

ENGENHARIA

DE

SOFTWARE

MODERNA

**Princípios e práticas para
desenvolvimento de software
com produtividade**



MARCO TULIO VALENTE

Versão 2020.1.0 - LeanPub

Direitos autorais protegidos pela Lei 9.610, de 10/02/1998. Versão para uso pessoal e individual, sendo proibida qualquer forma de redistribuição.

Para Cynthia, Daniel e Mariana.

Conteúdo

Prefácio	1
Público Alvo	3
Pré-requisitos	3
Website	3
Sobre o Autor	4
1 Introdução	5
1.1 Definições, Contexto e História	5
1.2 O que se Estuda em Engenharia de Software?	10
1.3 Sistemas ABC: Acute, Business e Casuais	26
1.4 Próximos Capítulos	28
Bibliografia	31
Exercícios de Fixação	31
2 Processos	33
2.1 Importância de Processos	33
2.2 Manifesto Ágil	35
2.3 Extreme Programming	40
2.4 Scrum	58
2.5 Kanban	66
2.6 Quando não Usar Métodos Ágeis?	73
2.7 Outros Métodos Iterativos	75
Bibliografia	79
Exercícios de Fixação	79
3 Modelos	83
3.1 Modelos de Software	83
3.2 UML	85
3.3 Diagramas de Classes	90

3.4	Diagramas de Pacotes	98
3.5	Diagramas de Sequência	100
	Bibliografia	106
	Exercícios de Fixação	107
4	Arquitetura	110
4.1	Introdução	110
4.2	Arquitetura em Camadas	113
4.3	Arquitetura MVC	116
4.4	Microsserviços	121
4.5	Arquiteturas Orientadas a Mensagens	128
4.6	Arquiteturas Publish/Subscribe	130
4.7	Outros Padrões Arquiteturais	132
4.8	Anti-padrões Arquiteturais	133
	Bibliografia	135
	Exercícios de Fixação	135
5	Refactoring	137
5.1	Introdução	137
5.2	Catálogo de Refactorings	140
5.3	Prática de Refactoring	155
5.4	Refactorings Automatizados	156
5.5	Code Smells	159
	Bibliografia	168
	Exercícios de Fixação	168
6	DevOps	172
6.1	Introdução	172
6.2	Controle de Versões	176
6.3	Integração Contínua	181
6.4	Entrega Contínua	188
6.5	Engenharia de Releases	188
	Bibliografia	191
	Exercícios de Fixação	191
A	Git	193
A.1	Init & Clone	193
A.2	Commit	194
A.3	Add	195
A.4	Push & Pull	197

A.5 Conflitos de Merge	199
A.6 Branches	201
A.7 Branches Remotos	205
A.8 Pull Requests	206
A.9 Squash	208
A.10 Forks	209
Bibliografia	210
Exercícios de Fixação	210

Prefácio

A inutilidade dos prefácios é um lugar comum da história dos prefácios, portanto serei breve. – Eduardo Giannetti

A ideia de escrever este livro surgiu no início de 2019, quando fui alocado para ministrar a disciplina Engenharia de Software, do Bacharelado em Ciência da Computação, da UFMG. Para preparar o curso, comecei com uma análise dos principais livros de Engenharia de Software. Para minha surpresa, percebi que eles tinham mudado pouco desde que cursei a disciplina na minha graduação: há mais de 25 anos!

Meu objetivo era escolher um livro que permitisse, no início de uma aula, dizer para os alunos: “hoje vamos estudar tal assunto, que corresponde a tal capítulo do livro texto”. No final da aula, gostaria de sugerir aos alunos: “para fixar a matéria que acabamos de ver, sugiro que façam tais exercícios”. No entanto, infelizmente, não encontrei esse livro. Em vez disso, tive que fazer uma extensa pesquisa e leitura de pelo menos 15 livros.

Como resultado, preparei mais de 600 slides, que considero conter o principal material que deve ser tratado em uma disciplina de graduação em Engenharia de Software, especificamente em cursos que possuem uma única disciplina na área. Porém, estudar apenas por slides não proporciona a mesma experiência de aprendizado obtida com a leitura atenta de um texto completo e contextualizado.

Assim, surgiu a ideia de transformar os slides em livro, que pudesse desempenhar o papel do sonhado livro texto, já visando futuras ofertas da disciplina. E que fosse útil também para outros professores, que devem enfrentar problemas semelhantes ao meu, quando têm que ministrar um curso de Engenharia de Software.

Gostaria, então, de destacar os seguintes pontos sobre o conteúdo, a organização e a estratégia usada na escrita do livro:

- Ele foi escrito para ser um livro moderno, com ênfase em técnicas e princípios que são largamente usados na construção de software nos dias de hoje.
- Por outro lado, o livro também cobre técnicas e princípios tradicionais, porém de forma rápida. O motivo é que achamos importante cobrir a história da área, mencionando o que deu certo e o que não deu. Achamos que essa visão é importante na formação e amadurecimento dos alunos.
- O livro inclui inúmeros exemplos e discussões de casos reais. Para isso, criei a seção “Mundo Real”, onde os assuntos são ilustrados com exemplos recentes e reais, provenientes de grandes empresas de software e também de artigos científicos, buscando-se apresentar o que se faz de melhor tanto na indústria como na academia.
- O livro foi escrito em Português, pois o objetivo é contribuir, primeiro, com os cursos e alunos brasileiros. Porém, optamos por não traduzir alguns termos — como refactoring, branches, sprint, etc — pois achamos que eles são usados, também sem tradução, pelos desenvolvedores brasileiros no seu dia a dia.
- Apesar de tratar de temas modernos, a intenção foi escrever um livro duradouro. Por isso, temas e tecnologias que ainda não passaram pelo teste do tempo não são abordados (ou são abordados de forma rápida). Um exemplo são os últimos frameworks e arquiteturas para implementação de sistemas, que tendem a mudar rapidamente. Também não acoplamos o livro a nenhuma linguagem de programação. Por exemplo, todos os trechos de código são mostrados em uma sintaxe bastante neutra.
- Ele não é um livro extenso, com diversos capítulos que, na prática, tratam de assuntos que não são importantes nos cursos atuais.
- Cada vez mais, engenheiros de software têm que escrever código. Hoje, há pouco espaço para dizer que eu não preciso programar, pois sou arquiteto ou analista. Por isso, nos capítulos de projeto, testes e refatoração procuramos seguir a recomendação de Linus Torvalds (criador do Linux): “falar é fácil, mas mostre-me o código”. Assim, esses capítulos incluem dezenas de exemplos de código, que simulam problemas e soluções típicos de sistemas reais.
- Ao longo da escrita, procuramos praticar o que enfatizamos no livro, principalmente no que diz respeito à prevalência atual de métodos de desenvolvimento

ágeis. Assim, o livro foi escrito seguindo princípios ágeis. Cada capítulo foi tratado com sendo um sprint; uma vez pronto, ele foi disponibilizado para uso, para receber críticas e sugestões. Como ocorre com software, acho arriscado, nos dias de hoje, passar anos escrevendo um manuscrito, trancado em minha sala, para só então torná-lo público.

Público Alvo

O livro destina-se a alunos de cursos de graduação. Ele foi escrito para ser adotado em cursos que possuem uma única disciplina de Engenharia de Software, com 60 horas. Porém, achamos também que ele pode ser usado — junto com outros livros — em cursos com duas ou mais disciplinas na área.

Além disso, ele pode ser adotado em cursos técnicos e tecnológicos. Na verdade, fizemos um esforço para usar uma linguagem clara, próxima à linguagem coloquial, exatamente para não criar barreiras à adoção do livro. Por fim, também escrevemos o livro pensando em profissionais da área, que estão em busca de aperfeiçoamento em temas e métodos modernos de Engenharia de Software. De forma nenhuma, este é um livro descolado do estado da prática em Engenharia de Software.

Pré-requisitos

Espera-se que os leitores tenham domínio de conceitos básicos de programação e de algoritmos e estruturas de dados. Além disso, recomendamos domínio de orientação a objetos. Supondo um curso de graduação de 4 anos, os alunos devem estar na metade final para que possam aproveitar o conteúdo do livro.

Website

O livro possui uma versão online, em HTML, disponível em:

<https://engsoftmoderna.info>

Nesse site, estão também disponíveis diversos exercícios de múltipla escolha e alguns roteiros de aula prática.

Não menos importante: para reportar qualquer erro, mesmo que um simples erro ortográfico, você pode usar este [formulário](#).

Marco Túlio Valente

Belo Horizonte, 23 de janeiro de 2020.

Sobre o Autor

Marco Túlio Valente é doutor em Ciência da Computação pela Universidade Federal de Minas Gerais, onde atualmente é Professor Associado do Departamento de Ciência da Computação. Seus interesses de pesquisa concentram-se em Engenharia de Software, especificamente nas áreas de Manutenção e Evolução de Software, Qualidade de Software e Mineração de Repositórios de Software. Ele é Bolsista de Produtividade Nível 1D do CNPq e bolsista do Programa Pesquisador Mineiro da FAPEMIG. É autor de mais de 100 artigos em periódicos e conferências internacionais, muitos deles em co-autoria com pesquisadores internacionais de renome. Orientou mais de 30 dissertações de mestrado e cerca de uma dezena de teses de doutorado. Desde 2010, coordena o Applied Software Engineering Research Group ([ASERG](#)), do DCC/UFMG. É responsável também pelo [CSIndexbr](#), sistema que oferece dados transparentes sobre a produção científica brasileira em Ciência da Computação.



Capítulo 1

Introdução

Our civilization runs on software. – Bjarne Stroustrup

Neste primeiro capítulo, vamos definir e contextualizar o que é Engenharia de Software (Seção 1.1) e dar uma visão geral dos principais assuntos estudados nesta área da Computação (Seção 1.2). O objetivo é propiciar ao leitor uma visão horizontal da área de Engenharia de Software, antes de nos aprofundarmos em temas específicos. Além disso, sendo Engenharia de Software uma área bastante ampla, vamos caracterizar os tipos de sistemas de software que podem se beneficiar das técnicas e conceitos apresentados neste livro (Seção 1.3). O objetivo é, logo no início, evitar falsas expectativas em relação ao conteúdo do trabalho. Por fim, iremos apresentar a estrutura e os assuntos tratados nos capítulos restantes do livro (Seção 1.4).

1.1 Definições, Contexto e História

No mundo moderno, tudo é software. Por exemplo, hoje em dia, empresas de qualquer tamanho dependem dos mais diversos sistemas de informação para automatizar seus processos. Governos interagem com cidadãos por meio de sistemas computacionais, por exemplo, para coletar impostos ou realizar eleições. Empresas vendem, por meio de sistemas de comércio eletrônico, uma gama imensa de produtos, diretamente para os consumidores. Software está também embarcado em diferentes dispositivos e produtos de engenharia, incluindo automóveis, aviões, satélites, robôs, etc. Por fim, software está contribuindo para renovar indústrias

e serviços tradicionais, como telecomunicações, transporte em grandes centros urbanos, hospedagem, lazer e publicidade.

Portanto, devido a sua relevância no nosso mundo, não é surpresa que exista uma área da Computação destinada a investigar os desafios e propor soluções que permitam desenvolver sistemas de software — principalmente aqueles mais complexos e de maior tamanho — de forma produtiva e com qualidade. Essa área é chamada de **Engenharia de Software**.

Engenharia de Software trata da aplicação de abordagens sistemáticas, disciplinadas e quantificáveis para desenvolver, operar, manter e evoluir software. Ou seja, Engenharia de Software é a área da Computação que se preocupa em propor e aplicar princípios de engenharia na construção de software.

Historicamente, a área surgiu no final da década de 60 do século passado. Nas duas décadas anteriores, os primeiros computadores modernos foram projetados e começaram a ser usados principalmente para resolução de problemas científicos. Ou seja, nessa época software não era uma preocupação central, mas sim construir máquinas que pudessem executar alguns poucos programas. Em resumo, computadores eram usados por poucos e para resolver apenas problemas científicos.

No entanto, progressos contínuos nas tecnologias de construção de hardware mudaram de forma rápida esse cenário. No final da década de 60, computadores já eram mais populares, já estavam presentes em várias universidades norte-americanas e europeias e já chegavam também em algumas grandes empresas. Os cientistas da computação dessa época se viram diante de um novo desafio: como os computadores estavam se tornando mais populares, novas aplicações não apenas se tornavam possíveis, mas começaram a ser demandadas pelos usuários dos grandes computadores da época. Na verdade, os computadores eram grandes no sentido físico e não em poder de processamento, se comparado com os computadores atuais. Dentre essas novas aplicações, as principais eram sistemas comerciais, como folha de pagamento, controle de clientes, controle de estoques, etc.

Conferência da OTAN: Em outubro de 1968, um grupo de cerca de 50 renomados Cientistas da Computação se reuniu durante uma semana em Garmisch, na Alemanha, em uma conferência patrocinada por um comitê científico da OTAN, a organização militar que congrega os países do Atlântico Norte (veja uma foto da reunião na próxima figura). O objetivo da conferência era chamar a atenção para um “problema crucial do uso de computadores, o chamado software”. A conferência produziu um relatório, com mais de 130 páginas, que afirmava a necessidade de que software fosse construído com base em princípios práticos e teóricos, tal como ocorre em ramos tradicionais e bem estabelecidos da Engenharia. Para deixar essa

proposta mais clara, decidiu-se cunhar o termo Engenharia de Software. Por isso, a Conferência da OTAN é considerada o marco histórico de criação da área de Engenharia de Software.



Figura 1.1: Cientistas na conferência da OTAN de 1968 sobre Engenharia de Software. Reprodução gentilmente autorizada pelo Prof. Robert McClure.

O comentário a seguir, de um dos participantes da Conferência da OTAN, ilustra os desafios que esperavam a recém criada área de pesquisa:

“O problema é que certas classes de sistemas estão colocando demandas sobre nós que estão além das nossas capacidades e das teorias e métodos de projeto que conhecemos no presente tempo. Em algumas aplicações não existe uma crise, como rotinas de ordenação e folhas de pagamento, por exemplo. Porém, estamos tendo dificuldades com grandes aplicações. Não podemos esperar que a produção de tais sistemas seja fácil.”

Passado mais de meio século da Conferência da OTAN, os avanços obtidos em técnicas e métodos para construção de software são notáveis. Hoje, já se tem conhecimento de que software — na maioria das vezes — não deve ser construído em fases estritamente sequenciais, como ocorre com produtos tradicionais de engenharia, tais como Engenharia Civil, Engenharia Mecânica, Engenharia Eletrônica, etc. Já existem também padrões que podem ser usados por Engenheiros de Software em seus novos sistemas, de forma que eles não precisam “reinventar a roda” toda vez que enfrentam um novo problema de projeto. Bibliotecas e frameworks para os mais diversos fins estão largamente disponíveis, de forma que desenvolvedores de software podem reusar código sem se preocupar com detalhes inerentes a tarefas como implementar interfaces gráficas, criar estruturas de dados, acessar bancos de

dados, criptografar mensagens, etc. Prosseguindo, as mais variadas técnicas de testes podem (e devem) ser usadas para garantir que os sistemas em construção tenham qualidade e que falhas não ocorram quando eles entrarem em produção e forem usados por clientes reais. Sabe-se também que sistemas envelhecem, como outros produtos de engenharia. Logo, software também precisa de manutenção, não apenas corretiva, para corrigir bugs reportados por usuários, mas também para garantir que os sistemas continuem fáceis de manter e entender, mesmo com o passar dos anos.

Não existe bala de prata

Como começamos a afirmar no parágrafo anterior, desenvolvimento de software é diferente de qualquer outro produto de Engenharia, principalmente quando se compara software com hardware. Frederick Brooks, Prêmio Turing em Computação (1999) e um dos pioneiros da área de Engenharia de Software, foi um dos primeiros a chamar a atenção para esse fato. Em 1987, em um ensaio intitulado *Não Existe Bala de Prata: Essência e Acidentes em Engenharia de Software* ([link](#)), ele discorreu sobre as particularidades da área de Engenharia de Software.

Segundo Brooks, existem dois tipos de dificuldades em desenvolvimento de software: **dificuldades essenciais** e **dificuldades accidentais**. As essenciais são da natureza da área e dificilmente serão superadas por qualquer nova tecnologia ou método que se invente.

Daí a menção à bala de prata no título do ensaio. Diz a lenda que uma bala de prata é a única maneira de matar um lobisomem, desde que usada em uma noite de lua cheia. Ou seja, por causa das dificuldades essenciais, não podemos esperar soluções milagrosas em Engenharia de Software, na forma de balas de prata. O interessante é que, mesmo conhecendo o ensaio de Brooks, sempre surgem novas tecnologias que são vendidas como se fossem balas de prata.

Segundo Brooks, as dificuldades essenciais são as seguintes:

- Complexidade: dentre as construções que o homem se propõe a realizar, software é uma das mais desafiadoras e mais complexas que existe. Na verdade, como dissemos antes, mesmo construções de engenharia tradicional, como um satélite, uma usina nuclear ou um foguete, são cada vez mais dependentes de software.
- Conformidade: pela sua natureza software tem que se adaptar ao seu ambiente, que muda a todo momento no mundo moderno. Por exemplo, se as leis para recolhimento de impostos mudam, normalmente espera-se que os sistemas sejam rapidamente adaptados à nova legislação. Brooks comenta

que isso não ocorre, por exemplo, na Física, pois as leis da natureza não mudam de acordo com os caprichos dos homens.

- Facilidade de mudanças (*changeability*): que consiste na necessidade de evoluir sempre, incorporando novas funcionalidades. Na verdade, quanto mais bem sucedido for um sistema de software, mais demanda por mudanças ele recebe.
- Invisibilidade: devido à sua natureza abstrata, é difícil visualizar o tamanho e consequentemente estimar o desafio de construir um sistema de software.

As dificuldades (2), (3) e (4) são específicas de sistemas de software, isto é, elas não ocorrem em outros produtos de Engenharia, pelo menos na mesma intensidade. Por exemplo, quando a legislação ambiental muda, os fabricantes de automóveis tem anos para se conformar às novas leis. Adicionalmente, carros não são alterados, pelo menos de forma essencial, com novas funcionalidades, após serem vendidos. Por fim, um carro é um produto físico e visível, possuindo peso, altura, largura, número de assentos, forma geométrica, etc; o que facilita sua avaliação e especificação por consumidores finais.

Ainda segundo Brooks, desenvolvimento de software enfrenta também dificuldades acidentais. No entanto, elas estão associadas a problemas tecnológicos, que os Engenheiros de Software podem resolver, se devidamente treinados e caso tenham acesso às devidas tecnologias e recursos. Como exemplo, podemos citar as seguintes dificuldades: um compilador que produz mensagens de erro obscuras, uma IDE que possui muitos bugs e frequentemente sofre crashes, um framework que não possui documentação, uma aplicação Web com uma interface pouco intuitiva, etc. Todas essas dificuldades dizem respeito à solução adotada e, portanto, não são uma característica inerente dos sistemas mencionados.

Mundo Real: Para ilustrar a complexidade envolvida na construção de sistemas de software reais, vamos dar alguns números sobre o tamanho desses sistemas, em linhas de código. Por exemplo, o sistema operacional Linux, em sua versão 4.1.3, de 2017, possui cerca de 25 milhões de linhas de código e contribuições de quase 1.700 engenheiros ([link](#)). Para mencionar um segundo exemplo, os sistemas do Google somavam 2 bilhões de linhas de código, distribuídas por 9 milhões de arquivos, em janeiro de 2015 ([link](#)). Nesta época, cerca de 40 mil solicitações de mudanças de código (commits) eram realizadas, em média, por dia, pelos cerca de 25 mil Engenheiros de Software empregados pelo Google.

