=2
```

Now, check this new terminal's logs to see how Docker Compose boots up an extra instance for the multiplication and gamification services. You can also verify this in Consul. See Figure 8-43.

ID	Node	Address	Node Checks	Service Checks
gamification-470	learnmicro	bf054157ab0f:8081	✓ 1	✓ 1
gamification-742	learnmicro	6fc197469219:8081	✓ 1	✓ 1

Figure 8-43. Consul UI: two containers for Gamification

Thanks to Consul Discovery, our gateway pattern, the Spring Cloud load balancer, and RabbitMQ consumer's load balancing, we'll get again our system to properly balance the load across multiple instances. You can verify that either by solving a few challenges from the UI or by executing directly some HTTP calls to the Gateway service. If you go for the terminal option, you can run this HTTPie command multiple times:

```
$ http POST :8000/attempts factorA=15 factorB=20 userAlias=test-docker-containers guess=300
```

In the logs, you'll see how both `multiplication_1` and `multiplication_2` handle requests from the API. The same happens for `gamification_1` and `gamification_2`, which also take different messages from the broker's queue. See Listing 8-70.

Listing 8-70. Scalability in Action with Docker Containers

```

multiplication_1 | 2020-07-30 09:48:34.559 INFO [,85acf6d095516f55,956486d186
a612dd,true] 1 --- [nio-8080-exec-8] m.b.m.c.ChallengeAttemptController      :
Received new attempt from test-docker-containers
logs_1           | [multiplication ] [85acf6d095516f55,31829523bbc1d6ea]
INFO 09:48:34.559 [http-nio-8080-exec-8] m.b.m.c.ChallengeAttemptController -
Received new attempt from test-docker-containers
gamification_1   | 2020-07-30 09:48:34.570 INFO [,85acf6d095516f55,44508dd6f0
9c83ba,true] 1 --- [ntContainer#0-1] m.b.gamification.game.GameEventHandler    :
Challenge Solved Event received: 7
gamification_1   | 2020-07-30 09:48:34.572 INFO [,85acf6d095516f55,44508dd6f09c
83ba,true] 1 --- [ntContainer#0-1] m.b.gamification.game.GameServiceImpl     : User
test-docker-containers scored 10 points for attempt id 7
logs_1           | [gamification ] [85acf6d095516f55,8bdd9b6
febc1eda8] INFO 09:48:34.570 [org.springframework.amqp.rabbit.
RabbitListenerEndpointContainer#0-1] m.b.g.game.GameEventHandler - Challenge
Solved Event received: 7
logs_1           | [gamification ] [85acf6d095516f55,247a930
d09b3b7e5] INFO 09:48:34.572 [org.springframework.amqp.rabbit.
RabbitListenerEndpointContainer#0-1] m.b.g.game.GameServiceImpl - User test-
docker-containers scored 10 points for attempt id 7
multiplication_2 | 2020-07-30 09:48:35.332 INFO [,fa0177a130683114,f2c2809dd9
a6bc44,true] 1 --- [nio-8080-exec-1] m.b.m.c.ChallengeAttemptController      :
Received new attempt from test-docker-containers
logs_1           | [multiplication ] [fa0177a130683114,f5b7991f5b1518a6]
INFO 09:48:35.332 [http-nio-8080-exec-1] m.b.m.c.ChallengeAttemptController -
Received new attempt from test-docker-containers
gamification_2   | 2020-07-30 09:48:35.344 INFO [,fa0177a130683114,298af219a0
741f96,true] 1 --- [ntContainer#0-1] m.b.gamification.game.GameEventHandler    :
Challenge Solved Event received: 7
gamification_2   | 2020-07-30 09:48:35.358 INFO [,fa0177a130683114,298af219a074
1f96,true] 1 --- [ntContainer#0-1] m.b.gamification.game.GameServiceImpl     : User
test-docker-containers scored 10 points for attempt id 7
logs_1           | [gamification ] [fa0177a130683114,2b9ce6c
ab6366dfb] INFO 09:48:35.344 [org.springframework.amqp.rabbit.
RabbitListenerEndpointContainer#0-1] m.b.g.game.GameEventHandler - Challenge
Solved Event received: 7

```

```
logs_1           | [gamification    ] [fa0177a130683114,536fbc8
035a2e3a2] INFO 09:48:35.358 [org.springframework.amqp.rabbit.
RabbitListenerEndpointContainer#0-1] m.b.g.game.GameServiceImpl - User test-
docker-containers scored 10 points for attempt id 7
```

Sharing Docker Images

All images we built so far are stored in our local machine. That doesn't help us achieve the "build once, deploy everywhere" strategy we aimed for. However, we're very close.

We already know Docker Hub, the public registry from where we downloaded the official images for RabbitMQ and Consul, and the base images for our microservices. Therefore, if we upload our own images there, they will be available to everybody. If you're fine with that, you can create a free account at hub.docker.com and start uploading (*pushing* in Docker terms) your custom images. In case you want to restrict access to your images, they also offer plans for setting up private repositories, hosted in their cloud. Actually, the Docker Hub is not the only option you have to store Docker images. You can also deploy your own registry following the instructions on the "Deploy a registry server" page (<https://tpd.io/own-registry>), or choose one of the online solutions offered by the different cloud vendors, like Amazon's ECR or Google Cloud's Container Registry.

In a Docker registry, you can keep multiple versions of your images using tags. For example, our Spring Boot images got the version number from the `pom.xml` file, so they got the default version created by the Initializer (e.g., `multiplication:0.0.1-SNAPSHOT`). We can keep our versioning strategy in Maven, but we could also set tags manually using the `docker tag` command. Besides, we can use multiple tags to refer to the same Docker image. A common practice is to add the tag `latest` to our Docker images to point to the most recent version of an image in the registry. See the list of available tags (<https://tpd.io/consul-tags>) for the Consul image as an example of Docker image versioning.

To connect our Docker's command-line tool with a registry, we use the `docker login` command. If we want to connect to a private host, we must add the host address. Otherwise, if we're connecting to the Hub, we can use the plain command. See Listing 8-71.

Listing 8-71. Logging In to Docker Hub

```
$ docker login
```

Login with your Docker ID to push and pull images from Docker Hub. If you don't have a Docker ID, head over to <https://hub.docker.com> to create one.