1.2 O que se Estuda em Engenharia de Software?

Para responder a essa pergunta, vamos nos basear no *Guide to the Software Engineering Body of Knowledge*, também conhecido pela sigla SWEBOK ([link](#)). Trata-se de um documento, organizado pela IEEE Computer Society (uma sociedade científica internacional), com o apoio de diversos pesquisadores e de profissionais da indústria. O objetivo do SWEBOK é precisamente documentar o corpo de conhecimento que caracteriza a área que hoje chamamos de Engenharia de Software.

O SWEBOK define 12 áreas de conhecimento em Engenharia de Software:

1. Engenharia de Requisitos
2. Projeto de Software
3. Construção de Software
4. Testes de Software
5. Manutenção de Software
6. Gerência de Configuração
7. Gerência de Projetos
8. Processos de Software
9. Modelos de Software
10. Qualidade de Software
11. Prática Profissional
12. Aspectos Econômicos

Na verdade, o SWEBOK inclui mais três áreas de conhecimento: Fundamentos de Computação, Fundamentos de Matemática e Fundamentos de Engenharia. No entanto, sendo áreas de fronteira, elas não serão tratadas neste capítulo.

No restante desta seção, vamos brevemente discutir e comentar sobre cada uma das 12 áreas listadas acima. O nosso objetivo é propiciar ao leitor um panorama do conhecimento que se adquiriu ao longo dos anos em Engenharia de Software e, assim, informá-lo sobre *o que* se estuda nessa área.

1.2.1 Engenharia de Requisitos

Os requisitos de um sistema definem *o que* ele deve fazer e *como* ele deve operar. Assim, a Engenharia de Requisitos inclui o conjunto de atividades realizadas com o objetivo de definir, analisar, documentar e validar os requisitos de um sistema. Em uma primeira classificação, os requisitos podem ser **funcionais** ou **não-funcionais**.

Requisitos funcionais definem *o que* um sistema deve fazer; isto é, quais funcionalidades ou serviços ele deve implementar.

Já os requisitos não-funcionais definem *como* um sistema deve operar, sob quais restrições e com qual qualidade de serviço. São exemplos de requisitos não-funcionais: desempenho, disponibilidade, tolerância a falhas, segurança, privacidade, interoperabilidade, capacidade, manutenibilidade, usabilidade, dentre outros.

Por exemplo, suponha um sistema de *home-banking*. Neste caso, os requisitos funcionais incluem informar o saldo da conta, informar o extrato, realizar transferência entre contas, pagar um boleto bancário, cancelar um cartão de débito, dentre outros. Já os requisitos não-funcionais, dentre outros, incluem:

- Desempenho: informar o saldo da conta em menos de 5 segundos;
- Disponibilidade: estar no ar 99.99% do tempo;
- Tolerância a falhas: continuar operando mesmo se um centro de dados cair;
- Segurança: criptografar todos os dados trocados com as agências;
- Privacidade: não disponibilizar para terceiros dados de clientes;
- Interoperabilidade: integrar-se com os sistemas do Banco Central;
- Capacidade: ser capaz de armazenar dados de 1 milhão de clientes;
- Usabilidade: ter uma versão para deficientes visuais.

1.2.2 Projeto de Software

Durante o projeto de um sistema de software, são definidas suas principais unidades de código, porém apenas no nível de interfaces, incluindo **interfaces providas** e **interfaces requeridas**. Interfaces providas são aqueles serviços que uma unidade de código torna público, para uso pelo resto do sistema. Interfaces requeridas são aquelas interfaces das quais uma unidade de código depende para funcionar.

Portanto, durante o projeto de um sistema de software, não entramos em detalhes de implementação de cada unidade de código, tais como detalhes de implementação

dos métodos de uma classe, caso o sistema seja implementado em uma linguagem orientada a objetos.

Por exemplo, durante o projeto de um sistema de *home-banking*, pode-se propor uma classe para representar contas bancárias, como a seguinte:

```
class ContaBancaria {  
    private Cliente cliente;  
    private double saldo;  
    public double getSaldo() { ... }  
    public String getNomeCliente() { ... }  
    public String getExtrato (Date inicio) { ... }  
    ...  
}
```

Primeiro, é importante mencionar que a implementação acima é bem simples, pois o nosso objetivo é didático, isto é, diferenciar projeto de software de sua implementação. Para atingir esse objetivo, o importante é mencionar que `ContaBancaria` oferece uma interface para as demais classes do sistema, na forma de três métodos públicos, que constituem a interface provida pela classe. Por outro lado, `ContaBancaria` também depende de uma outra classe, `Cliente`; logo, a interface de `Cliente` é uma interface requerida por `ContaBancaria`. Muitas vezes, interfaces requeridas são chamadas de dependências. Isto é, `ContaBancaria` possui uma dependência para `Cliente`.

Quando o projeto é realizado em um nível mais alto e as unidades de código possuem maior granularidade — são pacotes, por exemplo — ele é classificado como um projeto arquitetural. Ou seja, **arquitetura de software** trata da organização de um sistema em um nível de abstração mais alto do que aquele que envolve classes ou construções semelhantes.

1.2.3 Construção de Software

Construção trata da implementação, isto é, codificação do sistema. Neste momento, existem diversas decisões que precisam ser tomadas, como, por exemplo: definir os algoritmos e estruturas de dados que serão usados, definir os frameworks e bibliotecas de terceiros que serão usados; definir técnicas para tratamento de exceções; definir padrões de nomes, layout e documentação de código e, por último, mas não menos importante, definir as ferramentas que serão usadas no desenvolvimento, incluindo compiladores, ambientes integrados de desenvolvimento (IDEs), depuradores, sistemas gerenciadores de bancos de dados, ferramentas para construção de interfaces, etc.

1.2.4 Testes de Software

Teste consiste na execução de um programa com um conjunto finito de casos, com o objetivo de verificar se ele possui o comportamento esperado. A seguinte frase, bastante famosa, de Edsger W. Dijkstra — também prêmio Turing em Computação (1982) — sintetiza não apenas os benefícios de testes, mas também suas limitações:

“Testes de software mostram a presença de bugs, mas não a sua ausência.”

Pelo menos três pontos podem ser comentados sobre testes, ainda nesta Introdução.

Primeiro, existem diversos tipos de testes. Por exemplo, **testes de unidade** (quando se testa uma pequena unidade do código, como uma classe), **testes de integração** (quando se testa uma unidade de maior granularidade, como um conjunto de classes), **testes de performance** (quando se submete o sistema a uma carga de processamento, para verificar seu desempenho), **testes de usabilidade** (quando o objetivo é verificar a usabilidade da interface do sistema), etc.

Segundo, testes podem ser usados tanto para verificação como para validação de sistemas. Verificação tem como o objetivo garantir que um sistema atende à sua especificação. Já com validação, o objetivo é garantir que um sistema atende às necessidades de seus clientes. A diferença entre os conceitos só faz sentido porque pode ocorrer de a especificação de um sistema não expressar as necessidades de seus clientes. Por exemplo, essa diferença pode ser causada por um erro na fase de levantamento de requisitos; isto é, os desenvolvedores não entenderam os requisitos do sistema ou o cliente não foi capaz de explicá-los precisamente.

Existem duas frases, muito usadas, que resumem as diferenças entre verificação e validação:

- **Verificação:** estamos implementando o sistema corretamente? Isto é, de acordo com seus requisitos.
- **Validação:** estamos implementando o sistema correto? Isto é, aquele que os clientes ou o mercado está querendo.

Assim, quando se realiza um teste de um método, para verificar se ele retorna o resultado especificado, estamos realizando uma atividade de verificação. Por outro lado, quando realizamos um teste funcional e de aceitação, ao lado do cliente, isto é, mostrando para ele os resultados e funcionalidades do sistema, estamos realizando uma atividade de validação.

Terceiro, é importante definir e distinguir três conceitos relacionados a testes: **defeitos**, **bugs** e **falhas**. Para ilustrar a diferença entre eles, suponha o seguinte código para calcular a área de um círculo, dependendo de uma condição:

```
if (condicao)
    area = pi * raio * raio * raio;
```

Esse código possui um defeito, pois a área de um círculo é “pi vezes raio ao quadrado”, e não ao cubo. Bug é um termo mais informal, usado com objetivos às vezes diversos. Mas, o uso mais comum é como sinônimo de defeito. Por fim, uma falha ocorre quando um código com defeito for executado — por exemplo, a condição do `if` do programa acima for verdadeira — e, com isso, levar o programa a apresentar um resultado incorreto. Portanto, nem todo defeito ou bug ocasiona falhas, pois pode ser que o código defeituoso nunca seja executado.

Resumindo: código defeituoso é aquele que não está de acordo com a sua especificação. Se esse código for executado e de fato levar o programa a apresentar um resultado incorreto, diz-se que ocorreu uma falha.

Aprofundamento: Na literatura sobre testes, às vezes são mencionados os termos **erro** e **falta (fault)**. Quando isso ocorre, o significado é o mesmo daquele que adotamos para *defeito* neste livro. Por exemplo, o *IEEE Standard Glossary of Software Engineering Terminology* ([link](#)) define que falta é um “passo, processo ou definição de dados incorretos em um programa de computador; os termos erro e bug são [também] usados para expressar esse significado”. Resumindo, *defeito*, *erro*, *falta* e *bug* são sinônimos.

Mundo Real: Existe uma lista enorme de falhas de software, com consequências graves, tanto em termos financeiros como de vidas humanas. Um dos exemplos mais famosos é a explosão do foguete francês Ariane 5, lançado em 1996, de Kourou, na Guiana Francesa. Cerca de 30 segundos após o lançamento, o foguete explodiu devido a um comportamento inesperado de um dos sistemas de bordo, causando um prejuízo de cerca de meio bilhão de dólares. Interessante, o defeito que causou a falha no sistema de bordo do Ariane 5 foi bem específico, relativamente simples e restrito a poucas linhas de código, implementadas na linguagem de programação ADA, até hoje muito usada no desenvolvimento de software militar e espacial. Essas linhas eram

responsáveis pela conversão de um número real, em ponto flutuante, com 64 bits, para um número inteiro, com 16 bits. Durante os testes e, provavelmente, lançamentos anteriores do foguete, essa conversão sempre foi bem sucedida: o número real sempre “cabia” em um inteiro. Porém, na data da explosão, alguma situação nunca testada previamente exigiu a conversão de um número maior do que o maior inteiro que pode ser representado em 16 bits. Com isso, gerou-se um resultado espúrio, que fez com que o sistema de controle do foguete funcionasse de forma errática, causando a explosão.

1.2.5 Manutenção e Evolução de Software

Assim como sistemas tradicionais de Engenharia, software também precisa de manutenção. Neste livro, vamos usar a seguinte classificação para os tipos de manutenção que podem ser realizadas em sistemas de software: **corretiva, preventiva, adaptativa, refactoring e evolutiva**.

Manutenção corretiva tem como objetivo corrigir bugs reportados por usuários ou outros desenvolvedores.

Por sua vez, manutenção preventiva tem com objetivo corrigir bugs latentes no código, que ainda não causaram falhas junto aos usuários do sistema.

Manutenção adaptativa tem como objetivo adaptar um sistema a uma mudança em seu ambiente, incluindo tecnologia, legislação, regras de integração com outros sistemas ou demandas de novos clientes. Como exemplos de manutenção adaptativa podemos citar:

- A migração de um sistema de Python 2.7 para Python 3.0.
- A customização de um sistema para atender a requisitos de um novo cliente — isto é, quando se instala um sistema em um cliente é comum ter que realizar algumas alterações, para atender a particularidades de seu negócio.
- A adaptação de um sistema para atender a uma mudança de legislação ou outra mudança contextual.

Refactorings são modificações realizadas em um software preservando seu comportamento e visando exclusivamente a melhoria de seu código ou projeto. São exemplos de refactorings operações como renomeação de um método ou variável (para um nome mais intuitivo e fácil de lembrar), divisão de um método longo em dois métodos menores (para facilitar o entendimento) ou movimentação de um método para uma classe mais adequada.

Manutenção evolutiva é aquela realizada para incluir uma nova funcionalidade ou introduzir aperfeiçoamentos importantes em funcionalidades existentes. Sistemas de software podem ser usados por décadas exatamente porque eles sofrem manutenções evolutivas, que preservam o seu valor para os clientes. Por exemplo, diversos sistemas bancários usados hoje em dia foram criados nas décadas de 70 e 80, em linguagens como COBOL. No entanto, eles já sofreram diversas evoluções e melhorias. Hoje, esses sistemas possuem interfaces Web e para celulares, que se integram aos módulos principais, implementados há dezenas de anos.

Sistemas legados são sistemas antigos, baseados em linguagens, sistemas operacionais e bancos de dados tecnologicamente ultrapassados. Por esse motivo, a manutenção desses sistemas costuma ser mais custosa e arriscada. Porém, é importante ressaltar que legado não significa irrelevante, pois muitas vezes esses sistemas realizam operações críticas para seus clientes.

Mundo Real: Um exemplo de manutenção preventiva foram as atividades de manutenção realizadas por diversas empresas antes da virada do último milênio, de 1999 para 2000. Nessa época, diversos sistemas armazenavam o ano de uma data com dois dígitos, isto é, as datas tinham o formato DD-MM-AA. As empresas ficaram receosas de que, em 2000 e nos anos seguintes, algumas operações envolvendo datas retornassem valores incorretos, pois uma subtração 00 - 99, por exemplo, poderia dar um resultado inesperado. As empresas montaram então grupos de trabalho para realizar manutenções em seus sistemas e converter todas as datas para o formato DD-MM-AAAA. Como essas atividades foram realizadas antes da virada do milênio, elas são um exemplo de manutenção preventiva.

Aprofundamento: Na literatura, existem classificações alternativas para os tipos de manutenção de software. Uma delas, proposta por Lientz & Swanson, em 1978 ([link](#)), classifica manutenção nas seguintes categorias: (1) Corretiva, exatamente como usado e definido neste livro; (2) Perfectiva, refere-se à adição de novas funcionalidades; neste livro, optamos por chamá-la de manutenção evolutiva; (3) Adaptativa, refere-se a mudanças no ambiente operacional do software, como um novo hardware ou sistema operacional; logo, não inclui, por exemplo, customizações para novos clientes, como proposto neste livro; (4) Preventiva, refere-se a mudanças que visam incrementar a manutenibilidade

de um sistema; neste livro, optamos pelo termo mais comum hoje em dia, que é refactoring, e que iremos estudar no Capítulo 9.

1.2.6 Gerência de Configuração

Atualmente, é inconcebível desenvolver um software sem um sistema de controle de versões, como git. Esses sistemas armazenam todas as versões de um software, não só do código fonte, mas também de documentação, manuais, páginas web, relatórios, etc. Eles também permitem restaurar uma determinada versão. Por exemplo, se foi realizada uma mudança no código que introduziu um bug crítico, pode-se com relativa facilidade recuperar e retornar para a versão antiga, anterior à introdução do bug.

No entanto, gerência de configuração é mais do que apenas usar um sistema com o git. Ela inclui a definição de um conjunto de políticas para gerenciar as diversas versões de um sistema. Por exemplo, preocupa-se com o esquema usado para identificar as releases de um software; isto é, as versões de um sistema que serão liberadas para seus clientes finais. Um time de desenvolvedores pode definir que as releases de uma determinada biblioteca que eles estão desenvolvendo serão identificadas no formato $x.y.z$, onde x , y e z são inteiros. Um incremento em z ocorre quando se lança uma nova release com apenas correções de bugs (normalmente, chamada de *patch*); um incremento em y ocorre quando se lança uma release da biblioteca com pequenas funcionalidades (normalmente, chamada de versão *minor*); por fim, um incremento em x ocorre quando se lança uma release com funcionalidades muito diferentes daquelas da última release (normalmente, chamada de versão *major*). Esse esquema de numeração de releases é conhecido como **versionamento semântico**.

1.2.7 Gerência de Projetos

Desenvolvimento de software requer o uso de práticas e atividades de gerência de projetos, por exemplo, para negociação de contratos com clientes (com definição de prazos, valores, cronogramas, etc), gerência de recursos humanos (incluindo contratação, treinamento, políticas de promoção, remuneração, etc), gerência de riscos, acompanhamento da concorrência, marketing, finanças, etc. Em um projeto, normalmente usa-se o termo **stakeholder** para designar todas as partes interessadas no mesmo; ou seja, os stakeholders são aqueles que afetam ou que são afetados pelo projeto, podendo ser pessoas físicas ou organizações. Por exemplo, stakeholders comuns em projetos de software incluem, obviamente, seus desenvolvedores e seus clientes; mas também, gerentes da equipe de desenvolvimento,

empresas subcontratadas, fornecedores de qualquer natureza, talvez algum nível de governo, etc.

Existe uma frase muito conhecida, também de Frederick Brooks, que captura uma peculiaridade de projetos de software. Segundo Brooks:

“A inclusão de novos desenvolvedores em um projeto que está atrasado contribui para torná-lo ainda mais atrasado.”

Essa frase ficou tão famosa, que ela é hoje conhecida como **Lei de Brooks**. Basicamente, esse efeito acontece porque os novos desenvolvedores terão primeiro que entender e compreender todo o sistema, sua arquitetura e seu projeto (*design*), antes de começarem a produzir código útil. Além disso, equipes maiores implicam em um maior esforço de comunicação e coordenação para tomar e explicar decisões. Por exemplo, se um time tem 3 desenvolvedores (d_1, d_2, d_3), existem 3 canais de comunicação possíveis (d_1-d_2, d_1-d_3 e d_2-d_3); se ele cresce para 4 desenvolvedores, o número de canais duplica, para 6 canais. Se ele cresce para 10 desenvolvedores, passam a existir 45 canais de comunicação. Por isso, modernamente, software é desenvolvido em times pequenos, com uma dezena de engenheiros, se tanto.

Tradução: Em Português, a palavra *projeto* pode se referir tanto a *design* como a *project*. Por exemplo, em uma subseção anterior introduzimos questões de projeto de software, isto é, *software design*, tratando de conceitos como interfaces, dependências, arquitetura, etc. Na presente seção, acabamos de discutir questões de gerência de projetos de software, isto é, *software project management*, tais como prazos, contratos, Lei de Brooks, etc. No restante deste livro, iremos traduzir apenas o uso mais comum em cada capítulo e manter o uso menos comum em inglês. Por exemplo, no Capítulo 2 (Processos de Desenvolvimento), usaremos projeto com tradução de *project*, pois é o uso mais comum neste capítulo. Já no Capítulo 5 (Princípios de Projeto) e no Capítulo 6 (Padrões de Projeto), *design* será traduzido para projeto, pois é o uso mais comum nesses capítulos, aparecendo inclusive no título dos mesmos.

Aprofundamento: A Lei de Brooks foi proposta em um livro clássico do autor sobre gerenciamento de projetos de software, chamado *The Mythical Man-Month*, cuja primeira edição foi publicada em 1975 ([link](#)). Nesse livro, Brooks reporta as lições que aprendeu no início da sua carreira, como gerente responsável pelos primeiros sistemas operacionais da IBM. Em 1995, uma

segunda edição do livro foi lançada, em comemoração aos seus 20 anos. Essa edição incluiu um novo capítulo, com o artigo *No Silver Bullet Essence and Accidents of Software Engineering*, publicado originalmente em 1987 (e que já comentamos nesta Introdução). Em 1999, Frederick Brooks ganhou o Prêmio Turing, considerado o Prêmio Nobel da Computação.