```
Username: [your username]
```

```
Password: [your password]
```

```
Login Succeeded
```

Once you're logged in, you can push images to the registry. Keep in mind that, to make it work, you have to tag them with your username as a prefix, since that's the Docker Hub's naming convention. Let's change the name of one of our images following the expected pattern. Additionally, we'll modify the version identifier to 0.0.1. In this example, the registered username is learnmicro.

```
$ docker tag multiplication:0.0.1-SNAPSHOT learnmicro/multiplication:0.0.1
```

Now, you can push this image to the registry using the `docker push` command. See Listing 8-72 for an example.

Listing 8-72. Pushing an Image to the Public Registry, the Docker Hub

```
$ docker push learnmicro/multiplication:0.0.1
```

```
The push refers to repository [docker.io/learnmicro/multiplication]
abf6a2c86136: Pushed
9474e9c2336c: Pushing
[=====>] 37.97MB/58.48MB
9474e9c2336c: Pushing
[=====>] 10.44MB/58.48MB
5cd38b221a5e: Pushed
d12f80e4be7c: Pushed
c789281314b6: Pushed
2611af6e99a7: Pushing
[=====>] 7.23MB
02a647e64beb: Pushed
1ca774f177fc: Pushed
9474e9c2336c: Pushing
[=====>] 39.05MB/58.48MB
8713409436f4: Pushing
[==>] 10.55MB/154.9MB
8713409436f4: Pushing
[==>] 11.67MB/154.9MB
```

CHAPTER 8 COMMON PATTERNS IN MICROSERVICE ARCHITECTURES

```
7fbc81c9d125: Waiting  
8713409436f4: Pushing  
[====> ] 12.78MB/154.9MB  
9474e9c2336c: Pushed  
6c918f7851dc: Pushed  
8682f9a74649: Pushed  
d3a6da143c91: Pushed  
83f4287e1f04: Pushed  
7ef368776582: Pushed  
0.0.1: digest: sha256:ef9bbbed14b5e349f1ab05cffff92e60a8a99e01c412341a3232fc93ae  
eccfdc size: 4716
```

From this moment on, anybody with access to the registry will be able to pull our image and use it as a container. If we use the hub's public registry like in our example, the image becomes publicly available. If you're curious, you can verify that this image is really online by visiting its Docker Hub's link (<https://hub.docker.com/r/learnmicro/multiplication>). See Figure 8-44.

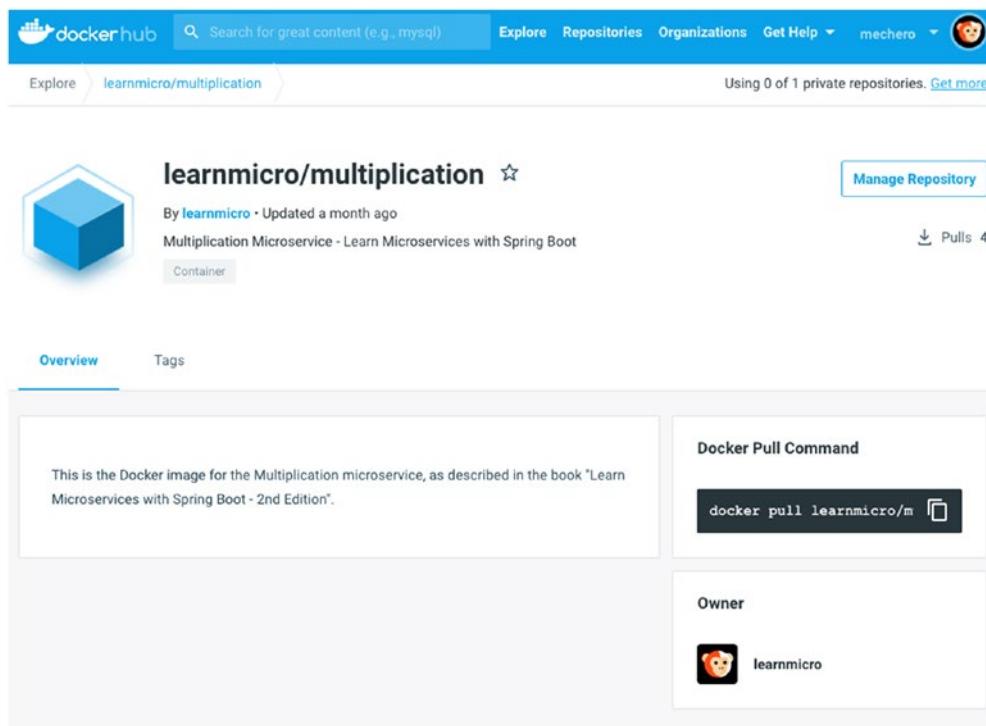


Figure 8-44. Multiplication's Docker image in the Docker Hub

Actually, all the Docker images that we described earlier are already available in the public registry under my account with the prefix `learnmicro/`. All these first versions are tagged as `0.0.1`. That makes possible for any Docker user to get a version of the complete system up and running without building anything at all. They just need to use a version of the same `docker-compose.yml` file we used in Listing 8-68, with image name replacements to point to existing images in the public registry. See Listing 8-73 for the required changes.

Listing 8-73. Changing the `docker-compose.yml` File to Use Publicly Available Images