1.2.8 Processos de Desenvolvimento de Software

Um processo de desenvolvimento define quais atividades e etapas devem ser seguidas para construir e entregar um sistema de software. Uma analogia pode ser feita, por exemplo, com a construção de prédios, que ocorre de acordo com algumas etapas: fundação, alvenaria, cobertura, instalações hidráulicas, instalações elétricas, acabamentos, pintura, etc.

Historicamente, existem dois grandes tipos de processos que podem ser adotados na construção de sistemas de software:

- **Processos Waterfall** (ou em cascata)
- **Processos Ágeis** (ou incrementais ou iterativos).

Processos Waterfall foram os primeiros a serem propostos, ainda na década de 70, quando a Engenharia de Software começava a ganhar envergadura. De forma compreensível, eles foram inspirados nos processos usados em engenharias tradicionais, os quais são largamente sequenciais, como ilustrado no exemplo do prédio, usado no parágrafo inicial desta seção. Processos Waterfall foram muito usados até a década de 1990 e grande parte desse sucesso deve-se a uma padronização lançada pelo Departamento de Defesa Norte-Americano, em 1985. Basicamente, eles estabeleceram que todo software comprado ou contratado pelo Departamento de Defesa deveria ser construído usando Waterfall.

Processos Waterfall — também chamados de **processos dirigidos por planejamento** (*plan-driven*) — propõem que a construção de um sistema deve ser feita em etapas sequenciais, como em uma cascata de água, onde a água vai escorrendo de um nível para o outro. Essas etapas são as seguintes: levantamento de requisitos, análise (ou projeto de alto nível), projeto detalhado, codificação e testes. Finalizado esse pipeline, o sistema é liberado para produção, isto é, para uso efetivo pelos seus usuários, conforme ilustrado na próxima figura.

No entanto, processos Waterfall, a partir do final da década de 90, passaram a ser muito criticados, devido aos atrasos e problemas recorrentes em projetos

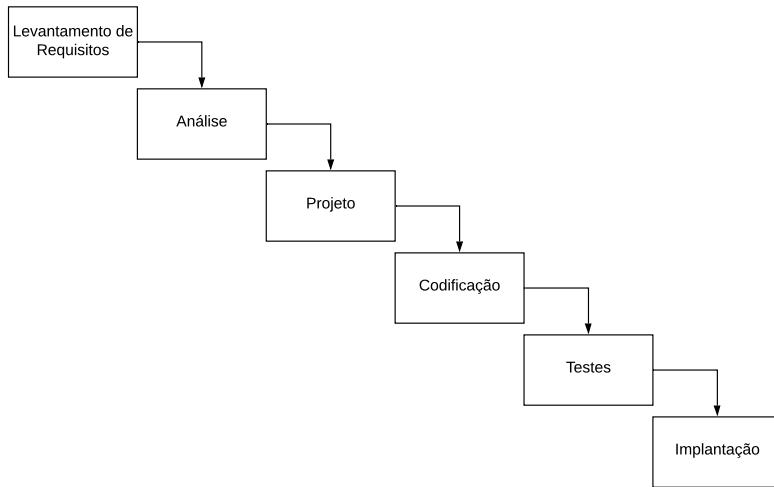


Figura 1.2: Fases de um processo Waterfall.

de software, que ocorriam com frequência nessa época. O principal problema é que Waterfall pressupõe um levantamento completo de requisitos, depois um projeto detalhado, depois uma implementação completa, etc. Para só então validar o sistema com os usuários, o que pode acontecer anos após o início do projeto. No entanto, neste período de tempo, o mundo pode ter mudado, bem como as necessidades dos clientes, que podem não mais precisar do sistema que ajudaram a especificar anos antes. Assim, reunidos em uma cidade de Utah, Estados Unidos, em fevereiro de 2001, um grupo de 17 Engenheiros de Software propôs um modo alternativo para construção de software, que eles chamaram de Ágil — nome do manifesto que eles produziram nesta reunião ([link](#)). Contrastando com processos Waterfall, a ideia de processos ágeis é que um sistema seja construído de forma incremental e iterativa. Pequenos incrementos de funcionalidade são produzidos, em intervalos de cerca de um mês e, logo em seguida, validados pelos usuários. Uma vez que o incremento produzido esteja aprovado, o ciclo se repete.

Processos ágeis tiveram um profundo impacto na indústria de software. Hoje, eles são usados pelas mais diferentes organizações que produzem software, desde pequenas empresas até as grandes companhias da Internet. Diversos métodos que concretizam os princípios ágeis foram propostos, tais como **XP**, **Scrum**, **Kanban** e **Lean Development**. Esses métodos também ajudaram a disseminar diversas práticas de desenvolvimento de software, como **testes automatizados**, **test-driven development** (isto é, escrever os testes primeiro, antes do próprio

código) e **integração contínua** (*continuous integration*). Integração contínua recomenda que desenvolvedores integrem o código que produzem imediatamente, se possível todo dia. O objetivo é evitar que desenvolvedores fiquem muito tempo trabalhando localmente, em sua máquina, sem integrar o código que estão produzindo no repositório principal do projeto. Quando o time de desenvolvimento é maior, isso aumenta as chances de conflitos de integração, que ocorrem quando dois desenvolvedores alteram em paralelo os mesmos trechos de código. O primeiro desenvolvedor a integrar seu código será bem sucedido; enquanto que o segundo desenvolvedor será informado de que o trecho já foi modificado pelo primeiro.

1.2.9 Modelos de Software

Um modelo oferece uma representação em mais alto nível de um sistema do que o seu código fonte. O objetivo é permitir que desenvolvedores possam analisar propriedades e características essenciais de um sistema, de modo mais fácil e rápido, sem ter que mergulhar nos detalhes do código.

Modelos podem ser criados antes do código, por exemplo, ainda na fase de projeto. Nesse caso, eles são usados para apoiar **Engenharia Avante** (*Forward Engineering*); isto é, primeiro cria-se um modelo para ter um entendimento de mais alto nível de um sistema, antes de partir para a implementação do código. Por outro lado, eles podem ser criados para ajudar a entender uma base de código existente; nesse caso, eles são um instrumento de **Engenharia Reversa** (*Reverse Engineering*). Em ambos os casos, modelos são uma forma de documentar o código de um sistema.

Frequentemente, modelos de software são baseados em notações gráficas. Por exemplo, **UML** (*Unified Modelling Language*) é uma notação que define mais de uma dezena de diagramas gráficos para representar propriedades estruturais e comportamentais de um sistema. Na próxima figura, mostra-se um diagrama UML — chamado Diagrama de Classes — para o exemplo de código usado na seção sobre Projeto de Software. Nesse diagrama, as caixas retangulares representam classes do sistema, incluindo seus atributos e métodos. As setas são usadas para denotar relacionamentos entre as classes. Existem editores para criar diagramas UML, que podem ser usados, por exemplo, em um cenário de Engenharia Avante.

1.2.10 Qualidade de Software

Qualidade é um objetivo recorrente em produtos de engenharia. Fabricantes de automóveis, celulares, computadores, empresas de construção civil, etc, todos almejam e dizem que possuem produtos de qualidade. Esse contexto não é diferente quando o produto em questão é um software. Segundo uma classificação

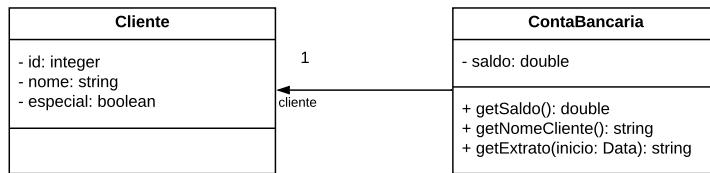


Figura 1.3: Exemplo de Diagrama de Classe UML com duas classes, `Cliente` e `ContaBancaria`.

inicialmente proposta por Bertrand Meyer [link](#), qualidade de software pode ser avaliada em duas dimensões: **qualidade externa** ou **qualidade interna**.

Qualidade externa considera fatores que podem ser aferidos sem analisar o código de um sistema. Assim, a qualidade externa de um software pode ser avaliada mesmo por usuários comuns, que não são especialistas em Engenharia de Software. Como exemplo, temos os seguintes fatores (ou atributos) de qualidade externa:

- Correção: o software atende à sua especificação? Nas situações normais, ele funciona como esperado?
- Robustez: o software continua funcionando mesmo quando ocorrem eventos anormais, como uma falha de comunicação ou de disco? Por exemplo, um software robusto não pode sofrer um *crash* (abortar) caso tais eventos anormais ocorram. Ele deve pelo menos avisar por qual motivo não está conseguindo funcionar conforme previsto.
- Eficiência: o software faz bom uso de recursos computacionais? Ou ele precisa de um hardware extremamente poderoso e caro para funcionar?
- Portabilidade: é possível portar esse software para outras plataformas e sistemas operacionais? Ele, por exemplo, possui versões para os principais sistemas operacionais, como Windows, Linux e Mac OS? Ou então, se for um app, ele possui versões para Android e iOS.
- Facilidade de Uso: o software possui uma interface amigável, mensagens de erro claras, suporta mais de uma língua, etc? Pode ser também usado por pessoas com alguma deficiência, como visual ou auditiva?
- Compatibilidade: o software é compatível com os principais formatos de dados de sua área? Por exemplo, se o software for uma planilha eletrônica, ele importa arquivos em formatos XLS e CSV?

Por outro lado, qualidade interna considera propriedades e características relacionadas com a implementação de um sistema. Portanto, a qualidade interna de um sistema somente pode ser avaliada por um especialista em Engenharia de Software e não por usuários leigos. São exemplos de fatores (ou atributos) de qualidade interna: modularidade, legibilidade do código, manutenibilidade e testabilidade.

Para garantir qualidade de software, diversas estratégias podem ser usadas. Primeiro, **métricas** podem ser usadas para acompanhar o desenvolvimento de um produto de software, incluindo métricas de código fonte e métricas de processo. Um exemplo de métrica de código é o número de linhas de um programa, que pode ser usado para dar uma ideia de seu tamanho. Métricas de processo incluem, por exemplo, o número de defeitos reportados em produção por usuários finais em um certo intervalo de tempo.

Existem ainda práticas que podem ser adotadas para garantir a produção de software com qualidade. Modernamente, por exemplo, diversas organizações usam **revisões de código**, isto é, o código produzido por um desenvolvedor somente entra em produção depois de ser revisado e inspecionado por um outro desenvolvedor do time. O objetivo é detectar possíveis *bugs* antecipadamente, antes de o sistema entrar em produção. Além disso, revisões de código servem para garantir a qualidade interna do código — isto é, sua manutenibilidade, legibilidade, modularidade, etc — e para disseminar boas práticas de Engenharia de Software entre os membros de um time de desenvolvimento.

A próxima figura mostra um exemplo de revisão de código, referente a um exemplo que usamos na seção sobre Testes de Software. Assumindo que a empresa que produziu esse código adotasse revisões de código, ele teria que ser analisado por um outro desenvolvedor, chamado de revisor, antes de entrar em produção. Esse revisor poderia perceber o bug e anotar o código com uma dúvida, antes de aprová-lo. Em seguida, o responsável pelo código poderia concordar que, de fato, existe um bug, corrigir o código e submetê-lo de novo para revisão. Finalmente, ele seria aprovado pelo revisor. Existem diversas ferramentas para apoiar processos de revisão de código. No exemplo, usamos a ferramenta fornecida pelo GitHub.

1.2.11 Prática Profissional

Como afirmado na frase de Bjarne Stroustrup que abre este capítulo, *nossa sociedade funciona a base de software*. Isso gera diversas oportunidades para os profissionais da área, mas também implica em responsabilidades e pontos de preocupação. Questões sobre a prática profissional em Engenharia de Software iniciam-se no momento da formação, em nível de graduação, envolvendo a definição de currículos de referência e a necessidade de cursos específicos para a área,

feature1

```
...     ... @@ -0,0 +1,2 @@
1  + if (enabled)
2  +     area = pi * raio * raio * raio;
```

 **mtov** 12 minutes ago Author Owner

Não seria $\pi * raio * raio$?

Figura 1.4: Exemplo de revisão de código, com um comentário procurando esclarecer um possível bug detectado no código.

que constituam alternativas aos cursos de Ciência da Computação, Sistemas de Informação e Engenharia de Computação. Não menos importante, existem também questões sobre a formação em nível técnico e tecnológico, anterior à formação universitária. Após a etapa de formação, existem questões sobre a regulamentação da profissão, por exemplo.

Por fim, mas muito atual e relevante, existem questionamentos sobre o papel e a **responsabilidade ética** dos profissionais formados em Computação, em uma sociedade onde os relacionamentos humanos são cada vez mais mediados por algoritmos e sistemas de software. Neste sentido, as principais sociedades científicas da área possuem códigos que procuram ajudar os profissionais de Computação — não necessariamente apenas Engenheiros de Software — a exercer seu ofício de forma ética. Como exemplos, temos o Código de Ética da ACM ([link](#)) e da IEEE Computer Society ([link](#)). Esse último é interessante porque é específico para a prática de Engenharia de Software. Por exemplo, ele prescreve que:

“Engenheiros de Software devem se comprometer em fazer da análise, especificação, projeto, desenvolvimento, teste e manutenção de software uma profissão benéfica e respeitada.”

No Brasil, existe o Código de Ética da Sociedade Brasileira de Computação (SBC), que por ser sintético, mas ao mesmo tempo claro, resolvemos reproduzir a seguir:

São deveres dos profissionais de Informática:

Art. 1º: Contribuir para o bem-estar social, promovendo, sempre que possível, a inclusão de todos setores da sociedade.

Art. 2º: Exercer o trabalho profissional com responsabilidade, dedicação, honestidade e justiça, buscando sempre a melhor solução.

Art. 3º: Esforçar-se para adquirir continuamente competência técnica e profissional, mantendo-se sempre atualizado com os avanços da profissão.

Art. 4º: Atuar dentro dos limites de sua competência profissional e orientar-se por elevado espírito público.

Art. 5º: Guardar sigilo profissional das informações a que tiver acesso em decorrência das atividades exercidas.

Art. 6º: Conduzir as atividades profissionais sem discriminação, seja de raça, sexo, religião, nacionalidade, cor da pele, idade, estado civil ou qualquer outra condição humana.

Art. 7º: Respeitar a legislação vigente, o interesse social e os direitos de terceiros.

Art. 8º: Honrar compromissos, contratos, termos de responsabilidade, direitos de propriedade, copyrights e patentes.

Art. 9º: Pautar sua relação com os colegas de profissão nos princípios de consideração, respeito, apreço, solidariedade e da harmonia da classe.

Art. 10º: Não praticar atos que possam comprometer a honra, a dignidade e privacidade de qualquer pessoa.

Art. 11º: Nunca apropriar-se de trabalho intelectual, iniciativas ou soluções encontradas por outras pessoas.

Art. 12º: Zelar pelo cumprimento deste código.

– Código de Ética da Sociedade Brasileira de Computação (SBC, 2013) ([link](#))

Mundo Real: O Stack Overflow realiza anualmente um survey com usuários da plataforma de perguntas e respostas. Em 2018, esse survey foi respondido por mais de 100 mil desenvolvedores, dos mais variados países. Dentre as perguntas, um grupo se referia a questões éticas ([link](#)). Uma delas perguntava se desenvolvedores têm a obrigação de considerar as implicações éticas do código que produzem. Quase 80% dos respondentes disseram que sim. Uma outra pergunta foi a seguinte: Quem, em última análise, é responsável por um código que colabora para um comportamento antiético? Nesse caso, 57.5% responderam que é a alta gerência da organização ou empresa, enquanto

que 23% disseram que é o próprio desenvolvedor. Quando perguntados se concordariam em escrever um código com dúvidas éticas, 58% responderam que não e 37% responderam que dependeria do código requisitado.

1.2.12 Aspectos Econômicos

Diversas decisões e questões econômicas se entrelaçam com o desenvolvimento de sistemas. Por exemplo, uma startup de software deve decidir qual o modelo de rentabilização pretende adotar, se baseado em assinaturas ou em anúncios. Desenvolvedores de apps para celulares têm que decidir sobre o preço que irão cobrar pela sua aplicação, o que, dentre outras variáveis, requer conhecimento sobre o preço das apps concorrentes. Por isso, não é surpresa que grandes companhias de software atualmente empreguem economistas, para avaliarem os aspectos econômicos dos sistemas que produzem.

Para discutir um caso mais concreto, em economia existe uma preocupação frequente com os custos de oportunidade de uma decisão. Isto é, toda decisão possui um custo de oportunidade, que são as oportunidades preteridas quando se descartou uma das decisões alternativas; em outras palavras, quando se descarta uma decisão Y em detrimento de uma decisão X, os eventuais benefícios de Y passaram a ser oportunidades perdidas. Por exemplo, suponha que o principal sistema de sua empresa tenha uma lista de bugs B para ser corrigida. Existem benefícios em corrigir B? Claro, isso vai deixar os clientes atuais mais satisfeitos; eles não vão pensar em migrar para sistemas concorrentes, etc. Porém, existe também um custo de oportunidade nessa decisão. Especificamente, em vez de corrigir B, a empresa poderia investir em novas funcionalidades F, que poderiam ajudar a ampliar a base de clientes. O que é melhor? Corrigir os bugs ou implementar novas funcionalidades? No fundo, essa é uma decisão econômica.

1.3 Sistemas ABC: Acute, Business e Casuais

Atualmente, como estamos ressaltando nesta Introdução, software permeia as mais distintas atividades humanas. Ou seja, temos software de todos os tamanhos, em todas as atividades, com os mais diferentes requisitos funcionais e não-funcionais, desenvolvidos por 1-2 desenvolvedores ou por grandes corporações da Internet, etc. O risco é então achar que existe um único modo de desenvolver software. Em outras palavras, que todo software deve ser construído usando o mesmo processo de desenvolvimento, os mesmos princípios de projeto, os mesmos mecanismos de garantia de qualidade, etc.

Uma classificação proposta por Bertrand Meyer ([link](#)) ajuda a distinguir e entender os diferentes sistemas de software que podem ser construídos e os princípios de Engenharia de Software mais recomendados para cada uma das categorias propostas. Segundo essa classificação, existem três tipos principais de software:

- **Sistemas A** (Acute)
- **Sistemas B** (Business)
- **Sistemas C** (Casuais)

Vamos discutir primeiro os Sistemas C e A (isto é, os sistemas em cada um dos extremos da classificação) e depois os Sistemas B.

Sistemas C (Casuais) não sofrem pressão para terem níveis altos de qualidade. São sistemas que podem ter alguns bugs, os quais não vão comprometer fundamentalmente o seu funcionamento. Como exemplo, podemos citar um script feito para um trabalho acadêmico, um programa de conversão de dados (que vai ser usado uma única vez, para converter os dados para um novo banco de dados que está sendo comprado pela empresa), um sistema para controlar os sócios do Diretório Acadêmico da universidade, um sistema para gerenciar as salas disponíveis para reuniões em uma empresa, etc. Por isso, Sistemas C não precisam ter níveis altos de qualidade interna; por exemplo, podem ter parte do código duplicado. Também não precisam ter desempenho ou uma boa interface. Em geral, são desenvolvidos por 1-2 programadores; ou seja, são sistemas pequenos e não críticos. Por tudo isso, eles não se beneficiam tanto das práticas, técnicas e processos estudados neste livro. Pelo contrário, no caso de Sistemas C, o maior risco é **over-engineering**, ou seja, o uso de recursos mais sofisticados em um contexto que não demanda tanta preocupação. Como se diz coloquialmente, Engenharia de Software nesse contexto equivale a “usar uma bala de canhão, para matar formigas”.

No outro extremo, temos os Sistemas A (de *acute*, ou de missão crítica). São sistemas onde qualquer falha pode causar um imenso prejuízo, incluindo a perda de vidas humanas. São sistemas para controlar um carro autônomo, uma usina nuclear, um avião, os equipamentos de uma UTI, um trem de metrô, etc. O exemplo do sistema de controle do foguete Ariane 5, usado na seção sobre Testes de Software, é um exemplo de Sistema A. O desenvolvimento desses sistemas deve ser feito de acordo com processos rígidos, incluindo rigorosa revisão de código e certificação por organizações externas. É comum exigir redundância não apenas em hardware, mas também no próprio software. Por exemplo, o sistema roda de forma paralela em duas máquinas e uma decisão somente é tomada caso ambas

instâncias cheguem ao mesmo resultado. Por fim, esses sistemas muitas vezes são especificados em uma linguagem formal, baseada em teoria de conjuntos ou lógica.

Aviso: Por tudo que foi afirmado no parágrafo anterior, **sistemas A (isto é, de missão crítica) não serão tratados neste livro.**

Sobram os sistemas B (Business), que são exatamente aqueles que vão se beneficiar dos conceitos estudados neste livro. Esses sistemas incluem as mais variadas aplicações corporativas (financeiras, recursos humanos, logística, vendas, contabilidade, etc), sistemas Web dos mais variados tipos, desde sistemas com poucas páginas até grandes redes sociais ou sistemas de busca. Outras aplicações incluem bibliotecas e frameworks de software, aplicações de uso geral (como editores de texto, planilhas, editores de imagens, etc) e sistemas de software básico (como compiladores, gerenciadores de bancos de dados, IDEs, etc). Nesses sistemas, as técnicas de Engenharia de Software estudadas neste livro podem contribuir com dois benefícios principais: (1) elas podem tornar mais produtivo o desenvolvimento de Sistemas B; (b) elas podem propiciar a construção de Sistemas B com melhor qualidade, tanto interna (por exemplo, sistemas mais fáceis de serem mantidos) como externa (por exemplo, sistemas com menor quantidade de bugs em produção).

1.4 Próximos Capítulos

Este livro possui **10 capítulos e um apêndice**:

Capítulo 2: Processos, com foco em processos ágeis de desenvolvimento, especificamente XP, Scrum e Kanban. Tomamos a decisão de focar em métodos ágeis porque eles são largamente usados hoje em dia no desenvolvimento dos mais variados tipos de sistemas, dos mais variados domínios e tamanhos. Tratamos também de processos tradicionais, como Waterfall e o Processo Unificado, porém de forma resumida e, também, para fazer o contraste com métodos ágeis.

Capítulo 3: Requisitos, que inicia com uma discussão sobre a importância de requisitos e os principais tipos de requisitos. Então, apresentamos duas técnicas para levantamento e validação de requisitos: Histórias de Usuário (usadas com métodos ágeis) e Casos de Uso (uma técnica tradicional, que é mais usada com métodos dirigidos por planejamento e documentação). Por fim, apresentamos dois assuntos que, apesar de importantes e atuais, não são ainda tratados nos livros tradicionais: Produto Mínimo Viável (MVPs) e Testes A/B. Argumentamos que

esses dois conceitos não são importantes apenas em startups, mas também em empresas que desenvolvem software para mercados mais estáveis.

Capítulo 4: Modelos, que tem foco no uso de UML para elaboração de esboços (*sketches*) de software. Modernamente, concordamos que UML não é mais usada para os fins que ela foi concebida na década de 90, ou seja, para criação de modelos detalhados de software. Praticamente, não existem mais casos de empresas que investem meses — ou anos — na elaboração de diagramas gráficos antes de começar a implementar qualquer linha de código. Porém, se não tratássemos de UML no livro ficaríamos com a sensação de que “após o banho, jogamos o bebê fora, junto com a água da bacia”. Se por um lado não faz sentido estudar todos os diagramas da UML em detalhes, por outro lado existem elementos importantes em alguns desses diagramas. Além disso, desenvolvedores, com frequência, elaboram pequenos esboços de software, por exemplo, para comunicar e discutir ideias de design com outros desenvolvedores. Assim, conhecimento básico de UML pode ser interessante para criar esses esboços, inclusive para evitar a criação de uma nova linguagem de modelagem.

Capítulo 5: Princípios de Projeto, que trata de dois temas que devem ser do conhecimento de todo projetista de software. São eles: (1) propriedades (ou considerações) importantes em projeto de software, incluindo integridade conceitual, ocultamento de informação, coesão e acoplamento; (2) princípios de projeto, os quais constituem recomendações mais específicas para construção de bons projetos de software, tais como responsabilidade única, prefira composição a herança, aberto/fechado, Demeter, etc.

Capítulo 6: Padrões de Projeto, os quais constituem um catálogo de soluções para problemas comuns de projeto de software. Neste capítulo, vamos estudar os principais padrões de projeto definidos no livro clássico sobre o tema. A discussão de cada padrão será dividida em três partes: (1) um contexto, isto é, um sistema onde o padrão pode ser útil; (2) um problema no projeto desse sistema; (3) uma solução para esse problema por meio de padrões. Iremos também apresentar diversos exemplos de código, para facilitar o entendimento e a discussão prática do uso de cada padrão. O código completo de alguns exemplos mais complexos será disponibilizado no GitHub.

Capítulo 7: Arquitetura, que inicia com uma apresentação e discussão sobre Arquitetura de Software. O objetivo é deixar claro que arquitetura deve ser vista como projeto em alto nível, envolvendo pacotes, camadas ou serviços, em vez de classes individuais. Em seguida, discutimos cinco padrões arquiteturais: arquitetura em camadas (incluindo 3-camadas), arquitetura MVC (incluindo

single-page applications), microserviços, arquiteturas orientadas por filas de mensagens e arquiteturas publish/subscribe. Essas duas últimas são comuns na construção de sistemas distribuídos fracamente acoplados. Por fim, apresentamos um anti-padrão arquitetural, chamado *big ball of mud*, que é um termo usado para designar sistemas sem organização arquitetural. Esses sistemas poderiam até possuir uma arquitetura no seu início, mas depois o projeto arquitetural deles foi sendo abandonado, transformando os sistemas em um “spaghetti” de dependências entre os seus módulos.

Capítulo 8: Testes, com ênfase em testes de unidade, usando frameworks como o JUnit. O capítulo inclui dezenas de exemplos de testes de unidade e também discute diversos aspectos desses testes. Por exemplo, discutimos bons princípios para escrita de testes de unidade e também test smells, isto é, padrões de testes que não são recomendados. Em seguida, tratamos de testabilidade, isto é, discutimos a importância de escrever código que possa ser facilmente testado. O capítulo inclui uma seção inteira sobre mocks e stubs, os quais são objetos que viabilizam o teste de unidade de código com dependências mais complexas, como dependências para bancos de dados e outros sistemas externos. Finalizada a discussão sobre testes de unidade, também discutimos, porém de forma mais resumida, dois outros tipos de testes: testes de integração e testes de sistema. Esses testes verificam propriedades de unidades maiores de código, como as classes responsáveis por um determinado serviço ou funcionalidade (testes de integração) ou mesmo todas as classes de um sistema (testes de sistema). Para terminar, incluímos uma discussão sobre outros testes, como testes caixa preta (ou testes funcionais), testes caixa branca (ou testes estruturais), testes de aceitação e também testes para verificar requisitos não-funcionais, como desempenho, falhas e usabilidade.

Capítulo 9: Refactoring, cujo principal conteúdo é uma apresentação dos principais refactorings que podem ser realizados para melhor a qualidade interna de um sistema de software. A apresentação inclui vários exemplos de código fonte, alguns deles de refactorings reais, realizados em sistemas de código aberto. O objetivo é transmitir ao leitor uma experiência prática de refatoração, que o ajude a desenvolver o hábito de frequentemente alocar tempo para melhor a qualidade interna do código que ele vai desenvolver. No capítulo, também apresentamos uma lista de code smells, isto é, indicadores de que uma estrutura de código não está “cheirando bem” e que, portanto, poderia ser objeto de uma refatoração.

Capítulo 10: DevOps, que é um movimento que tenta aproximar os times de desenvolvimento (Devs) e de operações (Ops) de uma empresa desenvolvedora de software. O time de operações é responsável por manter o software em funcionamento, sendo formado por administradores de rede, administradores de bancos

de dados, técnicos de suporte, etc. Em uma cultura tradicional, esses dois times tendem a atuar de forma independente. Ou seja, o time de desenvolvimento desenvolve o sistema e depois “joga ele por cima da parede” (*throw it over the wall*) que separa o departamento de desenvolvimento do departamento de operações. Ou seja, os dois times não conversam nem no momento da “passagem de bastão” de uma área para outra. Para resolver esse problema, DevOps propõe uma interação constante entre as áreas Devs e Ops, desde os primeiros dias do desenvolvimento. O objetivo é acelerar a entrada em produção de um sistema. Além de uma introdução a DevOps, vamos estudar no Capítulo 9 algumas práticas essenciais quando uma empresa adota essa cultura, incluindo Controle de Versões, Integração Contínua e Deployment/Entrega Contínua.

Apêndice A: Git, que apresenta e mostra exemplos de uso dos principais comandos do sistema git. Atualmente, é inconcebível não usar controle de versões em qualquer sistema, mesmo naqueles mais simples. Por isso, fizemos questão de acrescentar esse apêndice no livro. Git é o sistema de controle de versões mais usado atualmente.

Bibliografia

Pierre Bourque, Richard Fairley. Guide to the Software Engineering Body of Knowledge, Version 3.0, IEEE Computer Society, 2014.

Armando Fox, David Patterson. Construindo Software como Serviço: Uma Abordagem Ágil Usando Computação em Nuvem. Strawberry Canyon LLC. 1a edição, 2014.

Frederick Brooks. O Mítico Homem-Mês. Ensaios Sobre Engenharia de Software. Alta Books, 1a edição, 2018.

Exercícios de Fixação

1. Segundo Frederick Brooks, desenvolvimento de software enfrenta dificuldades essenciais (para as quais não há bala de prata) e accidentais (para as quais existe uma solução melhor). Dê um exemplo de dificuldade accidental que já tenha experimentado ao desenvolver programas, mesmo que pequenos. Sugestão: elas podem estar relacionadas a ferramentas que tenha usado, como compiladores, IDEs, bancos de dados, sistemas operacionais, etc.
2. Diferencie requisitos funcionais de requisitos não-funcionais.
3. Explique porque testes podem ser considerados tanto uma atividade de verificação como de validação de software. Qual tipo de teste é mais adequado se

o objetivo for verificação? Qual tipo de teste é mais adequado se o objetivo for validar um sistema de software?

4. Por que testes não conseguem provar a *ausência* de bugs?
5. Suponha um programa que tenha uma única entrada: um inteiro de 64 bits. Em um teste exaustivo, temos que testar esse programa com todos os possíveis inteiros (logo, 2^{64}). Se cada teste levar 1 nanosegundo (10^{-9} segundos), quanto tempo levará esse teste exaustivo? (Exercício baseado em um comentário do livro de Fox & Patterson, [link](#))
6. Se considerarmos seu contexto histórico, por que foi natural que os primeiros processos de desenvolvimento de software tivessem características sequenciais e que fossem baseados em planejamento e documentação detalhados?
7. Alguns estudos mostram que os custos com manutenção e evolução podem alcançar 80% ou mais dos custos totais alocados a um sistema de software, durante todo o seu ciclo de vida. Explique porque esse valor é tão alto.
8. Refactoring é normalmente definido como uma transformação de código que preserva comportamento. Qual o significado da expressão *preservar comportamento*? Na prática, qual restrição ela impõe a uma operação de refactoring?
9. Dê exemplos de sistemas A (*Acute*, ou críticos) e B (*Business*, ou comerciais) com os quais já tenha interagido.
10. Dê exemplos de sistemas C (casuais) que você já tenha desenvolvido.
11. Em 2015, descobriu-se que o software instalado em mais de 11 milhões de carros da Volkswagen detectava quando eles estavam sendo testados em um laboratório de certificação. Nessas situações, o carro emitia poluentes dentro das normas legais. Fora do laboratório, emitia-se mais poluentes, para melhorar o desempenho. Ou seja, o código provavelmente incluía uma estrutura de decisão como a seguinte (meramente ilustrativa, para fins deste exercício):

```
if "Carro sendo testado em um laboratório"
    "Emita poluentes dentro das normas"
else
    "Emita poluentes fora das normas"
```

O que você faria se seu chefe pedisse para escrever um *if* como o acima? (para mais informações sobre esse episódio com automóveis Volkswagen, consulte essa página da [Wikipedia](#)).

Capítulo 2

Processos

In software development, perfect is a verb, not an adjective. There is no perfect process. There is no perfect design. There are no perfect stories. You can, however, perfect your process, your design, and your stories. – Kent Beck

Este capítulo inicia com uma apresentação sobre a importância de processos de software (Seção 2.1). Em seguida, discutimos questões gerais e preliminares sobre processos ágeis de desenvolvimento de software (Seção 2.2), incluindo uma apresentação sobre o contexto histórico que motivou o surgimento desse tipo de processo. As próximas seções tratam de três métodos ágeis: Extreme Programming (Seção 2.3), Scrum (Seção 2.4) e Kanban (Seção 2.5). Depois, temos uma seção dedicada a discutir quando métodos ágeis não são recomendados (Seção 2.6). Por fim, na Seção 2.7, discutimos alguns processos tradicionais, principalmente o Processo Unificado.

2.1 Importância de Processos

A produção de um carro em uma fábrica de automóveis segue um processo bem definido. Sem estender muito a explicação, primeiro, as chapas de aço são cortadas e prensadas, para ganhar a forma de portas, tetos e capôs. Depois, o carro é pintado e instalaram-se painel, bancos, cintos de segurança e toda a fiação. Por fim, instala-se a parte mecânica, incluindo motor, suspensão e freios.

Assim como carros, software também é produzido de acordo com um **processo**, embora certamente menos mecânico e mais dependente de esforço intelectual. Um

processo de desenvolvimento de software define um conjunto de passos, tarefas, eventos e práticas que devem ser seguidos por desenvolvedores de software, na produção de um sistema.

Alguns críticos de processos de software costumam fazer a seguinte pergunta: por que eu preciso seguir um processo? E complementam, perguntando também o seguinte: Qual processo Linus Torvalds usou na implementação do sistema operacional Linux? Ou que Donald Knuth usou na implementação do formatador de textos TeX?

Na verdade, a segunda parte da pergunta não faz muito sentido, pois tanto o Linux (no seu início) e o TeX são projetos individuais, liderados por um único desenvolvedor. Nesses casos, a adoção de um processo é menos importante. Ou, dizendo de outra forma, o processo em tais projetos é pessoal, composto pelos princípios, práticas e decisões tomadas pelo seu único desenvolvedor; e que terão impacto apenas sobre ele mesmo.

Porém, os sistemas de software atuais são por demais complexos para serem desenvolvidos por uma única pessoa. Por isso, casos de sistemas desenvolvidos por heróis serão cada vez mais raros. Na prática, os sistemas modernos — e que nos interessam neste livro — são desenvolvidos em **equipes**.

E essas equipes, para produzir software com qualidade e produtividade, precisam de um ordenamento, mesmo que mínimo. Por isso, empresas dão tanto valor a processos de software. Eles são o instrumento de que as empresas dispõem para coordenar, motivar, organizar e avaliar o trabalho de seus desenvolvedores, de forma a garantir que eles trabalhem com produtividade e produzam sistemas alinhados com os objetivos da organização. Sem um processo — mesmo que simplificado e leve, como os processos ágeis que estudaremos neste capítulo — existe o risco de que os times de desenvolvimento passem a trabalhar de forma descoordenada, gerando produtos sem valor para o negócio da empresa. Por fim, processos são importantes não apenas para a empresa, mas também para os desenvolvedores, pois permitem que eles tomem consciência das tarefas e resultados que se esperam deles. Sem um processo, os desenvolvedores podem se sentir perdidos, trabalhando de forma errática e sem alinhamento com os demais membros do time de desenvolvimento.

Neste capítulo, vamos estudar alguns processos de Software. Na verdade, no Capítulo 1 já comentamos sobre Processos Waterfall e Ágeis. Na próxima seção, vamos retomar essa discussão e, em seguida, descrever alguns métodos de desenvolvimento de software.

2.2 Manifesto Ágil

Os primeiros processos de desenvolvimento de software — do tipo Waterfall, propostos ainda na década de 70 — eram estritamente sequenciais, começando com uma fase de especificação de requisitos até chegar às fases finais de implementação, testes e manutenção do sistema.

Se considerarmos o contexto histórico, essa primeira visão de processo era natural, visto que projetos de Engenharia tradicional também são sequenciais e precedidos de um planejamento detalhado. Todas as fases também geram documentações detalhadas do produto que está sendo desenvolvido. Por isso, nada mais natural que a nascente Engenharia de Software se espelhasse nos processos de áreas mais tradicionais, como a Engenharia Eletrônica, Civil, Mecânica, Aeronáutica, etc.

No entanto, após cerca de uma década, começou-se a perceber que software é diferente de outros produtos de Engenharia. Essa percepção foi ficando clara devido aos problemas frequentes enfrentados por projetos de software nas décadas de 70 a 90. Por exemplo, os cronogramas e orçamentos desses projetos não eram obedecidos. Não raro, projetos inteiros eram cancelados, após um ou mais anos de trabalho, sem entregar um sistema funcional para os clientes.

Em 1994, um relatório produzido pela empresa de consultoria Standish Group revelou informações mais detalhadas sobre os projetos de software da época. Por exemplo, o relatório, que ficou conhecido pelo sugestivo nome de CHAOS Report ([link](#)), mostrou que mais de 55% dos projetos estourava os prazos planejados entre 51% e 200%; pelo menos 12% estouraram os prazos acima de 200%, conforme mostra o próximo gráfico:

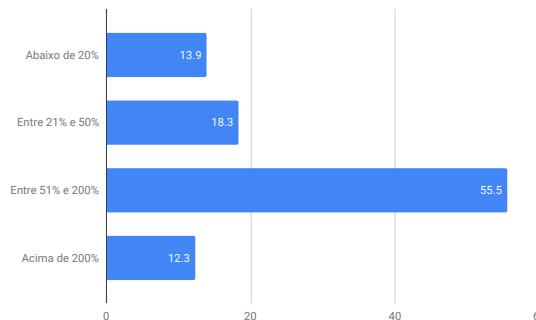


Figura 2.1: CHAOS Report (1994): percentual de projetos que estourava seus **prazos** (para cada faixa de estouro).

Os resultados em termos de custos não eram mais animadores: quase 40% dos projetos ultrapassava o orçamento entre 51% e 200%, como mostra o gráfico:

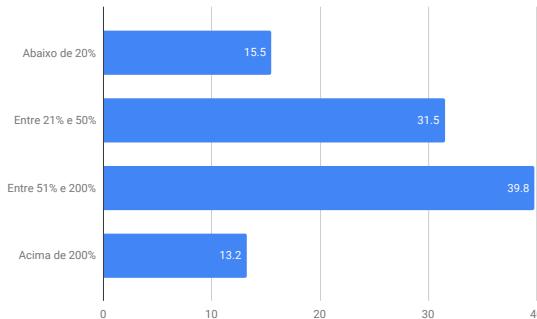


Figura 2.2: CHAOS Report (1994): percentual de projetos que estourava seus **orçamentos** (para cada faixa de estouro).