```
version: "3"

services:
  frontend:
    image: learnmicro/challenges-frontend:0.0.1
    # ...
  multiplication:
    image: learnmicro/multiplication:0.0.1
    # ...
  gamification:
    image: learnmicro/gamification:0.0.1
    # ...
  gateway:
    image: learnmicro/gateway:0.0.1
    # ...
  logs:
    image: learnmicro/logs:0.0.1
    # ...
  consul-importer:
    image: learnmicro/consul-importer:0.0.1
    # ...
  consul-dev:
    # same as before
  rabbitmq-dev:
    # same as before

networks:
  # ...
```

We achieved our goal for this section. Deploying our application became easy, given that the only requisite is having Docker support. That opens us a lot of possibilities since most of the platforms to manage and orchestrate distributed systems support Docker container deployments. In the next section, we'll learn some basics about *platforms*.

Platforms and Cloud-Native Microservices

Throughout this chapter, we've been covering a few patterns that are the foundations of a proper microservice architecture: routing, service discovery, load balancing, health reporting, centralized logging, centralized configuration, distributed tracing, and containerization.

If we take a moment to analyze the components in our system, we'll realize how complex it's becoming to support three main functional parts: the Web UI and the Multiplication and Gamification back-end domains. Even though we employed popular implementations for many of these patterns, we still had to configure them, or even deploy some extra components, to make our system work.

Besides, we didn't cover yet the *clustering strategies*. If we deploy all our applications in a single machine, there is a high risk that something goes wrong. Ideally, we want to replicate components and spread them across multiple physical servers. Luckily for us, there are tools to manage and orchestrate the different components across a cluster of servers, either in your own hardware or in the cloud. The most popular alternatives work at the container level or the application level, and we'll describe them separately.

Container Platforms

First, let's focus on container platforms like Kubernetes, Apache Mesos, or Docker Swarm. In these platforms, we deploy containers either directly or by using wrapping structures with extra configuration intended for the specific tool. As an example, a deployment unit in Kubernetes is a *pod*, and its definition (to keep it simple, a *deployment*) may define a list of Docker containers to deploy (usually only one) and some extra parameters to set the allocated resources, connect the pod to a network, or add external configuration and storage.

Besides, these platforms usually integrate patterns that should be already familiar to us. Again, let's use Kubernetes as an example since it's one of the most popular options. This list gives a high-level perspective of some of its features:

- *Container orchestration* across multiple nodes forming a cluster.
When we deploy a unit of work in Kubernetes (pod), the platform decides where to instantiate it. If the complete node dies or we bring it down gracefully, Kubernetes finds another place to put this unit of work, based on our configuration for concurrent instances.
- *Routing*: Kubernetes uses *ingress controllers* that allow us to route traffic to the deployed services.
- *Load balancing*: All pod instances in Kubernetes are usually configured to use the same DNS address. Then, there is a component called *kube-proxy* that takes care of balancing the load across the pods. Other services just need to call a common DNS alias, e.g., `multiplication.myk8scluster.io`. This is a server-side discovery and load balancing strategy, applied per server component.
- *Self-healing*: Kubernetes uses HTTP probes to determine whether the services are alive and ready. In case they aren't, we can configure it to get rid of those *zombie* instances and start new ones to satisfy our redundancy configuration.
- *Networking*: Similarly to Docker Compose, Kubernetes uses exposed ports and provides different network topologies we can configure.
- *Centralized configuration*: The container platform offers solutions like *ConfigMaps*, so we can separate the configuration layer from the application and therefore change that per environment.

On top of that, Kubernetes has other built-in functionalities on aspects such as security, cluster administration, and distributed storage management.

Therefore, we could deploy our system in Kubernetes and benefit from all these features. Moreover, we could get rid of some patterns we built and leave these responsibilities to Kubernetes.

People who know how to configure and manage a Kubernetes cluster would probably never recommend you deploy bare containers like we did with Docker Compose; instead, they would start directly with a Kubernetes setup. However, the

extra complexity that a container orchestration platform introduces should never be underestimated. If we know the tooling very well, surely we can have our system up and running quickly. Otherwise, we'll have to dive into a lot of documentation with custom YAML syntax definitions.

In any case, I recommend you learn how one of these container platforms works and try to deploy our system in there to get into the practical aspects. They're popular in many organizations since they abstract all the infrastructure layer from the applications. Developers can focus on the process from coding to building a container, and the infrastructure team can focus on managing the Kubernetes clusters in different environments.