Em 2001, um grupo de profissionais da indústria se reuniu na cidade de Snowbird, no estado norte-americano de Utah, para discutir e propor uma alternativa aos processos do tipo Waterfall que então predominavam. Essencialmente, eles passaram a defender que software é diferente de produtos tradicionais de Engenharia. Por isso, software também demanda um processo de desenvolvimento diferente.

Por exemplo, os requisitos de um software mudam com frequência, mais do que os requisitos de um computador (hardware), de um avião ou de uma ponte. Além disso, os clientes frequentemente não têm uma ideia precisa do que querem. Ou seja, corre-se o risco de projetar por anos um produto que depois de pronto não será mais necessário, ou porque o mundo mudou ou porque os planos e as necessidades dos clientes mudaram. Eles diagnosticaram ainda problemas referentes aos documentos prescritos por processos do tipo Waterfall, incluindo documentos de requisitos, fluxogramas, diagramas, etc. Esses documentos eram detalhados, pesados e extensos; assim, rapidamente se tornavam obsoletos, pois quando os requisitos mudavam os desenvolvedores não propagavam as alterações para a documentação, mas apenas para o código fonte.

Então eles decidiram lançar as bases para um novo conceito de processo de software, as quais foram registradas em um documento que chamaram de **Manifesto Ágil**. Por ser curto, iremos reproduzir o texto do manifesto a seguir:

Por meio desse trabalho, passamos a valorizar:
Indivíduos e interações, mais do que processos e ferramentas
Software em funcionamento, mais do que documentação abrangente
Colaboração com o cliente, mais do que negociação de contratos
Resposta a mudanças, mais do que seguir um plano.

A característica principal de processos ágeis é a adoção de **ciclos curtos e iterativos de desenvolvimento**, por meio dos quais um sistema é implementado de forma gradativa; começando por aquilo que é mais urgente para o cliente. De início, implementa-se uma primeira versão do sistema, com as funcionalidades que segundo o cliente são para “ontem”, isto é, possuem prioridade máxima. Em seguida, essa versão é validada pelo cliente. Se ela for aprovada, um novo ciclo — ou **iteração** — inicia-se, com mais algumas funcionalidades, também priorizadas pelos clientes. Normalmente, esses ciclos são curtos, com duração de um mês, talvez até um pouco menos. Assim, o sistema vai sendo construído de forma incremental; sendo cada incremento devidamente aprovado pelos clientes. O desenvolvimento termina quando o cliente decide que todos os requisitos estão implementados.

As próximas figuras comparam desenvolvimento em Waterfall e Ágil.



Figura 2.3: Desenvolvimento usando um Processo Waterfall. O sistema fica pronto apenas no final.

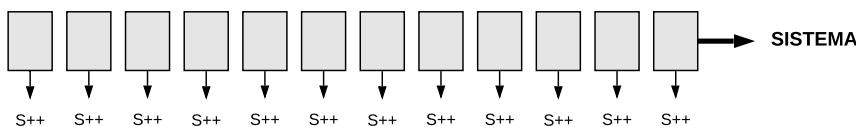


Figura 2.4: Desenvolvimento usando um Processo Ágil. A cada iteração (representada pelos retângulos) gera-se um incremento no sistema (S++), que já pode ser validado e testado pelos usuários finais.

No entanto, a figura anterior pode sugerir que, em desenvolvimento ágil, cada

iteração é um mini-waterfall, incluindo todas as fases de um processo Waterfall. Isso não é verdadeiro; em geral, as iterações em métodos ágeis não são um pipeline de tarefas, como em Waterfall (mais sobre isso nas próximas seções). A figura também pode sugerir que ao final de cada iteração tem-se que colocar um sistema em produção, para uso pelos usuários finais. Isso também não é verdade. De fato, o objetivo é entregar um sistema funcional, isto é, que realize tarefas úteis. Porém, a decisão de colocá-lo em produção envolve outras variáveis, como riscos para o negócio da empresa, disponibilidade de servidores, campanhas de marketing, elaboração de manuais, treinamento de usuários, etc.

Outras características de processos ágeis incluem:

- **Menor ênfase em documentação**, ou seja, apenas o essencial deve ser documentado.
- **Menor ênfase em planos detalhados**, pois muitas vezes nem o cliente, nem os Engenheiros de Software têm, no início de um projeto, uma ideia clara dos requisitos que devem ser implementados. Esse entendimento vai surgir ao longo do caminho, à medida que incrementos de produto sejam produzidos e validados. Em outras palavras, o importante em desenvolvimento ágil é conseguir avançar, mesmo em ambientes com informações imperfeitas, parciais e sujeitas a mudanças.
- **Inexistência de uma fase dedicada a design** (*big upfront design*). Em vez disso, o design também é incremental. Ele evolui à medida que o sistema vai nascendo, ao final de cada iteração.
- **Desenvolvimento em times pequenos**, com cerca de uma dezena de desenvolvedores. Ou, em outras palavras, times que possam ser alimentados com duas pizzas, conforme popularizado pelo CEO da Amazon, Jeff Bezos.
- **Ênfase em novas práticas de desenvolvimento** (pelo menos, para o início dos anos 2000), como **programação em pares**, **testes automatizados** e **integração contínua**.

Devido a essas características, processos ágeis são considerados **processos leves**, com poucas prescrições e documentos.

No entanto, as características acima são genéricas e abrangentes; por isso alguns métodos foram propostos para ajudar desenvolvedores a adotar os princípios ágeis, de forma mais concreta. O interessante é que todos eles foram propostos, pelo menos na primeira versão, antes do encontro de Utah, em 2001, que lançou o Manifesto Ágil.

Neste capítulo, vamos estudar três métodos ágeis:

- **Extreme Programming (XP)**, proposto por Kent Beck, em um livro lançado em 1999 ([link](#)). Uma segunda edição do livro, incluindo uma grande revisão, foi lançada em 2004. Neste capítulo, vamos nos basear nessa edição mais recente.
- **Scrum**, proposto por Jeffrey Sutherland e Ken Schwaber, em um artigo publicado em 1995 ([link](#)).
- **Kanban**, cujas origens remontam a um sistema de controle de produção que começou a ser usado nas fábricas da Toyota, ainda na década de 50 ([link](#)). Nos últimos 10 anos, Kanban tem sido gradativamente adaptado para uso no desenvolvimento de software.

Aprofundamento: Neste livro, usamos os termos **processos** e **métodos**.

Processo é o conjunto de passos, etapas e tarefas que se usa para construir um software. Toda organização usa um processo para desenvolver seus sistemas, o qual pode ser ágil ou waterfall, por exemplo. Ou, talvez, esse processo pode ser caótico. Porém, o ponto que queremos reforçar é que sempre existe um processo. Já método, no nosso contexto, define e especifica um determinado processo de desenvolvimento (a palavra método tem sua origem no grego, onde significa “caminho para se chegar a um objetivo”). Assim, XP, Scrum e Kanban são métodos ágeis ou, de modo mais extenso, são métodos que definem práticas, atividades, eventos e técnicas compatíveis com princípios ágeis de desenvolvimento de software. Aproveitando que estamos tratando de definições, frequentemente usa-se também o termo **metodologia** quando se fala de processos de software. Por exemplo, é comum ver referências a metodologias para desenvolvimento de software, metodologias ágeis, metodologia orientada a objetos, etc. A palavra metodologia, no sentido estrito, denota o “ramo da lógica que se ocupa dos métodos das diferentes ciências”, segundo o Dicionário Houaiss. No entanto, a palavra também pode ser usada como sinônimo de método, segundo o mesmo dicionário. Apesar disso, neste livro evitamos usar o termo metodologia e tentamos empregar sempre o termo método.

Aviso: Todo método de desenvolvimento deve ser entendido como um conjunto de recomendações; cabe a uma organização analisar cada uma e decidir se ela faz sentido no seu contexto. Como resultado, a organização pode ainda

decidir por adaptar essas recomendações para atender às suas necessidades. Logo, provavelmente, não existem duas organizações que seguem exatamente o mesmo processo de desenvolvimento, mesmo que elas digam que estão desenvolvendo usando Scrum, por exemplo.

Mundo Real: O sucesso e impacto de processos ágeis foi impressionante. Hoje, a grande maioria das empresas que desenvolvem software, independente de seu tamanho ou do foco de seu negócio, usam princípios ágeis, em maior ou menor escala. Para citar alguns dados, em 2018, o Stack Overflow survey incluiu uma pergunta sobre o método de desenvolvimento mais usado pelos respondentes ([link](#)). Essa pergunta recebeu 57 mil respostas de desenvolvedores profissionais e a grande maioria deles mencionou métodos ou práticas ágeis, incluindo aquelas que vamos estudar neste capítulo, como Scrum (63% das respostas), Kanban (36%) e Extreme Programming (16%). Apenas 15% dos participantes marcaram Waterfall como resposta.

2.3 Extreme Programming

Segundo seu autor, XP é um método leve recomendado para desenvolver software com requisitos vagos ou sujeitos a mudanças; isto é, basicamente sistemas comerciais, na classificação que adotamos no Capítulo 1. Sendo um método ágil, XP possui todas as características que mencionamos na seção anterior, isto é: adota ciclos curtos e iterativos de desenvolvimento, concede menos ênfase a documentação e planos detalhados, propõe que o design de um sistema também seja definido de forma incremental e sugere que as equipes de desenvolvimento sejam pequenas.

Porém, XP não é um método prescritivo, que define um passo a passo detalhado para construção de software. Em vez disso, XP é definido por meio de um conjunto de **valores, princípios e práticas de desenvolvimento**. Ou seja, XP é inicialmente definido de forma abstrata, usando-se de valores e princípios que devem fazer parte da cultura e dos hábitos de times de desenvolvimento de software. Depois, esses valores e princípios são concretizados em uma lista de práticas de desenvolvimento. Frequentemente, quando decidem adotar XP, desenvolvedores e organizações concentram-se nas práticas. Porém, os valores e princípios são componentes chaves do método, pois são eles que dão sentido às práticas propostas em XP. Sendo mais claro, se uma organização não está

preparada para trabalhar no modelo mental de XP — representado pelos seus valores e princípios — recomenda-se também não adotar suas práticas.

Neste capítulo, vamos primeiro apresentar os valores e princípios de XP. Veja uma lista deles a seguir:

- **Valores:** comunicação, simplicidade, feedback, coragem, respeito e qualidade de vida.
- **Princípios:** humanidade, economicidade, benefícios mútuos, melhorias contínuas, falhas acontecem, baby steps, responsabilidade pessoal

Em seguida, vamos descrever as práticas. Para facilitar a explicação delas, resolvemos organizá-las em três grupos: práticas sobre o processo de desenvolvimento, práticas de programação e práticas de gerenciamento de projetos. Veja a seguir uma lista das práticas em cada grupo:

- **Práticas sobre o Processo de Desenvolvimento:** representante dos clientes, histórias dos usuários, iterações, releases, planejamento de releases, planejamento de iterações, planning poker, slack.
- **Práticas de Programação:** design incremental, programação pareada, desenvolvimento dirigido por testes (TDD), build automatizado, integração contínua.
- **Práticas de Gerenciamento de Projetos:** métricas, ambiente de trabalho, contratos com escopo aberto.

2.3.1 Valores

XP defende que o desenvolvimento de projetos de software seja norteado por três valores principais: comunicação, simplicidade e feedback. Na verdade, argumenta-se que esses valores são universais, para convívio humano. Ou seja, eles não servem apenas para guiar projetos de desenvolvimento, mas a própria vida em sociedade. Uma boa **comunicação** é importante em qualquer projeto, não apenas para evitar, mas também para aprender com erros. O segundo valor de XP é **simplicidade**, pois em todo sistema complexo e desafiador existem sistemas ou subsistemas mais simples, que às vezes não são considerados. Por último, existem riscos em todos os projetos de software: os requisitos mudam, a tecnologia muda, a equipe de desenvolvimento muda, o mundo muda, etc. Um valor que ajuda a controlar tais riscos é estar aberto ao **feedback** dos stakeholders, a fim de que correções de rota sejam implementadas o quanto antes. Em outras palavras, é difícil desenvolver o sistema de software “certo” em uma primeira e única tentativa. Frederick Brooks tem uma frase conhecida sobre esse fenômeno:

“Planeje-se para jogar fora partes de seu sistema, pois você fará isso.”

Por isso, feedback é um valor essencial para garantir que as partes ou versões que serão descartadas sejam identificadas o quanto antes, de forma a diminuir prejuízos e retrabalho. Além dos três valores mencionados, XP também defende outros valores, como **coragem, respeito e qualidade de vida**.

2.3.2 Princípios

Os valores que mencionamos são abstratos e universais. Por outro lado, as práticas que vamos mencionar mais adiante são procedimentos concretos e pragmáticos. Assim, para unir esses dois extremos, XP defende que projetos de software devem seguir um conjunto de princípios. A imagem que se apresenta é de um rio: de um lado estão os valores e de outro as práticas. Os princípios — que descreveremos agora — fazem o papel de uma ponte ligando esses dois lados. Alguns dos principais princípios de XP são os seguintes:

Humanidade (“humanity”, em inglês). Software é uma atividade intensiva no uso de capital humano. O principal recurso de uma empresa de software não são seus bens físicos — computadores, prédios, móveis ou conexões de Internet, por exemplo — mas sim seus colaboradores. Um termo que reflete bem esse princípio é *peopleware*, o qual foi cunhado por Tom DeMarco, em um livro com o mesmo título ([link](#)). A ideia é que a gestão de pessoas — incluindo fatores como expectativas, crescimento, motivação, transparência, responsabilidade, etc — é um princípio chave para o sucesso de projetos de software.

Economicidade (*economics*, em inglês). Se por um lado, *peopleware* é fundamental, por outro lado software é uma atividade cara, que demanda a alocação de recursos financeiros consideráveis. Logo, tem-se que ter consciência de que o outro lado, isto é, quem está pagando as contas do projeto, espera resultados econômicos e financeiros. Por isso, na grande maioria dos casos, software não pode ser desenvolvido apenas para satisfazer a vaidade intelectual de seus desenvolvedores. Software não é uma obra de arte, mas algo que tem que gerar resultados econômicos, como defendido por esse princípio de XP.

Benefícios Mútuos. XP defende que as decisões tomadas em um projeto de software têm que beneficiar múltiplos stakeholders. Por exemplo, o contratante do software deve garantir um bom ambiente de trabalho (*peopleware*); em contrapartida, a equipe deve entregar um sistema que agregue valor ao seu negócio (*economicidade*). Mais um exemplo: ao escrever testes um desenvolvedor se beneficia, pois eles ajudam a detectar bugs no seu código; mas testes também ajudam

outros desenvolvedores, que futuramente terão mais segurança de que o seu código não vai introduzir regressões — isto é, bugs — em código que está funcionando. Um terceiro e último exemplo: refactoring é uma atividade que torna o código mais limpo e fácil de entender, tanto para quem o escreveu, como para quem futuramente terá que mantê-lo. A frase “todo negócio tem que ser bom para os dois lados” resume bem esse terceiro princípio de XP.

Melhorias Contínuas (no livro de XP, o nome original é *improvements*): Como expressa a frase de Kent Beck que abre este capítulo, nenhum processo de desenvolvimento de software é perfeito. Por isso, é mais seguro trabalhar com um sistema que vai sendo continuamente aprimorado, a cada iteração, com o feedback dos clientes e de todos os membros do time. Pelo mesmo motivo, XP não recomenda investir um grande montante de tempo em um design inicial e completo. Em vez disso, o design do sistema também é incremental, melhorando a cada iteração. Por fim, as próprias práticas de desenvolvimento podem ser aprimoradas; para isso, o time deve reservar tempo para refletir sobre elas.

Falhas Acontecem. Desenvolvimento de software não é uma atividade livre de riscos. Como discutido no Capítulo 1, software é uma das mais complexas construções humanas. Logo, falhas são esperadas em projetos de desenvolvimento de software. No contexto desse princípio, falhas incluem bugs, funcionalidades que não se mostraram interessantes para os usuários finais e requisitos não-funcionais que não estão sendo plenamente atendidos, como desempenho, usabilidade, privacidade, disponibilidade, etc. Evidentemente, XP não advoga que essas falhas devem ser acobertadas. Porém, elas não devem ser usadas para punir membros de um time. Pelo contrário, falhas fazem parte do jogo, se um time pretende avançar, inovar e entregar software com rapidez.

Baby Steps. É melhor progressos seguros, testados e validados, mesmo que pequenos, do que grandes implementações com riscos de serem descartadas pelos usuários. O mesmo vale para testes (que são úteis mesmo quando as unidades testadas são de menor granularidade), integração de código (é melhor integrar diariamente, do que passar pelo stress de fazer uma grande integração após semanas de trabalho) e refatorações (que devem ocorrer em pequenos passos, quando é mais fácil verificar que o comportamento do sistema está sendo preservado). Em resumo, o importante é garantir melhorias contínuas, não importando que sejam pequenas, desde que na direção correta. Essas pequenas melhorias são melhores do que grandes revoluções, as quais costumam não apresentar resultados positivos, pelo menos quando se trata de desenvolvimento de software.

Responsabilidade Pessoal (que usamos como tradução para *accepted responsibility*)

bility). De acordo com esse princípio, desenvolvedores devem ter uma ideia clara de seu papel e responsabilidade na equipe. O motivo é que responsabilidade não pode ser transferida, sem que a outra parte a aceite. Por isso, XP defende que o engenheiro de software que implementa uma *história* — termo que o método usa para requisitos — deve ser também aquele que vai testá-la e mantê-la.

Mundo Real: Um dos primeiros sistemas a adotar XP foi um sistema de folha de pagamentos da fabricante de automóveis Chrysler, chamado Chrysler Comprehensive Compensation (C3) ([link](#)). O projeto desse sistema começou no início de 1995 e, como não apresentou resultados concretos, ele foi reiniciado no ano seguinte, sob a liderança de Kent Beck. Outro membro conhecido da comunidade ágil, Martin Fowler, participou do projeto, como consultor. No desenvolvimento do sistema C3, foram usadas e testadas diversas ideias do método que poucos anos depois receberia o nome de XP.

2.3.3 Práticas sobre o Processo de Desenvolvimento

XP — como outros métodos ágeis — recomenda o envolvimento dos clientes com o projeto. Ou seja, além de desenvolvedores, os times incluem pelo menos um **representante dos clientes**, que deve entender do domínio do sistema que será construído. Uma das funções desse representante é escrever as **histórias de usuário** (*user stories*), que é o nome que XP dá para os documentos que descrevem os requisitos do sistema a ser implementado. No entanto, histórias são documentos resumidos, com apenas duas ou três sentenças, com as quais o representante dos clientes define o que ele deseja que o sistema faça, usando sua própria linguagem.

Iremos aprofundar o estudo sobre histórias de usuários no Capítulo 3. Mas, por enquanto, gostaríamos de adiantar que as histórias são escritas em cartões de papel, normalmente à mão. Ou seja, em vez de documentos de requisitos detalhados, histórias são documentos simples, que focam nas funcionalidades do sistema, sempre na visão de seus usuários.

Como exemplo, mostramos a seguir uma história de um sistema de perguntas e respostas — semelhante ao famoso Stack Overflow ([link](#)) — que usaremos neste capítulo para explicar XP. Observe que a história tem um título (Postar Pergunta) e uma breve descrição, que não ocupa mais do que duas ou três sentenças. Costuma-se dizer que histórias são um lembrete para que depois esse requisito seja verbalmente detalhado pelo representante dos clientes.