Application Platforms

Now, let's cover a different type of platforms: application runtime platforms. These offer an even higher level of abstraction. Basically, we can write our code, build a .jar file, and then push it directly to an environment, making it ready for use. The application platform takes care of everything else: containerizing the app (if needed), running it on a cluster node, providing load balancing and routing, securing the access, etc. These platforms can even aggregate logs, and provide other tools such as message brokers and databases-as-a-service.

On this level, we can find solutions like Heroku or CloudFoundry. There are alternatives for us to manage these platforms in our own servers, but the most popular options are the cloud-provided solutions. The reason is that we can put our product or service alive in just a few minutes, without taking care of many of the pattern implementations or infrastructure aspects.

Cloud Providers

To complete the landscape of platforms and tools, we have to mention cloud solutions such as AWS, Google, Azure, OpenShift, etc. Many of these also offer implementations for the patterns we covered in this chapter: gateway, service discovery, centralized logs, containerization, etc.

Furthermore, they usually provide a managed Kubernetes service too. That means that, if we prefer to work at a container platform level, we can do that without needing to set up this platform manually. Of course, that means we'll have to pay for this service, on top of the cloud resources we use (machine instances, storage, etc.).

See Figure 8-45 for a first example of how we could deploy a system like ours in a cloud provider. In this first case, we choose to pay only for some low-level services such as storage and virtual machines, but we set up our own installations of Kubernetes, the databases, and the Docker registry. This means we avoid paying for extra managed services, but we have to maintain all these tools ourselves.

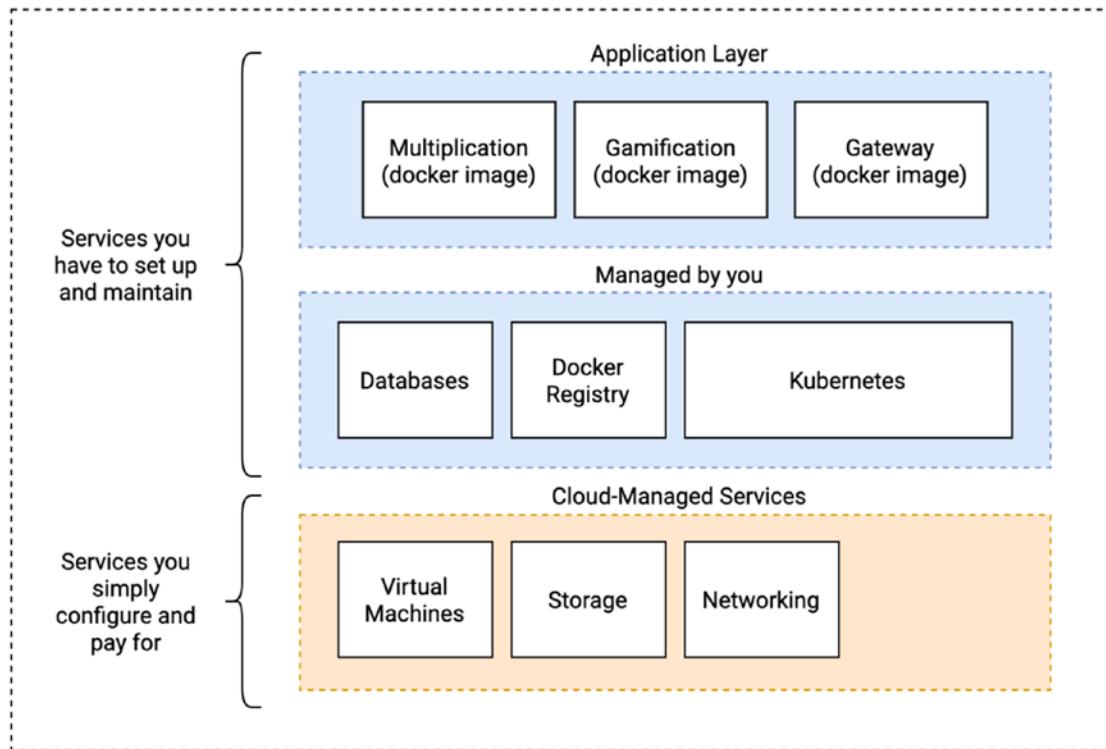


Figure 8-45. Using a cloud provider: example 1

Check now Figure 8-46 for an alternative setup. In this second case, we could use some additional managed services from the cloud provider: Kubernetes, a gateway, a Docker registry, etc. As an example, in AWS, we can use a gateway-as-a-service solution called Amazon API Gateway to route traffic directly to our containers, or we could also choose an Amazon Elastic Kubernetes Service with its own routing implementation. In any of these cases, we avoid having to implement these patterns and maintain these tools, at the expense of paying more for these cloud-managed services. Nevertheless, take into account that, in the long run, you might save money with this approach because you need people to maintain all the tools if you decide to go that way.

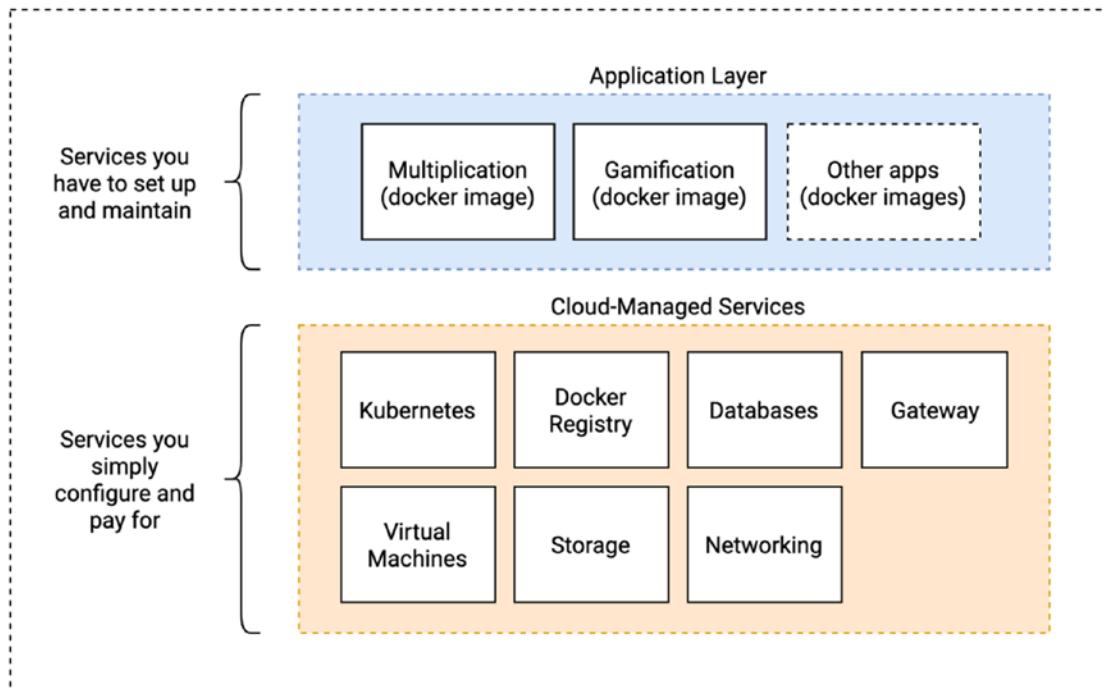


Figure 8-46. Using a cloud provider: example 2

Making a Decision

Given that there are plenty of options, we should analyze the pros and cons of each level of abstraction for our specific situation. As you can imagine, high-level abstractions are more expensive than building the solution ourselves at a lower level. On the other hand, if we choose the cheapest option, we may spend much more money on setting that up, maintaining, and evolving it. Besides, if we're planning to deploy our system to the cloud, we should compare the costs of each vendor since there might be substantial differences.

Usually, a good idea is to start a project using high-level solutions, which translates to managed services and/or application platforms. They might be more expensive and less customizable, but you can try your product or service much faster. Then, if the project goes well, you can decide to take ownership of those services if it's worth it in terms of costs.

Cloud-Native Microservices

No matter what option we choose to deploy our microservices, we know we should respect some good practices to make sure they'll work properly in the cloud (well, ideally, in any environment): data-layer isolation, stateless logic, scalability, resilience, simple logging, etc. We've been taking care of all these aspects while we were learning new topics within this book.

Many of these patterns we followed are usually included in the different definitions of cloud-native microservices. Therefore, we could put that label on our applications.

However, the term *cloud-native* is too ambitious, and sometimes confusing in my opinion. It's being used to pack a bunch of buzzwords and techniques across multiple aspects of software development: microservices, event-driven, continuous deployment, infrastructure-as-code, automation, containers, cloud solutions, etc.

The problem with cloud-native as a broad-scope classification of applications is that it can lead people to think they need all the included patterns and methodologies to achieve the aimed target. Microservices? Sure, it's the new standard. Event-driven? Why not. Infrastructure as code? Go for it. It looks like only if we can check all the boxes, we can say we do cloud-native applications. All these patterns and techniques offer benefits, but do you need all of them for your service or product? Maybe not. You can build a well-structured monolith, make a container out of it, and deploy it to the cloud in minutes. On top of that, you can automate all the process to build the monolith and take it to production. Is that a *cloud-native nonolith*? You won't find that definition anywhere, but that doesn't mean it's not a proper solution for your specific case.