Postar Pergunta

Um usuário, quando logado no sistema, deve ser capaz de postar perguntas. Como é um site sobre programação, as perguntas podem incluir blocos de código, os quais devem ser apresentados com um layout diferenciado.

Depois de escritas pelo representante dos clientes, as histórias são estimadas pelos desenvolvedores. Ou seja, são os desenvolvedores que definem, mesmo que preliminarmente, quanto tempo será necessário para implementar as histórias escritas pelo representante dos clientes. Frequentemente, a duração de uma história é estimada em **story points**, em vez de horas ou homens/hora. Nesses casos, usa-se uma escala inteira para classificar histórias como possuindo um certo número de story points. O objetivo é definir uma ordem relativa entre as histórias. As histórias mais simples são estimadas como tendo tamanho igual a 1 story point; histórias que são cerca de duas vezes mais complexas do que as primeiras são estimadas como tendo 2 story points e assim por diante. Muitas vezes, usa-se também uma sequência de Fibonacci para definir a escala de possíveis story points, como em 1, 2, 3, 5, 8, 13 story points. Nesse caso, o objetivo é criar uma escala que torne as tarefas progressivamente mais difíceis e, ao mesmo tempo, permita ao time realizar comparações similares à seguinte: será que o esforço para implementar essa tarefa que planejamos estimar com 8 story points é equivalente ao esforço de implementar uma tarefa na escala anterior (5 story points) e mais uma tarefa na próxima escala inferior (3 pontos)? Se isso for verdade, 8 story points é uma boa estimativa. Caso contrário, o melhor é estimar a história com 5 story points.

Aprofundamento: Uma técnica usada para estimar o tamanho de histórias é conhecida como **Planning Poker**. Ela funciona assim: o representante dos clientes seleciona uma história e a lê para os desenvolvedores. Após a leitura, os desenvolvedores interagem com o representante dos clientes para tirar possíveis dúvidas e conhecer melhor a história. Feito isso, cada desenvolvedor faz sua estimativa para o tamanho da história, de forma independente. Depois disso, eles ao mesmo tempo levantam cartões com a estimativa que pensaram, em story points. Esses cartões foram distribuídos antes e deles constam os números 1, 2, 3, 5, etc. Se houver consenso, o tamanho da história está estimado e passa-se para a próxima história. Senão, o time deve iniciar uma discussão, para esclarecer a razão das diferentes estimativas. Por exemplo, os desenvolvedores responsáveis pelas estimativas mais discrepantes podem

explicar o motivo da sua proposta. Feito isso, realiza-se uma nova votação e o processo se repete, até que o consenso seja alcançado.

A implementação das histórias ocorre em **iterações**, as quais têm uma duração fixa e bem definida, variando de uma a três semanas, por exemplo. As iterações, por sua vez, formam ciclos mais longos, chamados de **releases**, de dois a três meses, por exemplo. A **velocidade** de um time é o número de story points que ele consegue implementar em uma iteração. Sugere-se que o representante dos clientes escreva histórias que requeiram pelo menos uma release para serem implementadas. Ou seja, em XP, o horizonte de planejamento é uma release, isto é, alguns meses.

Aviso: Em XP, a palavra release tem um sentido diferente daquele que se usa em gerência de configuração. Em gerência de configuração, uma release é uma versão de um sistema que será disponibilizada para seus usuários finais. Como já mencionamos em um aviso anterior, não necessariamente a versão do sistema ao final de uma release de XP precisa entrar em produção.

Em resumo, para começar a usar XP precisamos de:

- Definir a duração de uma iteração.
- Definir o número de iterações de uma release.
- Um conjunto de histórias, escritas pelo representante dos clientes.
- Estimativas para cada história, feitas pelos desenvolvedores.
- Definir a velocidade do time, isto é, o número de story points que ele consegue implementar por iteração.

Uma vez definidos os parâmetros e documentos acima, o representante do cliente deve priorizar as histórias. Para isso, ele deve definir quais histórias serão implementadas nas iterações da primeira release. Nesta priorização, deve-se respeitar a velocidade do time de desenvolvimento. Por exemplo, suponha que a velocidade de um time seja de 25 story points por iteração. Nesse caso, o representante do cliente não pode alocar histórias para uma iteração cujo somatório de story points ultrapasse esse limite. A tarefa de alocar histórias a iterações e releases é chamada de **planejamento de releases** (ou então *planning game*, que foi o nome adotado na primeira edição do livro de XP).

Por exemplo, suponha o fórum de perguntas e respostas que mencionamos antes. A próxima tabela resume o resultado de um possível planejamento de releases. Nessa

tabela, estamos assumindo que o representante dos clientes escreveu 8 histórias, que cada release possui duas iterações e que a velocidade do time é de 21 story points por iteração (veja que o somatório dos story points de cada iteração é exatamente igual a 21).

História	Story Points	Iteração	Release
Cadastrar usuário	8	1	1
Postar perguntas	5	1	1
Postar respostas	3	1	1
Tela de abertura	5	1	1
Gamificar perguntas e respostas	5	2	1
Pesquisar perguntas e respostas	8	2	1
Adicionar tags em perguntas e respostas	5	2	1
Comentar perguntas e respostas	3	2	1

A tabela anterior serve para reforçar dois pontos já mencionados: (1) as histórias em XP representam funcionalidades do sistema que se pretende construir; isto é, a implementação do sistema é dirigida por suas funcionalidades; (2) os desenvolvedores não opinam sobre a ordem de implementação das histórias; isso é decidido pelo representante dos clientes, que deve ser alguém capacitado e com autoridade para definir o que é mais urgente e importante para a empresa que está contratando o desenvolvimento do sistema.

Uma vez realizado o planejamento de uma release, começam as iterações. Antes de mais nada, o time de desenvolvimento deve se reunir para realizar o **planejamento da iteração**. O objetivo desse planejamento é decompor as histórias de uma iteração em tarefas, as quais devem corresponder a atividades de programação que possam ser alocadas para um dos desenvolvedores do time. Por exemplo, a seguinte lista mostra as tarefas para a história “Postar Perguntas”, que é a primeira história que será implementada em nosso sistema de exemplo.

- Projetar e testar a interface Web, incluindo layout, CSS templates, etc.
- Instalar banco de dados, projetar e criar tabelas.
- Implementar a camada de acesso a dados.
- Instalar servidor e testar framework web.
- Implementar camada de controle, com operações para cadastrar, remover e atualizar perguntas.

- Implementar interface Web.

Como regra geral, as tarefas não devem ser complexas, devendo ser possível concluir-las em alguns dias.

Resumindo, um projeto XP é organizado em:

- releases, que são conjunto de iterações, com duração total de alguns meses.
- iterações, que são conjuntos de tarefas, com duração total de algumas semanas.
- tarefas, com duração de alguns dias.

Definidas as tarefas, o time deve decidir qual desenvolvedor será responsável por cada uma. Feito isso, começa de fato a iteração, com a implementação das tarefas.

Uma iteração termina quando todas as histórias estiverem implementadas e validadas pelo representante dos clientes. Assim, ao fim de uma iteração, as histórias devem ser mostradas para o representante dos clientes, que deve concordar que elas, de fato, atendem ao que ele especificou.

XP defende ainda que os times, durante uma iteração, programem algumas **folgas** (*slacks*), que são tarefas que podem ser adiadas, caso necessário. Como exemplos, podemos citar o estudo de uma nova tecnologia, a realização de um curso online, preparar uma documentação ou manual ou mesmo desenvolver um projeto paralelo. Algumas empresas, como o Google, por exemplo, são famosas por permitir que seus desenvolvedores usem 20% de seu tempo para desenvolver um projeto pessoal ([link](#)). No caso de XP, folgas têm dois objetivos principais: (1) criar um “buffer de segurança” em uma iteração, que possa ser usado caso alguma tarefa demande mais tempo do que o previsto; (2) permitir que os desenvolvedores respirem um pouco, pois o ritmo de trabalho em projetos de desenvolvimento de software costuma ser intenso e desgastante. Logo, os desenvolvedores precisam de um tempo para realizarem algumas tarefas onde não existe uma cobrança imediata de resultados.

Perguntas Frequentes

Vamos agora responder algumas perguntas sobre as práticas de XP que acabamos de explicar.

Qual a duração ideal de uma iteração? Difícil precisar, pois depende das características do time, da empresa contratante, da complexidade do sistema a ser desenvolvido, etc. Iterações curtas — por exemplo, de uma semana — propiciam feedback mais rápido. Porém, requerem um maior comprometimento dos clientes,

pois toda semana um novo incremento de produto deve ser validado. Além disso, requerem que as histórias sejam mais simples. Por outro lado, iterações mais longas — por exemplo, de um mês — permitem que o time planeje e conclua as tarefas com mais tranquilidade. Porém, demora-se um pouco mais para receber feedback dos clientes. Esse feedback pode ser importante quando os requisitos são pouco claros. Por isso, uma escolha de compromisso seria algo como 2 ou 3 semanas. Outra alternativa recomendada consiste em experimentar, isto é, testar e avaliar diferentes durações, antes de decidir.

O que o representante dos clientes faz durante as iterações? No início de uma release, cabe ao representante dos clientes escrever as histórias das iterações que farão parte dessa release. Depois, no final de cada iteração, cabe a ele validar e aprovar a implementação das histórias. Porém, durante as iterações, ele deve estar fisicamente disponível para tirar dúvidas do time. Veja que uma história é um documento muito resumido, logo é natural que surjam dúvidas durante a sua implementação. Por isso, o representante dos clientes deve estar sempre disponível para se reunir com os desenvolvedores, para tirar dúvidas e explicar detalhes relativos à implementação de histórias.

Como escolher o representante dos clientes? Antes de mais nada, deve ser alguém que conheça o domínio do sistema e que tenha autoridade para priorizar histórias. Conforme detalhado a seguir, existem pelo menos três perfis de representante dos clientes:

- Suponha o desenvolvimento interno de um sistema, isto é, o departamento de sistemas da empresa X está desenvolvendo um sistema para um outro departamento Y, da mesma empresa. Nesse caso, o representante dos clientes deve ser um funcionário do departamento Y.
- Suponha que o time de desenvolvimento foi contratado para desenvolver um sistema para a empresa X. Ou seja, trata-se de um desenvolvimento terceirizado. Nesse caso, o representante do cliente deve ser um funcionário da empresa X, com pleno domínio da área do sistema e que vai ser um dos seus principais usuários, quando ele ficar pronto.
- Suponha que o time de desenvolvimento de uma empresa X foi designado para fazer um sistema para um público externo à empresa. Por exemplo, um sistema como aquele usado neste capítulo, similar ao Stack Overflow. Logo, os clientes do sistema não são funcionários de X, mas sim clientes externos. Nesse caso, o representante dos clientes deve ser alguém da área de marketing, vendas ou negócios da empresa X. Em última instância, pode ser o dono da empresa. Em qualquer caso, a sugestão é que seja uma pessoa

próxima do problema e o mais distante possível da solução. Por isso mesmo, deve-se evitar que ele seja um desenvolvedor ou um gerente de projeto. O tipo de representante dos clientes que mencionamos neste item é, às vezes, chamado de um **user proxy**.

Como definir a velocidade do time? Não existe bala de prata para essa questão. Essa definição depende da experiência do time e de seus membros. Se eles já participaram de projetos semelhantes àquele que estão iniciando, certamente essa deve ser uma questão menos difícil. Caso contrário, tem-se que experimentar e ir calibrando a velocidade nas iterações seguintes.

Histórias podem incluir tarefas de instalação de infra-estrutura de software? Não, histórias são especificadas pelo representante dos clientes, que é um profissional leigo em Engenharia de Software. Portanto, ele não costuma ter conhecimento de infra-estrutura de software. No entanto, uma história pode dar origem a uma tarefa como “instalar e testar o banco de dados”. Resumindo, histórias estão associadas a requisitos funcionais; para implementá-las criam-se tarefas, que podem estar associadas a requisitos funcionais, não-funcionais ou tarefas técnicas, como instalação de bancos de dados, servidores, frameworks, etc.

A história X depende da história Y, mas o representante dos clientes priorizou Y antes de X. O que devo fazer? Por exemplo, suponha que no sistema de exemplo o representante dos clientes tenha alocado a história “Postar Pergunta” para a iteração 2 e a história “Postar Resposta” para a iteração 1. A pergunta então é a seguinte: o time deve respeitar essa alocação? Sim, pois a regra é clara: o representante dos clientes é a autoridade final quando trata-se de definir a ordem de implementação das histórias. Logo, pode-se perguntar em seguida: como que vamos postar respostas, sem ter as perguntas? Para isso, basta implementar algumas perguntas “fixas”, que não possam ser modificadas pelos usuários. Na iteração 1, quando o cliente abrir o sistema, essas perguntas vão aparecer por default, talvez com um layout bem simples, e então o cliente vai poder usar o sistema apenas para responder essas perguntas fixas.

Quando um projeto XP termina? Quando o representante dos clientes decide que as histórias já implementadas são suficientes e que não há mais nada de relevante que deva ser implementado.

2.3.4 Práticas de Programação

O nome Extreme Programming foi escolhido porque XP propõe um conjunto de práticas de programação inovadoras, principalmente para a época na qual foram propostas, no final da década de 90. Na verdade, XP é um método que dá grande

importância à tarefa de programação e produção de código. Essa importância tem que ser entendida no contexto da época, onde havia uma diferença entre analistas e programadores. Analistas eram encarregados de elaborar o projeto em alto nível de um sistema, definindo seus principais componentes, classes e interfaces. Para isso, recomendava-se o uso de uma linguagem de modelagem gráfica, como UML, que veremos em um dos próximos capítulos desse livro. Concluída a fase de análise e projeto, começava a fase de codificação, que ficava a cargo dos programadores. Assim, na prática, existia uma hierarquia nesses papéis, sendo o papel de analista o de maior prestígio. Métodos ágeis — e, particularmente, XP — acabaram com essa hierarquia e passaram a defender a produção de código com funcionalidades, logo nas primeiras semanas de um projeto.

Mas XP não apenas acabou com a grande fase de projeto e análise, logo no início dos projetos. O método também propôs um novo conjunto de práticas de programação, incluindo programação em pares, testes automatizados, desenvolvimento dirigido por testes (TDD), builds automatizados, integração contínua, etc. A maioria dessas práticas passou a ser largamente adotada pela indústria de software e hoje são praticamente obrigatórias na maioria dos projetos — mesmo naqueles que não usam um método ágil.

Nessa seção, vamos estudar as práticas de programação de XP.

Design Incremental. Como afirmado nos parágrafos anteriores, em XP não há uma fase de design e análise detalhados, conhecida como *big design upfront* (BDUF), a qual é uma das principais fases de processos do tipo Waterfall. A ideia é que o time deve reservar tempo para definir o design do sistema que está sendo desenvolvido. Porém, isso deve ser uma atividade contínua e incremental, em vez de estar concentrada no início do projeto, antes de qualquer codificação. No caso, simplicidade é o valor de XP que norteia essa opção por um design também incremental. Argumenta-se que quando o design é confinado no início do projeto, correm-se diversos riscos, pois os requisitos ainda não estão totalmente claros para o time, e nem mesmo para o representante dos clientes. Por exemplo, pode-se supervalorizar alguns requisitos, que mais tarde irão se revelar menos importantes; de forma inversa, pode-se subvalorizar outros requisitos, que depois, com o decorrer da implementação, irão assumir um maior protagonismo. Isso sem falar que novos requisitos podem surgir ao longo do projeto, tornando o design inicial desatualizado.

Por isso, XP defende que o momento ideal para pensar em *design* é quando ele se revelar importante. Frequentemente, duas frases são usadas para motivar e

justificar essa prática: *do the simplest thing that could possibly work e you aren't going to need it*, essa última conhecida pela sigla YAGNI.

Duas observações são importantes para melhor entender a proposta de design incremental. Primeiro, times experientes costumam ter uma boa aproximação do design logo na primeira iteração. Por exemplo, eles já sabem que se trata de um sistema com interface Web, com uma camada de lógica não trivial, mas também não tão complexa, e depois com uma camada de persistência e um banco de dados, certamente relacional. Ou seja, apenas a sentença anterior já define muito do design que deve ser adotado. Como uma segunda observação, nada impede que na primeira iteração o time crie uma tarefa técnica para discutir e refinar o design que vão começar a adotar no sistema.

Por fim, design incremental somente é possível caso seja adotado em conjunto com as demais práticas de XP, principalmente **refactoring**. XP defende que refactoring deve ser continuamente aplicado para melhorar a qualidade do design. Por isso, toda oportunidade de refatoração, visando facilitar o entendimento e a evolução do código, não pode ser deixada para depois.

Programação em Pares. Junto com design incremental, programação em pares é uma das práticas mais polêmicas de XP. Apesar de polêmica, a ideia é simples: toda tarefa de codificação — incluindo implementação de uma nova história, de um teste, ou a correção de um bug — deve ser realizada por dois desenvolvedores trabalhando juntos, compartilhando o mesmo teclado e monitor. Um dos desenvolvedores é o **líder** (ou *driver*) da sessão, ficando com o teclado e o mouse. Ao segundo desenvolvedor cabe a função de revisor e questionador, no bom sentido, do trabalho do líder. Às vezes, esse segundo desenvolvedor é chamado de **navegador**. A inspiração vem dos ralis automobilísticos, onde os pilotos são acompanhados de um navegador.

Com programação em pares espera-se melhorar a qualidade do código e do design, pois “duas cabeças pensam melhor do que uma.” Além disso, programação em pares contribui para disseminar o conhecimento sobre o código, que não fica nas mãos e na cabeça de apenas um desenvolvedor. Por exemplo, não é raro encontrar sistemas nos quais um determinado desenvolvedor tem dificuldade para sair de férias, pois apenas ele conhece uma parte crítica do código. Como uma terceira vantagem, programação por pares pode ser usada para treinar desenvolvedores menos experientes em tecnologias de desenvolvimento, algoritmos e estruturas de dados, padrões e princípios de design, escrita de testes, técnicas de depuração, etc.

Por outro lado, existem também custos econômicos derivados da adoção de programação em pares, já que dois programadores estão sendo alocados para

realizar uma tarefa que, a princípio, poderia ser realizada por apenas um. Além disso, muitos desenvolvedores não se sentem confortáveis com a prática. Para eles é desconfortável — do ponto de vista emocional e cognitivo — discutir cada linha de código e cada decisão de implementação com um colega. Para aliviar esse desconforto, XP propõe que os pares sejam trocados a cada sessão. Elas podem durar, por exemplo, 50 minutos, seguidos de uma pausa de 10 minutos para descanso. Na próxima sessão, os pares e papéis (líder vs revisor) são trocados. Assim, se em uma sessão você atuou como revisor do programador X, na sessão seguinte você passará a ser o líder, mas tendo outro desenvolvedor Y como revisor.

Mundo Real: Em 2008, dois pesquisadores da Microsoft Research, Andrew Begel e Nachiappan Nagappan, realizaram um survey com 106 desenvolvedores da empresa, para capturar a percepção deles sobre programação em pares ([link](#)). Quase 65% dos desenvolvedores responderam positivamente a uma primeira pergunta sobre se programação em pares estaria funcionando bem para eles (“*pair programming is working well for me*”). Quando perguntados sobre os benefícios de programação em pares, as respostas foram as seguintes: redução no número de bugs (62%), produção de código de melhor qualidade (45%), disseminação de conhecimento sobre o código (40%) e oportunidade de aprendizado com os pares (40%). Por outro lado, os custos da prática foram apontados como sendo seu principal problema (75%). Sobre as características do par ideal, a resposta mais comum foi complementaridade de habilidades (38%). Ou seja, desenvolvedores preferem parar com uma pessoa que o ajude a superar seus pontos fracos.