Conclusions

This chapter guided us through an amazing journey of patterns and tools for microservices. In each section, we analyzed the issues we were facing with the current implementation. Then, we learned about well-known patterns that can solve these challenges while they also help make our system scalable, resilient, and easier to analyze and deploy, among other features.

For most of these patterns, we chose solutions that can be easily integrated with Spring Boot, given that it was the choice for our practical case. Its autoconfiguration features helped us, for instance, to quickly set up the connection with Consul as a service

discovery registry and centralized configuration server. Nevertheless, these patterns apply to many different programming languages and frameworks to build microservices, so you can reuse all the learned concepts.

Our microservice architecture became mature, and everything started working together: the gateway routes traffic transparently to multiple instances of our microservices, which can be dynamically allocated depending on our requirements. All the log outputs are channeled to a central place, where we can also see the full trace of every single process.

We also introduced containerization with Docker, which helped prepare our services to be deployed easily to multiple environments. Besides, we learned how container platforms such as Kubernetes, and cloud-based services, can help achieve the nonfunctional requirements we aimed for: scalability, resilience, etc.

At this point, you might be asking yourself why we spent almost a full chapter (a long one) to learn about all these common microservice patterns if there were easier ways to achieve the same results with container and application platforms or with managed services in the cloud. The reason is simple: you need to know how the patterns work to fully understand the solutions you're applying. If you start directly with complete platforms or cloud solutions, you get only a high-level, vendor-specific view.

With this chapter, we finalize the implementation of our microservice architecture that we started in Chapter 6. Back then, we decided to stop including extra logic within the small monolith and create a new microservice for the Gamification domain. These three chapters helped us understand the reasons to move to microservices, how to isolate and communicate them properly, and what are the patterns we should consider if we want to be successful with our project.

Chapter's Achievements:

- You learned how to use a gateway to route traffic to your microservices and provide load balancing between their instances.
- You scaled up a microservice architecture using service discovery, the HTTP load balancer, and RabbitMQ queues.
- You made the system resilient by checking the health of each instance to find out when they don't work. Also, you introduced retries to avoid losing requests.
- You saw how to override configuration per environment with an external configuration server.

- You implemented centralized logs with distributed traces across microservices, so you can follow a process from end to end easily.
- You integrated all these patterns in our microservice architecture with the projects from the Spring Cloud family: Spring Cloud Gateway, Spring Cloud Load Balancer, Spring Cloud Consul (Discovery & Configuration), and Spring Cloud Sleuth.
- You learned how to create Docker images for our applications using Spring Boot 2.3 and Cloud Native Buildpacks.
- You saw how Docker and Compose can help deploy our microservice architecture anywhere. Besides, you saw how easy is to spin up new instances using Docker.
- You compared the approach we followed in the book with other alternatives such as container platforms and application platforms, which include already some of the patterns you need for a distributed architecture.
- You understood why we introduced the new patterns and tools in every step you made in this chapter.

Afterword

In this book, we covered the main topics related to microservices architecture. We started with a look inside a Spring Boot application, traveling from an empty project to a microservice properly structured in layers. While working with the different aspects of our application, such as REST services or JPA repositories, we unveiled the magic behind Spring Boot. Besides, to build this first application, we followed a test-driven development approach that can help you clarify your functional requirements in the future.

The book tried to explain from the beginning why it's a good idea to start with a small monolith. Actually, it's an idea supported by many people very experienced with microservices: *start with a single project, identify boundaries, and decide whether it's worthwhile to split your functionality*. What happens frequently is that it's difficult to understand the reasons to start with a modular monolith if you never worked with microservices and only with monolithic applications. However, at this point in time, I'm pretty sure you understand the pain you might suffer if you go for microservices from scratch. Setting up the ecosystem without having a strong knowledge baseline and a good plan will cause chaos, at the least. Service discovery, routing, load balancing, communication between services, error handling, distributed tracing, deployment—it's important that you know what you'll face on your way *before* you start your adventure with microservices.

The next time you're confronted with the tough decision of moving to microservices, you should be able to estimate adequately the complexity that comes with that type of architecture. Don't worry if the outcome of that analysis is that you want to stay with a single, well-modularized deployment unit. The goal of this book is not to promote microservices as the new architecture design to apply everywhere, but to give a realistic view on the patterns that come with distributed systems, and their pros and cons, so you can make sound decisions. It's perfectly possible that your project doesn't benefit much from microservices; maybe the development team is small enough to keep a single project under control, or perhaps your system is not divided into domains that have different nonfunctional requirements.

AFTERWORD

In this particular adventure, you built a web application to allow users to practice their mental calculation skills every day without any help, to train their brains.

Multiplication was our first microservice, but the real challenge started when we introduced the second one, Gamification: a service that reacts to events happening in the existing logic and calculates the score and badges to make the application look like a game. At that point, we applied the event-driven pattern not before exposing the differences between synchronous and asynchronous communication.

Then, we got deep into some core concepts of microservices: how they can find each other with service discovery allowing dynamic scaling, how to apply load balancing to add capacity and resilience, how to route traffic from the outside to the corresponding component using an API gateway, how to handle errors by checking application health and retrying requests, etc. We made it practical, but looking thoroughly at the concepts, so you can apply them to any other system using different technologies.

Nowadays, you can skip many of these pattern implementations and use container or application platforms. These will manage many aspects of the microservice ecosystem for you: you just push your Spring Boot application or Docker container to the cloud, set the number of instances and how they should be routed, and everything else is handled by the platform. However, you always need to understand what you're doing. Pushing your applications to the cloud without knowing what is happening behind the curtains is risky. If errors arise, it might be impossible for you to know which part of the magic is not doing its job. Ideally, this book helps you understand many platform-provided tools, because that's why we implemented many of those patterns in our system.

I recommend you continue your learning path following the same practical approach you did in this book. For example, you can choose a cloud solution like AWS and try to deploy the system there. Alternatively, you can learn how Kubernetes works and create the required configuration to play with our application there. While you do that, I also recommend you focus on good practices for continuous integration and deployment, such as building automated scripts that can compile your microservices, run the tests, and deploy your microservices.

Don't forget that the practical case study keeps evolving online. Visit <https://tpd.io/book-extra> for extra content and source code updates. As an example, the first guide available online helps you build end-to-end tests for our distributed system using Cucumber.

Now that you've reached the end of the book, I hope you have a better understanding of the topics and you can use them at work or in your personal projects. I've enjoyed writing this book a lot, and, most importantly, I've also learned along the way. Thanks.

Index

A

Advanced Message Queuing Protocol (AMQP), 231, 239
AMQP Starter, 241, 242
description, 233, 234
exchange types and routing, 234–236
gamification (*see* Gamification)
message acknowledgments and rejection, 236, 237
multiplication, 243–250
AmqpTemplate, 246–248, 253, 270, 273, 279
ApiClient Class, 80, 139
API-first Approach, 55
Application services, 43
AssertJ, 11, 20, 21, 66

B

BadgeProcessor interface, 177, 178
Badges, 155, 182, 183, 186, 194, 206, 211–213, 219, 226
BDDMockito’s given method, 65, 66, 125
Bean’s declaration method, 253
Behavior-driven development (BDD), 15, 16, 73
BronzeBadgeProcessor Implementation, 178, 179
Business model, 39

C

ChallengeAttempt class, 43, 115, 117
ChallengeComponent class, 201 integration, 89, 90
main structure of, 85–87
rendering, 87–89
source code, 82–84
ChallengeServiceImpl class, 128, 129, 193, 247
ChallengeServiceTest class, 124–126
ChallengeSolvedDTO class, 174, 192
Choreography, 206, 222
Chrome DevTools, 91, 92
Cloud-native Microservices, 413
componentDidMount function, 86, 88
Consul UI, 311, 336
Containerization application running, 401
configuration importer, 384, 386
Consul UI, 399, 400
Docker, 376–379
Docker compose, 387–391
docker-compose.yml File, 391, 394, 396, 397
Docker configuration, 393
Dockerizing microservices, 382
Dockerizing UI, 383, 384
high-level overview, 382
hypervisors, 373

INDEX

- Containerization (*cont.*)
 - logs container, 398
 - RabbitMQ configuration, 395
 - running docker-compose ps, 393
 - scaling up, 401–403
 - service discovery, 374
 - sharing Docker images, 404–408
 - Spring Boot and Buildpacks, 380, 381
 - virtual machine deployment, 374, 375
 - Controller layer
 - ByteBuddyInterceptor class, 134
 - Hibernate objects, 134
 - JsonConfiguration class, 135