Mais recentemente, diversas empresas passaram a adotar a prática de **revisão de código**. A ideia é que todo código produzido por um desenvolvedor tem que ser revisado e comentado por um outro desenvolvedor, porém de forma offline e assíncrona. Ou seja, nesses casos, o revisor não trabalha fisicamente ao lado do líder. Iremos comentar mais sobre revisão de código no capítulo sobre Qualidade de Software.

Propriedade Coletiva do Código. A ideia é que qualquer desenvolvedor — ou par de desenvolvedores trabalhando junto — pode modificar qualquer parte do código, seja para implementar uma nova feature, para corrigir um bug ou para aplicar um refactoring. Por exemplo, se você descobriu um bug em algum ponto do código, vá em frente e corrija-o. Para isso, você não precisa de autorização de quem implementou esse código ou de quem realizou a última manutenção nele.

Testes Automatizados. Essa é uma das práticas de XP que alcançou maior sucesso. A ideia é que testes manuais — um ser humano executando o programa, fornecendo entradas e checando as saídas produzidas — é um procedimento custoso e que não pode ser reproduzido a todo momento. Logo, XP propõe a implementação de programas — chamados de testes — para executar pequenas unidades de um sistema, como métodos, e verificar se as saídas produzidas são aquelas esperadas. Essa prática prosperou porque, mais ou menos na mesma época de sua proposição, foram desenvolvidos os primeiros frameworks de testes de unidade — como o JUnit, cuja primeira versão, implementada por Kent Beck e Erich Gamma, é de 1997. No momento, não iremos estender a discussão sobre testes automatizados porque vamos ter um capítulo exclusivo para isso no livro.

Desenvolvimento Dirigido por Testes (TDD). Essa é outra prática de programação inovadora proposta por XP. A ideia é simples: se em XP todo método deve possuir testes, por que não escrevê-los primeiro? Isto é, implementa-se o teste de um método e, só então, o seu código. TDD, que também é conhecido como *test-first programming*, possui duas motivações principais: (1) evitar que a escrita de testes seja sempre deixada para amanhã, pois eles são a primeira coisa que se deve implementar; (2) ao escrever um teste, o desenvolvedor se coloca no papel de cliente do método testado, isto é, ele primeiro pensa na sua interface, em como os clientes devem usá-lo, para então pensar na implementação. Com isso, incentiva-se a criação de métodos mais amigáveis, do ponto de vista da interface provida para os clientes. Não iremos alongar a explicação sobre TDD, pois dedicaremos uma seção inteira para essa prática no capítulo de Testes.

Build Automatizado. Build é o nome que se dá para a geração de uma versão de um sistema que seja executável e que possa ser colocada em produção. Logo, inclui não apenas a compilação do código, mas a execução de outras ferramentas como linkeditores e empacotadores de código em arquivos WAR, JAR, etc. No caso de XP, a execução dos testes é outra etapa fundamental do processo de build. Para automatizar esse processo, são usadas ferramentas, como o sistema Make, que faz parte das distribuições do sistema operacional Unix desde a década de 1970. Mais recentemente, surgiram outras ferramentas de build, como Ant, Maven, Gradle, Rake, MSBuild, etc. Primeiro, XP defende que o processo de build seja automatizado, sem nenhuma intervenção dos desenvolvedores. O objetivo é liberá-los das tarefas de rodar scripts, informar parâmetros de linhas de comando, configurar ferramentas, etc. Assim, eles podem focar apenas na implementação de histórias. Segundo, XP defende que o processo de build seja o mais rápido possível, para que os desenvolvedores recebam rapidamente feedback sobre possíveis problemas, como um erro de compilação. Na segunda versão

do livro de XP, recomenda-se um limite de 10 minutos para a conclusão de um build. No entanto, dependendo do tamanho do sistema e de sua linguagem de programação, pode ser difícil atender a esse limite. Por isso, o mais importante é focar na regra geral: sempre procurar automatizar e reduzir o tempo de build.

Integração Contínua. Sistemas de software são desenvolvidos com o apoio de sistemas de controle de versões (VCS, ou *Version Control System*), que armazenam o código fonte do sistema e de arquivos relacionados, como arquivos de configuração, documentação, etc. Hoje o sistema de controle de versão mais usado é o git, por exemplo. Quando se usa um VCS, desenvolvedores tem que primeiro baixar (*pull*) o código fonte para sua máquina local, antes de começar a trabalhar em uma tarefa. Feito isso, eles devem subir o código modificado (*push*). Esse último passo é chamado de **integração** da modificação no código principal, armazenado no VCS. Porém, entre um *pull* e um *push*, outro desenvolvedor pode ter modificado o mesmo trecho de código e realizado a sua integração. Nesse caso, quando o primeiro desenvolvedor tentar subir com seu código, o sistema de controle de versão irá impedir a integração, dizendo que existe um **conflito**.

Conflitos são ruins, porque eles devem ser resolvidos manualmente. Se um desenvolvedor A modificou a inicialização de uma variável x com o valor 10; e outro desenvolvedor B, trabalhando em paralelo com A, gostaria de inicializar x com 20, isso representa um conflito. Para resolvê-lo, A e B devem sentar e discutir qual o melhor valor para inicializar x . No entanto, esse é um cenário simples. Conflitos podem ser mais complexos, envolver grandes trechos de código e mais do que dois desenvolvedores. Por isso, a resolução de conflitos de integração costuma demandar um grande esforço, resultando no que se chama de **integration hell**.

Por outro lado, evitar completamente conflitos é impossível. Por exemplo, não é possível conciliar automaticamente os interesses de dois desenvolvedores quando um deles precisa inicializar x com 10 e outro com 20. Porém, pode-se pelo menos tentar diminuir o número e o tamanho dos conflitos. Para conseguir isso, a ideia de XP é simples: desenvolvedores devem integrar seu código sempre, se possível todos os dias. Essa prática é chamada de **integração contínua**. O objetivo é evitar que desenvolvedores passem muito tempo trabalhando localmente, em suas tarefas, sem integrar o código. E, com isso, pelo menos diminuir as chances e o tamanho dos conflitos.

Para garantir a qualidade do código que está sendo integrado com frequência quase diária, costuma-se usar também um **serviço de integração contínua**. Antes de realizar qualquer integração, esse serviço faz o build do código e executa os testes. O objetivo é garantir que o código não possui erros de compilação e que

ele passa em todos os testes. Existem diversos serviços de integração contínua, como Jenkins, TravisCI, CircleCI, etc. Por exemplo, quando se desenvolve um sistema usando o GitHub, pode-se ativar esses serviços na forma de plugins. Se o repositório GitHub for público, normalmente o serviço de integração contínua é gratuito; se for privado, deve-se pagar uma assinatura.

Iremos estudar mais sobre Integração Contínua no capítulo sobre DevOps.

Mundo Real: Em 2010, Laurie Williams, professora da Universidade da Carolina do Norte, nos EUA, pediu que 326 desenvolvedores respondessem a um questionário sobre a experiência deles com métodos ágeis ([link](#)). Em uma das questões, pedia-se aos participantes para ranquear a importância de práticas ágeis, usando uma escala de 1 a 5, onde o score 5 deveria ser dado apenas para práticas essenciais em desenvolvimento ágil. Três práticas ficaram empatadas em primeiro lugar, com score médio 4.5 e desvio padrão de 0.8. São elas: integração contínua, iterações curtas (menos de 30 dias) e definição de critérios para tarefas concluídas (*done criteria*). Por outro lado, dentre as práticas nas últimas posições podemos citar planning poker (score médio 3.1) e programação em pares (score médio 3.3).

2.3.5 Práticas de Gerenciamento de Projetos

Ambiente de Trabalho. XP defende que o projeto seja desenvolvido com um time pequeno, com menos de 10 desenvolvedores, por exemplo. Todos eles devem estar dedicados ao projeto. Isto é, deve-se evitar times fracionados, nos quais alguns desenvolvedores trabalham apenas alguns dias da semana no projeto e os outros dias em um outro projeto, por exemplo.

Além disso, XP defende que todos os desenvolvedores trabalhem em uma mesma sala, para facilitar comunicação e feedback. Também propõe que o espaço de trabalho seja informativo, isto é, que sejam, por exemplo, fixados cartazes nas paredes, com as histórias da iteração, incluindo seu estado: histórias pendentes, histórias em andamento e histórias concluídas. A ideia é permitir que o time possa visualizar e sentir o trabalho que está sendo realizado.

Outra preocupação de XP é garantir jornadas de trabalho sustentáveis. Empresas de desenvolvimento de software são conhecidas por exigirem longas jornadas de trabalho, com diversas horas extras e trabalho nos finais de semana. XP defende explicitamente que essa prática não é sustentável e que as jornadas de trabalho devem ser sempre próximas de 40 horas, mesmo na véspera de entregas. O

interessante é que XP é um método proposto por desenvolvedores, com grande experiência em projetos reais de desenvolvimento de software. Logo, eles já devem ter sentido na própria pele os efeitos nocivos de longas jornadas de trabalho. Dentre outros problemas, elas podem causar danos à saúde física e mental dos desenvolvedores, assim como incentivar a rotatividade do time, cujos membros vão estar sempre pensando em um novo emprego.

Contratos com Escopo Aberto. Quando uma empresa terceiriza o desenvolvimento, existem duas possibilidades de contrato: com escopo fechado ou com escopo aberto. Em contratos com escopo fechado, a empresa contratante define, mesmo que de forma mínima, os requisitos do sistema e a empresa contratada define um preço e um prazo de entrega. XP advoga que esses contratos são arriscados, pois os requisitos mudam e nem mesmo o cliente sabe antecipadamente o que ele quer que o sistema faça, de modo preciso. Assim, contratos com escopo fixo podem fazer com que a contratada entregue um sistema com problemas de qualidade e mesmo com alguns requisitos implementados de modo parcial ou com bugs, apenas para não ter que pagar possíveis multas. Por outro lado, quando o escopo é aberto, o pagamento ocorre por hora trabalhada. Por exemplo, combina-se que a contratada vai alocar um time com um certo número de desenvolvedores para trabalhar integralmente no projeto, usando as práticas de XP. Combina-se também um preço para a hora de cada desenvolvedor. O cliente define as histórias e faz a validação delas ao final de cada iteração. O contrato pode ser rescindido ou renovado após um certo número de meses, o que dá ao cliente a liberdade de mudar de empresa caso não esteja satisfeito com a qualidade do serviço prestado. Como usual em XP, o objetivo é abrir um fluxo de comunicação e feedback entre contratada e contratante, em vez de forçar a primeira a entregar um produto com problemas conhecidos, apenas para cumprir um contrato. Na verdade, contratos com escopo aberto são mais compatíveis com os princípios do Manifesto Ágil, que explicitamente valoriza “colaboração com o cliente, mais que negociação de contratos”.

Métricas de Processo. Para que gerentes e executivos possam acompanhar um projeto XP recomenda-se o uso de duas métricas principais: número de bugs em produção (que deve ser idealmente da ordem de poucos bugs por ano) e intervalo de tempo entre o início do desenvolvimento e o momento em que o projeto começar a gerar os seus primeiros resultados financeiros (que também deve ser pequeno, da ordem de um ano, por exemplo).

2.4 Scrum

Scrum é um método ágil, iterativo e incremental para gerenciamento de projetos. Foi proposto por Jeffrey Sutherland e Ken Schwaber, em um artigo publicado pela primeira vez em 1995 ([link](#)). Dentre os métodos ágeis, Scrum é o mais conhecido e usado. Provavelmente, parte do sucesso do método seja explicada pela existência de uma indústria associada à sua adoção, a qual inclui a produção de livros, diversos cursos, consultorias e certificações.

Uma pergunta que vamos responder logo no início desta seção diz respeito às diferenças entre Scrum e XP. Existem diversas pequenas diferenças, mas a principal delas é a seguinte:

- XP é um método ágil voltado exclusivamente para projetos de desenvolvimento de software. Para isso, XP inclui um conjunto de práticas de programação, como testes de unidade, programação em pares, integração contínua e design incremental, que foram estudadas na seção anterior.
- Scrum é um método ágil para gerenciamento de projetos, que não necessariamente precisam ser projetos de desenvolvimento de software. Por exemplo, a escrita deste livro — como comentaremos daqui a pouco — é um projeto que está sendo realizado usando conceitos de Scrum. Tendo um foco mais amplo que XP, Scrum não propõe nenhuma prática de programação.

Dentre os métodos ágeis, Scrum é também aquele que é melhor definido. Essa definição inclui um conjunto preciso de **papéis**, **artefatos** e **eventos**, que são listados a seguir. No resto desta seção, vamos explicar cada um deles.

- **Papéis:** Dono do Produto, Scrum Master, Desenvolvedor.
- **Artefatos:** Backlog do Produto, Backlog do Sprint, Quadro Scrum, Gráfico de Burndown.
- **Eventos:** Planejamento do Sprint, Sprint, Reuniões Diárias, Revisão do Sprint, Retrospectiva.

2.4.1 Papéis

Times Scrum são formados por um Dono de Produto (*Product Owner*), um Scrum Master e de três a nove desenvolvedores.

O **Dono do Produto** tem exatamente o mesmo papel do Representante dos Clientes em XP, por isso não vamos explicar de novo a sua função em detalhes. Mas ele, como o próprio nome indica, deve possuir a visão do produto que será

construído, sendo responsável também por maximizar o retorno do investimento feito no projeto. Assim como em XP, cabe ao Dono do Produto escrever as histórias do usuário e, por isso, ele deve estar sempre disponível para tirar dúvidas do time.

O **Scrum Master** é um papel característico e único de Scrum. Trata-se do especialista em Scrum do time, sendo responsável por garantir que as regras do método estão sendo seguidas. Para isso, ele deve continuamente treinar e explicar os princípios de Scrum para os demais membros do time. Ele também deve desempenhar funções de um facilitador dos trabalhos e removedor de impedimentos. Por exemplo, suponha que um time esteja enfrentando problemas com um dos servidores de bancos de dados, cujos discos estão apresentando problemas todos os dias. Cabe ao Scrum Master intervir junto aos níveis adequados da empresa para garantir que esse problema de hardware não atrapalhe o avanço do time. Por outro lado, ele não é um gerente de projeto tradicional. Por exemplo, ele não é o líder do time, pois todos em um time Scrum tem o mesmo nível hierárquico.

Costuma-se dizer que times Scrum são **cross-funcionais**, isto é, eles devem incluir — além do Dono do Produto e do Scrum Master — todos os especialistas necessários para desenvolver o produto, de forma a não depender de membros externos. No caso de projetos de software, isso inclui desenvolvedores *front-end*, desenvolvedores *back-end*, especialistas em bancos de dados, projetistas de interfaces, etc. Cabe a esses especialistas tomar todas as decisões técnicas do projeto, incluindo definição da linguagem de programação, arquitetura e frameworks que serão usados no desenvolvimento. Cabe a eles também estimar o tamanho das histórias definidas pelo Dono do Produto, usando uma unidade como story points, de modo semelhante ao que vimos em XP.

2.4.2 Principais Artefatos e Eventos

Em Scrum, os dois artefatos principais são o Backlog do Produto e o Backlog do Sprint e os principais eventos são sprints e o planejamento de sprints, conforme descreveremos a seguir.

- O **Backlog do Produto** é uma lista de histórias, ordenada por prioridades. Assim como em XP, as histórias são escritas e priorizadas pelo Dono do Produto e constituem uma descrição resumida das funcionalidades que devem ser implementadas no projeto. É importante mencionar ainda que o Backlog do Produto é um artefato dinâmico, isto é, ele deve ser continuamente atualizado, de forma a refletir mudanças nos requisitos e na visão do produto. Por exemplo, à medida que o desenvolvimento avança, ideias de novas funcionalidades podem surgir, enquanto outras podem perder importância.

Todas essas atualizações devem ser realizadas pelo Dono do Produto. Na verdade, é o fato de ser o dono do Backlog do Produto que faz o Dono do Produto receber esse nome.

- **Sprint** é o nome dado por Scrum para uma iteração. Ou seja, como todo método ágil, Scrum é um método iterativo, no qual o desenvolvimento é dividido em sprints, de até um mês. Ao final de um sprint, deve-se entregar um produto com valor tangível para o cliente. O resultado de um sprint é frequentemente chamado de um produto potencialmente pronto para entrar em produção (*potentially shippable product*). Lembre que o adjetivo potencial não torna a entrada em produção obrigatória, conforme discutido na Seção 2.2.
- **O Planejamento do Sprint** é uma reunião na qual todo o time se reúne para decidir as histórias que serão implementadas no sprint que vai se iniciar. Portanto, ele é o evento que marca o início de um sprint. Essa reunião é dividida em duas partes. A primeira é comandada pelo Dono do Produto. Ele propõe histórias para o sprint e o restante do time decide se tem **velocidade** para implementá-las. A segunda parte é comandada pelos desenvolvedores. Nela, eles quebram as histórias em tarefas e estimam a duração delas. No entanto, o Dono do Produto deve continuar presente nessa parte final, para tirar dúvidas sobre as histórias selecionadas para o sprint. Por exemplo, pode-se decidir cancelar uma história, pois ela se revelou mais complexa ao ser quebrada em tarefas.
- **O Backlog do Sprint** é o artefato gerado ao final do Planejamento do Sprint. Ele é uma lista com as tarefas do sprint, bem como inclui a duração das mesmas. Como o Backlog do Produto, o Backlog do Sprint também é dinâmico. Por exemplo, tarefas podem se mostrar desnecessárias e outras podem surgir, ao longo do sprint. Pode-se também alterar a estimativa de horas previstas para uma tarefa. Porém, o que não pode ser alterado é o **objetivo do sprint** (*sprint goal*), isto é, a lista de histórias que o dono do produto selecionou para o sprint e que o time de desenvolvimento se comprometeu a implementar na duração do mesmo. Assim, Scrum é um método adaptativo a mudanças, mas desde que elas ocorram entre sprints. Ou seja, no time-box de um sprint, a equipe de desenvolvimento deve ter tranquilidade e segurança para trabalhar com uma lista fechada de histórias.

Terminada a reunião de planejamento, tem início o sprint. Ou seja, o time começa a trabalhar na implementação das tarefas do backlog. Além de cross-funcionais,

os times Scrum são **auto-organizáveis**, isto é, eles têm autonomia para decidir como e por quem as histórias serão implementadas.

Ao lado do Backlog do Sprint, costuma-se anexar um quadro com tarefas *a fazer, em andamento e finalizadas*. Esse quadro — também chamado de **Quadro Scrum (Scrum Board)** — pode ser fixado nas paredes do ambiente de trabalho, permitindo que o time tenha diariamente uma sensação visual sobre o andamento do sprint. Veja um exemplo na próxima figura.

Backlog	To Do	Doing	Testing	Done

Figura 2.5: Exemplo de Quadro Scrum, mostrando as histórias selecionadas para o sprint e as tarefas nas quais elas foram quebradas. Cada tarefa nesse quadro pode estar em um dos seguintes estados: a fazer, em andamento, em teste ou concluída.