 - retrieve statistics, 132
 - serialization, 133, 135, 136
 - suboptimal configuration, 137
 - Cross-origin resource
 - sharing (CORS), 93, 94
 - CrudRepository interface, 121
-
- ## D
- Data access objects (DAOs), 120
 - Data-driven
 - decision-making (DDDM), 154
 - Data layer
 - API
 - controller layer, 132–137
 - displaying last attempts, 130
 - service layer, 130–132
 - ChallengeAttempt, 103, 104
 - conceptual model, 102, 103
 - domain model, 104, 105
 - DTOs, 104
 - entities, 115–120
 - feature, 144–146
 - H2, hibernate, and JPA, 106, 107
 - repositories, 120–124
 - requirements, 102
 - Spring Boot Data JPA
 - data source (auto)configuration, 113–115
 - dependencies and autoconfiguration, 107–111
 - technology stack, 111–113
 - SQL *vs.* NoSQL, 105, 106
 - storing users and attempts
 - BDDMockito's given method, 125
 - ChallengeServiceImpl class, 128, 129
 - ChallengeServiceTest class, 125, 126
 - repository classes, 129
 - returnsFirstArg() utility method, 126
 - test cases, 124
 - user entity, 127
 - userRepository, 127
 - three-tier architecture, 101, 102
 - user interface, 138
 - ApiClient class, 139
 - app.css file, 138, 139
 - App.js file, 138
 - ChallengeComponent class, 141–143
 - index.css file, 138
 - LastAttemptsComponent, 140
 - setState method, 143
 - Data transfer objects (DTOs), 48
 - Dependency injection, 10
 - Docker, 376–379
 - Dockerizing Microservices, 382
 - Docker Hub, 404
 - Document Object Model (DOM), 80
 - Domain-driven design (DDD), 40

E

Enterprise service bus (ESB), 216–218
 Event-driven architecture, 4
 AMQP, 239
 AMQP Starter, 241, 242
 description, 233, 234
 exchange types and routing, 234–236
 gamification (*see* Gamification)
 message acknowledgments and rejection, 236, 237
 multiplication, 243–250
 asynchronous messaging, 223–225
 asynchronous process, 240
 ChallengeSolvedDTO object, 220
 choreography, 222
 definition, 218
 HTTP error response, 269
 logs of, 264
 message brokers
 ESB, 216–218
 high-level view, 216
 protocols, standards, and tools, 231, 232
 messaging patterns, 229
 data durability, 230, 231
 filtering, 230
 publish-subscribe, 230
 work queues, 230
 microservices, scaling up
 first test, 277, 278
 interface questions, 274, 276
 load balancing, 275, 276
 RabbitMQ queue, 276, 277
 second instance, 278, 279
 server mode, 275
 server.port parameter, 275

nonpersistent delivery mode, 270
 persistent delivery mode, 270
 pros and cons of, 226–229
 RabbitMQ (*see* RabbitMQ)
 RabbitMQ UI
 connections, 264, 265
 exchange detail, 262, 263
 exchange list, 262
 queue detail, 265
 queue list, 263
 single connection, 261
 reactive systems, 226
 REST *vs.* message, 219–221
 scenarios, 259, 260
 Spring Boot Application Logs, 260
 tasks, 241
 transactionality, 270–273
 Event-driven system, 6

F

findByAlias query method, 124
 FirstWonBadgeProcessor
 Implementation, 179–181

G

GameApiClient class, 196
 GameService interface, 173–177
 Gamification
 AMOP
 AmqpTemplate, 253
 Bean's declaration method, 253
 ChallengeSolvedEvent, 254
 changes, 258
 configuration class, 251–253
 newAttemptForUser method, 257
 queue and exchange names, 251

INDEX

- Gamification (*cont.*)
 RabbitListener annotations, 256
 RabbitMQ Consumer's Logic, 255
API exposure, 213
BadgeProcessor interface, 177, 178
BronzeBadgeProcessor
 Implementation, 179
ChallengeSolvedDTO class, 174
configuration, 188–190
controller, 186–188
data, 183–186
domain, 157, 166–168
 BadgeCard Domain/Data
 Class, 170, 171
 BadgeType Enum, 169, 170
 LeaderBoardRow, 171
 ScoreCard Domain/Data
 Class, 168, 169
fault tolerance, 203, 204
FirstWonBadgeProcessor
 Implementation, 179–181
game logic, 172, 173
GameService interface, 173–177
LeaderBoardService interface, 181–183
logical view, 205
logs, 267, 268
multiplication, 190–195
playing with system, 201–203
points, badges, and leaderboards,
 155, 156
RabbitMQ UI, 267
synchronous interfaces vs. eventual
 consistency, 206–211
tight coupling, 206
transactions, 211, 212
TTL, 268
user interface, 195–201
Gamification microservice, 283, 294, 332
- GamificationServiceClient
 service, 247
Gateway microservice, 290, 291
getStatsForUser method, 130, 131
- H**
- Hibernate, 108, 111
HttpMessageConverter interface, 58
Hystrix (circuit-breaking), 286
- I**
- Inversion of Control (IoC)
 container, 10
- J**
- Java, 12–15, 26
Java Management eXtensions (JMX), 300
Java Message Service (JMS), 231
Java Persistence API (JPA), 106
JsonConfiguration class, 135
JUnit 5, 16, 18, 21
JVM-based languages, 9
- K**
- Kafka, 217, 232, 280, 370
- L**
- LeaderBoardService Implementation,
 182, 183
LeaderBoardService interface, 181, 182
Load balancing, 157, 236, 276, 277, 280,
 283, 284, 286, 306
Logging, 14, 22, 23
Lombok, 23

M

Message Queuing Telemetry Transport (MQTT), [231](#)

Microservice architectures

- application platforms, [410](#)
- centralized logs
 - AMQPConfiguration Class, [363](#)
 - centralized logs, [366](#)
 - Consumer Class, [364](#)
 - high-level overview, [367](#)
 - log aggregation pattern, [356, 357](#)
 - LogBack Configuration, [365](#)
 - log centralization, [357–360](#)
 - logs microservice, [362](#)
- cloud-native Microservices, [413](#)
- cloud providers, [410, 411](#)
- clustering strategies, [408](#)
- configuration per environment
 - centralized configuration, [339, 347, 349, 351, 352, 354, 355](#)
- configurations in Consul, [340, 341](#)
- implementing centralized configuration, [343–345, 347](#)
- rabbitprod profile, [337](#)
- Spring Cloud Consul Config, [341, 343](#)
- container platforms, [408–410](#)
- distributed tracing
 - implementing, [371–373](#)
 - RabbitMQ messages, [368](#)
 - Spring Cloud Sleuth, [369, 370](#)
- gateway
 - cloud patterns, [286–288](#)
 - dependencies, [292](#)
 - first approach, [293](#)
 - high-level overview, [285](#)
 - load balancing, [297](#)
 - multiplication API URL, [295](#)

routes, predicates, and filters, [288](#)

routing configuration, [289, 290](#)

running, [296, 297](#)

WebConfiguration class, [294](#)

health

- adding Spring Boot Actuator, [302–304](#)
- high-level overview, [299](#)
- redundancy, [298](#)
- Spring Boot Actuator, [299–302](#)
- making decision, [412](#)

Microservices, [1, 6](#)

- advantages, [160, 161](#)
- architecture, [163, 164](#)
- disadvantages, [161, 162](#)
- fine-grained nonfunctional requirements, [160](#)

gamification

- API exposure, [213](#)
- BadgeProcessor interface, [177, 178](#)
- BronzeBadgeProcessor
 - Implementation, [178, 179](#)
- ChallengeSolvedDTO class, [174](#)
- data, [183–186](#)
- domain, [172](#)
- configuration, [188–190](#)
- controller, [186–188](#)
- domain, [157, 166–171](#)
- fault tolerance, [203, 204](#)
- FirstWonBadgeProcessor
 - Implementation, [179–181](#)
- game logic, [172, 173](#)
- GameService interface, [173–177](#)
- LeaderBoardService
 - Implementation, [182](#)
- LeaderBoardService interface, [181, 182](#)
- logical view, [205](#)

INDEX

Microservices (*cont.*)

- multiplication, 190–195
- playing with system, 201–203
- points, badges, and leaderboards, 155, 156
- Spring Boot Skeleton, 165, 166
- synchronous interfaces vs. eventual consistency, 206–211
- tight coupling, 206
- transactions, 211, 212
- user interface, 195–201
- horizontal scalability, 158, 159
- improvements, 156
- independent workflows, 157, 158
- interfaces, 165
- requirements, 154
- small monolith
 - definition, 150
 - embracing refactoring, 152
 - planning, 152–154
 - problems with, 150, 151
 - for small teams, 151, 152
- users and challenges
 - domains, 157
- Minimum viable product (MVP), 5
- Mockito, 17–20
- Mosquitto, 232

N

- newAttemptForUser method, 257
- Nonfunctional requirements, 5, 6, 414, 417
- Nonpersistent delivery mode, 270

O

- Object/relational mapping (ORM), 106
- Open source software (OSS), 286

P, Q

- Persistent delivery mode, 270
- Points, gamification, 155
- Project Lombok, 12–15

R

- RabbitHealthIndicator, 300, 301
- RabbitMQ, 11, 232, 233, 237–239, 241, 242, 244, 251, 254, 255, 261, 304, 337, 357, 360, 362, 369, 398
- React
 - ChallengeComponent class
 - integration, 89, 90
 - main structure of, 85–87
 - rendering, 87–89
 - source code, 82–84
 - CORS configuration, 93, 94
 - creation, 90, 91
 - debugging, 91, 93
 - definition, 75
 - deployment, 95–98
 - JavaScript Client, 80, 81
 - output, 77
 - playing with, 94
 - skeleton, 78–80
 - version of Node.js and npm, 76
- Reactive systems, 226
- refreshLeaderBoard function, 200
- Remote procedure call (RPC), 219
- REpresentational State Transfer (REST), 53
- RestTemplate class, 190
- returnsFirstArg() utility method, 126

S

- Service discovery and load balancing
 - adding Log Line, 327, 328

- consul, 310, 311
- in gateway, 323–327
- logs for gamification, 331
- logs for multiplication, 330
- multiplication and gamification, 329
- noninternal HTTP communication, 308
- no retry pattern, 335
- pattern overview, 305
- retry pattern, 333
- Spring Cloud Consul, 312–315, 317–319
- Spring Cloud load balancer, 319–322
- Service layer, 130–132
- Service registry, 304–308, 326, 328
- setState function, 86
- SimpleJpaRepository class, 121
- Skeleton Web App, 27–31
- Small monolith, 150
- Spring Boot, 10, 11, 239
- Spring Boot Actuator, 298–302
- Spring Boot and Buildpacks, 380, 381
- Spring Boot application, 4, 5
 - autoconfiguration
 - embedded Tomcat, 35
 - main starter, 33
 - tomcat starter dependencies, 32, 33
 - ServletWebServerFactory
 - Configuration Fragment, 34, 35
 - web starter dependencies, 32
- business logic
 - generate random
 - challenges, 44–48
 - verify attempts, 48–52
- domain objects, 39
- modeling domain
 - business entities, 39
 - classes, 41–43
 - domain-driven design, 39–41
- presentation layer
 - context to override defaults, 60
 - newly created API, 57
 - new request, 60
 - REST API, 53–55
 - Spring Boot application, 56
 - testing controllers, 61–65
 - validating data in controllers, 67–73
 - valid attempt test, 66, 67
 - WebMvcAutoConfiguration Class, 58
- Skeleton Web App, 27–31
- test-driven development, 26
- three-layer architecture, 36–39
- Spring Boot Data JPA
 - data source (auto)configuration, 113–115
 - dependencies and autoconfiguration, 107–111
 - technology stack, 111–113
- Spring Boot’s Maven plugin, 29, 31, 317, 337, 380
- Spring Boot test package, 21
- Spring Boot Web JSON
 - autoconfiguration, 59
- Spring Cloud consul, 312–315, 317–319
- Spring Cloud gateway, 11
- Spring Cloud load balancer, 11, 319–322
- Spring Cloud Sleuth, 369, 370
- Spring Core Web libraries, 11
- Spring ecosystem, 310
- Spring framework, 9
- Spring Initializr, 27, 290
- Spring integration, 10
- Spring security, 10
- StackOverflow, 155
- Streaming Text Oriented Messaging Protocol (STOMP), 231

INDEX

T

Test-driven development (TDD), [5, 15, 417](#)
Tomcat, [31, 32, 34](#)

U

User class, [115, 116](#)
UserController class, [194](#)
User interface (UI), [138](#)
ApiClient class, [139](#)

app.css file, [138, 139](#)
App.js file, [138](#)
ChallengeComponent class, [141–143](#)
index.css file, [138](#)
LastAttemptsComponent, [140](#)
setState method, [143](#)

V, W, X, Y, Z

Virtual machine (VM), [373–375, 411](#)