Uma decisão importante em projetos Scrum envolve os critérios para considerar uma história ou tarefa como concluídas (*done*). Esses critérios devem ser combinados com o time e ser do conhecimento de todos os seus membros. Por exemplo, em um projeto de desenvolvimento de software, para que uma história seja marcada como concluída pode-se exigir a implementação de testes de unidade, que devem estar todos passando, bem como a revisão do código por um outro membro do time. Além disso, o código deve ter sido integrado com sucesso no repositório do projeto. O objetivo desses critérios é evitar que os membros — de forma apressada e valendo-se de código de baixa qualidade — consigam mover suas tarefas para a coluna concluído.

Um outro artefato comum em Scrum é o **Gráfico de Burndown**. A cada dia do sprint, esse gráfico mostra quantas horas são necessárias para se implementar as tarefas que ainda não estão concluídas. Isto é, no dia X do sprint ele informa que restam tarefas a implementar que somam Y horas. Logo, a curva de um gráfico de burndown deve ser declinante, atingindo o valor zero ao final do sprint, caso

ele seja bem sucedido. Mostra-se a seguir um exemplo, assumindo-se um sprint com duração de 15 dias.

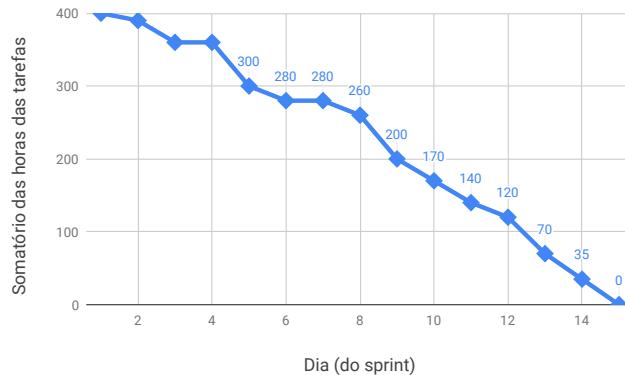


Figura 2.6: Exemplo de Gráfico de Burndown, assumindo um sprint com duração de 15 dias. O sprint foi bem sucedido, pois conseguiu-se implementar todas as tarefas previstas.

Mundo Real: Este livro está sendo escrito usando artefatos e eventos de Scrum. Claro que apenas alguns, pois o livro tem um único autor que, em certa medida, desempenha todos os papéis previstos por Scrum. Logo no início do projeto, os capítulos do livro foram planejados, constituindo assim o Backlog do Produto. A escrita de cada capítulo é considerada como sendo um sprint. Na reunião de Planejamento do Sprint, define-se a divisão do capítulo em seções, que são equivalentes às tarefas. Então começa-se a escrita de cada capítulo, isto é, tem início um sprint. Via de regra, os sprints são planejados para ter uma duração de três meses. Para ficar mais claro, mostra-se a seguir o backlog do sprint atual, bem como o estado de cada tarefa, exatamente no momento em que se está escrevendo este parágrafo:

História	A fazer	Em andamento	Concluídas
Cap. 2 - Processos de Desenvolvimento	Kanban Quando não usar Métodos Ágeis Outros Processos Exercícios	Scrum	Introdução Manifesto Ágil XP

Decidiu-se adotar um método ágil para escrita do livro para minimizar os riscos de desenvolver um produto que não atenda às necessidades de nossos clientes, que, nesta primeira versão, são estudantes e professores brasileiros de disciplinas de Engenharia de Software, principalmente em nível de graduação. Assim, ao final de cada sprint, um capítulo é lançado e divulgado publicamente, de forma a receber feedback. Com isso, evita-se uma solução Waterfall, onde o livro seria escrito por cerca de dois anos, sem receber qualquer feedback. Para finalizar, vamos comentar sobre o critério para conclusão de um capítulo, ou seja, para definir que um capítulo está finalizado (*done*). Esse critério requer a leitura e revisão completa do capítulo, pelo autor do livro. Concluída essa revisão, o capítulo é divulgado preliminarmente para os membros do Grupo de Pesquisa em Engenharia de Software Aplicada, do DCC/UFMG.

2.4.3 Outros Eventos

Vamos agora descrever mais três eventos Scrum, especificamente Reuniões Diárias, Revisão do Sprint e Retrospectiva.

Scrum propõe que sejam realizadas **Reuniões Diárias**, de cerca de 15 minutos, das quais devem participar todos os membros do time. Essas reuniões para serem rápidas devem ocorrer com os membros em pé, daí serem também conhecidas como **reuniões em pé** (*standup meetings*, ou ainda *daily scrum*). Nelas, cada membro do time deve responder a três perguntas: (1) o que ele fez no dia anterior; (2) o que ele pretende fazer no dia corrente; (3) e se ele está enfrentando algum problema mais sério, isto é, um impedimento, na sua tarefa. Essas reuniões têm como objetivo melhorar a comunicação entre os membros do time, fazendo com que eles compartilhem e socializem o andamento do projeto. Por exemplo, dois desenvolvedores podem tomar ciência, durante a reunião diária, que eles vão

começar a modificar o mesmo componente de código. Portanto, seria recomendável que eles se reunissem, separadamente do resto do time, para discutir as modificações que pretendem realizar. E, com isso, minimizar as chances de possíveis conflitos de integração.

A **Revisão do Sprint** (*Sprint Review*) é uma reunião para mostrar os resultados de um sprint. Dela devem participar todos os membros do time e idealmente outros stakeholders, convidados pelo Dono do Produto, que estejam envolvidos com o resultado do sprint. Durante essa reunião o time demonstra, ao vivo, o produto para os clientes. Como resultado, todas as histórias do sprint podem ser aprovadas pelo Dono do Produto. Por outro lado, caso ele detecte problema em alguma história, ela deve voltar para o Backlog do Produto, para ser retrabalhada em um próximo sprint. O mesmo deve ocorrer com as histórias que o time não conseguiu concluir durante o sprint.

A **Retrospectiva** é a última atividade de um sprint. Trata-se de uma reunião do time Scrum, com o objetivo de refletir sobre o sprint que está terminando e, se possível, identificar pontos de melhorias no processo, nas pessoas, nos relacionamentos e nas ferramentas usadas. Apenas para dar um exemplo, como resultado de uma retrospectiva, o time pode acordar sobre a importância de todos estarem presentes, pontualmente, nas reuniões diárias, pois nos últimos sprints alguns membros estão se atrasando. Veja, portanto, que uma retrospectiva não é uma reunião para “lavar a roupa suja” e para membros ficarem discutindo entre si. Se for necessário, isso deve ser feito em particular, em outras reuniões ou com a presença de gerentes da organização. Depois da retrospectiva, o ciclo se repete, com um novo sprint.

Uma característica marcante de todos os eventos Scrum é terem uma duração bem definida, que é chamada de **time-box** da atividade. Por isso, esse termo aparece sempre em documentos Scrum. Por exemplo, veja essa frase do Scrum Guide oficial: “o coração do método Scrum é um sprint, que tem um time-box de um mês ou menos e durante o qual um produto”done“, usável e que potencialmente pode ser colocado em produção é criado” ([link](#)). O objetivo da fixação de *time boxes* é criar um fluxo contínuo de trabalho, bem como fomentar o compromisso da equipe com o sucesso do sprint e evitar a perda de foco.

A próxima tabela mostra o time-box dos eventos Scrum. No caso de eventos com um time-box máximo (exemplo: planejamento do sprint), o valor recomendado refere-se a um sprint de um mês. Se o sprint for menor, o time-box sugerido deve ser também menor.

Evento	Time-box
Planejamento do Sprint	máximo de 8 horas
Sprint	menos de 1 mês
Reunião Diária	15 minutos
Revisão do Sprint	máximo de 4 horas
Retrospectiva	máximo de 3 horas

2.4.4 Perguntas Frequentes

Antes de concluir a seção, vamos responder algumas perguntas sobre Scrum:

O que significa a palavra Scrum? O nome não é uma sigla, mas uma referência à “reunião” de jogadores realizada em uma partida de rugby para decidir quem vai ficar com a bola, após uma infração involuntária.

O que é um squad? Esse termo é um sinônimo para time ágil ou time Scrum. O nome foi popularizado pela Spotify. Assim como times Scrum, squads são pequenos, cross-funcionais e auto-organizáveis. É comum ainda usar o nome tribo para denotar um conjunto de squads.

O Dono do Produto pode ser um comitê? Em outras palavras, pode existir mais de um Dono de Produto em um time Scrum? A resposta é não. Apenas um membro do time exerce essa função. O objetivo é evitar decisões por comitê, que tendem a gerar produtos lotados de funcionalidades, mas que foram implementadas apenas para atender a determinados membros do comitê. Porém, nada impede que o Dono do Produto faça a ponte entre o time e outros usuários com amplo domínio da área do produto que está sendo construído. Na verdade, essa é uma tarefa esperada de Donos de Produto, pois às vezes existem requisitos que são do domínio de apenas alguns colaboradores da organização. Cabe então ao Dono do Produto intermediar as conversas entre os desenvolvedores e tais usuários.

O Scrum Master deve exercer seu papel em tempo integral? Idealmente, sim. Porém, em times maduros, que conhecem e praticam Scrum há bastante tempo, às vezes não é necessário ter um Scrum Master em tempo integral. Nesses casos, existem duas possibilidades: (1) permitir que o Scrum Master desempenhe esse papel em mais de um time Scrum; (2) alocar a responsabilidade de Scrum Master a um dos membros do time. No entanto, caso a segunda alternativa seja adotada, o Scrum Master não deve ser também o Dono do Produto. A razão é que uma das responsabilidades do Scrum Master é exatamente acompanhar e auxiliar o Dono do Produto em suas tarefas de escrever e priorizar histórias do usuário.

O Scrum Master precisa ter um diploma de nível superior em um curso da área de Computação? Não, pois sua função envolve remover impedimentos e assegurar que o time esteja seguindo os princípios de Scrum. Portanto, ele não é um resolvidor de problemas técnicos, tais como bugs, uso correto de frameworks, implementação de funcionalidades, etc. Por outro lado, isso não impede que um desenvolvedor técnico, com um curso superior na área de Computação, assuma as funções de Scrum Master, como vimos na resposta anterior. Existem também certificações para Scrum Master, as quais podem ser requeridas por empresas que decidem adotar Scrum.

Além de histórias, quais outros itens podem fazer parte do Backlog do Produto? Também podem ser cadastrados no Backlog do Produto itens como bugs — principalmente aqueles mais complexos e que demandam dias para serem resolvidos — e também melhorias em histórias já implementadas.

Existem gerentes quando se usa Scrum? A resposta é sim! De fato, times Scrum são autônomos para implementar as histórias priorizadas pelo Dono do Produto. Porém, um projeto demanda diversas outras decisões que devem ser tomadas em um nível gerencial. Dentre elas, podemos citar as seguintes:

- Contratar e alocar membros para os times Scrum; ou seja, os desenvolvedores não tem autonomia para escolher em quais times eles vão trabalhar. Essa é uma decisão de mais alto nível e, portanto, tomada por gerentes.
- Decidir os objetivos e responsabilidades de cada time, incluindo, o sistema — ou parte de um sistema — que o time irá desenvolver usando Scrum. Por exemplo, um time não decide, por conta própria, que a organização precisa de um novo sistema de contabilidade e então começa a desenvolvê-lo. Essa é uma decisão estratégica da organização, que é tomada pelos seus gerentes e executivos.
- Gerenciar e administrar questões de recursos humanos, incluindo contratações de novos funcionários, desligamentos de funcionários, promoções, transferências, treinamentos, etc.
- Avaliar se os resultados produzidos pelos times Scrum estão, de fato, gerando benefícios e valor para a organização.

2.5 Kanban

A palavra japonesa *kanban* significa “cartão visual” ou “cartão de sinalização”. Desde a década de 50, o nome também é usado para denotar o processo de produção *just-in-time* usado em fábricas japonesas, principalmente naquelas da Toyota, onde

ele foi usado pela primeira vez. O processo também é conhecido como Sistema de Produção da Toyota (TPS) ou, mais recentemente, por manufatura *lean*. Em uma linha de montagem, os cartões são usados para controlar o fluxo de produção.

No caso de desenvolvimento de software, Kanban foi usado pela primeira vez na Microsoft, em 2004, como parte de um esforço liderado por David Anderson, então um funcionário da empresa ([link](#)). Segundo Anderson, “Kanban é um método que ajuda times de desenvolvimento a trabalhar em ritmo sustentável, eliminando desperdício, entregando valor com frequência e fomentando uma cultura de melhorias contínuas”.

Para começar a explicar Kanban, vamos usar uma comparação com Scrum. Primeiro, Kanban é mais simples do que Scrum, pois não usa nenhum dos eventos de Scrum, incluindo sprints. Também, não existe nenhum dos papéis (Dono do Produto, Scrum Master, etc.), pelo menos da forma rígida preconizada por Scrum. Por fim, não existe nenhum dos artefatos Scrum, com uma única e central exceção: o quadro de tarefas, que é chamado de **Quadro Kanban** (*Kanban Board*), e que inclui também o Backlog do Produto.

O Quadro Kanban é dividido em colunas, da seguinte forma:

- A primeira coluna é o backlog do produto. Como em Scrum, usuários escrevem as histórias, que vão para o Backlog.
- As demais colunas são os passos que devem ser seguidos para transformar uma história do usuário em uma funcionalidade executável. Por exemplo, pode-se ter colunas como Especificação, Implementação e Revisão de Código. A ideia, portanto, é que as histórias sejam processadas passo a passo, da esquerda para a direita, como em uma linha de montagem. Além disso, cada coluna é dividida em duas sub-colunas: “em execução” e “concluídas”. Por exemplo, a coluna implementação tem duas sub-colunas: tarefas em implementação e tarefas implementadas. As tarefas concluídas em um passo estão aguardando serem puxadas, por um membro do time, para o próximo passo. Por isso, Kanban é chamado de um sistema *pull*.

Mostra-se abaixo um exemplo de Quadro Kanban. Observe que existe uma história no backlog (H3), além de uma história (H2) que já foi puxada por algum membro do time para o passo de especificação. Existem ainda quatro tarefas (T6 a T9) que foram especificadas a partir de uma história anterior. Continuando, existem duas tarefas em implementação (T4 e T5) e existe uma tarefa implementada e aguardando ser puxada para revisão de código (T3). No último passo, existe uma tarefa em revisão (T2) e uma tarefa cujo processamento está concluído (T1).

Por enquanto, não se preocupe com a sigla WIP que aparece em todos os passos, exceto no backlog. Vamos explicá-la em breve. Além disso, representamos as histórias e tarefas pelas letras H e T, respectivamente. Porém, em um quadro real, ambas são cartões auto-adesivos com uma pequena descrição. O Quadro Kanban pode ser assim montado em uma das paredes do ambiente de trabalho do time.

Backlog	Especificação WIP		Implementação WIP		Revisão de Código WIP	
H3	em espec. H2	especificadas T6 T7 T8 T9	em implementação T4 T5	implementadas T3	em revisão T2	revisadas T1

Agora, mostraremos uma evolução do projeto. Isto é, alguns dias depois, o Quadro Kanban passou para o seguinte estado (as tarefas que avançaram no quadro estão em vermelho).

Backlog	Especificação WIP		Implementação WIP		Revisão de Código WIP	
H3	em espec.	especificadas T8 T9 T10 T11 T12	em implementação T4 T5 T6 T7	implementadas	em revisão T3	revisadas T1 T2

Veja que a história H2 desapareceu, pois foi quebrada em três tarefas (T10, T11 e T12). O objetivo da fase de especificação é exatamente transformar uma história em uma lista de tarefas. Continuando, T6 e T7 — que antes estavam esperando — entraram em implementação. Já T3 entrou na fase de revisão de código. Por fim, a revisão de T2 terminou. Veja ainda que, neste momento, não existe nenhuma tarefa implementada e aguardando ser puxada para revisão.

Como em outros métodos ágeis, times Kanban são auto-organizáveis. Isso significa que eles têm autonomia para definir qual tarefa vai ser puxada para o próximo passo. Eles também são cross-funcionais, isto é, devem incluir membros capazes de realizar todos os passos do Quadro Kanban.

Por fim, resta explicar o conceito de **Límites WIP** (*Working in Progress*). Via de regra, métodos de gerenciamento de projetos têm como objetivo garantir um ritmo sustentável de trabalho. Para isso, deve-se evitar duas situações extremas: (1) o time ficar ocioso boa parte do tempo, sem tarefa para realizar; ou (2) o time ficar sobrecarregado de trabalho e, por isso, não conseguir produzir software de qualidade. Para evitar a segunda situação — sobrecarga de trabalho — Kanban propõe um limite máximo de tarefas que podem estar em cada um dos passos

de um Quadro Kanban. Esse limite é conhecido pelo nome Limite WIP, isto é, trata-se do limite máximo de cartões presentes em cada passo, contando aqueles na primeira coluna (em andamento) e aqueles na segunda coluna (concluídos) do passo. A exceção é o último passo, onde o WIP aplica-se apenas à primeira sub-coluna, já que não faz sentido aplicar um limite ao número de tarefas concluídas pelo time.

Abaixo, reproduzimos o último Quadro Kanban, mas com os limites WIP. Eles são os números que aparecem abaixo do nome de cada passo, na primeira linha do quadro. Ou seja, nesse Quadro Kanban, admite-se um máximo de 2 histórias em especificação; 5 tarefas em implementação; e 3 tarefas em revisão. Vamos deixar para explicar o limite WIP do passo Especificação por último. Mas podemos ver que existem 4 tarefas em implementação (T4, T5, T6 e T7). Portanto, abaixo do WIP desse passo, que é igual a 5 tarefas. Em Revisão de Código, o limite é de 3 tarefas e também está sendo respeitado, pois existe apenas uma tarefa em revisão (T3). Veja que para fins de verificar o limite WIP contam-se as tarefas em andamento (1a sub-coluna de cada passo) e concluídas (2a sub-coluna de cada passo), com exceção do último passo, no qual consideram-se apenas as tarefas da primeira sub-coluna (T3, no exemplo).

Backlog	Especificação 2		Implementação 5		Revisão de Código 3	
H3	em espec. T8 T9 T10 T11 T12	especificadas	em implementação T4 T5 T6 T7	implementadas	em revisão T3	revisadas T1 T2

Agora vamos explicar o WIP do passo Especificação. Para verificar o WIP desse passo, deve-se somar as histórias em especificação (zero no quadro anterior) e as histórias que já foram especificadas. No caso, temos duas histórias especificadas. Ou seja, T8 e T9 são tarefas que resultaram da especificação de uma mesma história. E as tarefas T10, T11 e T12 são resultado da especificação de uma segunda história. Portanto, para fins do cálculo de WIP, temos duas histórias no passo, o que está dentro do seu limite, que também é 2. Para facilitar a visualização, costuma-se representar as tarefas resultantes da especificação de uma mesma história em uma única linha. Seguindo esse padrão, para calcular o WIP do passo Especificação, deve-se somar as histórias da primeira subcoluna (zero, no nosso exemplo) com o número de linhas na segunda coluna (duas).

Ainda no quadro anterior, e considerando os limites WIP, tem-se que:

- A história H3, que está no backlog, não pode ser puxada para o passo Especificação, pois o WIP desse passo está no limite.

- Uma das tarefas já especificadas (T8 a T12) pode ser puxada para implementação, pois o WIP desse passo está em 4, enquanto o limite é 5.
- Uma ou mais tarefas em implementação (T4 a T7) podem ser finalizadas, o que não altera o WIP do passo.
- A revisão de T3 pode ser finalizada.

Apenas reforçando, o objetivo dos limites WIP é evitar que os times Kanban fiquem sobrecarregados de trabalho. Quando um desenvolvedor tem muitas tarefas para realizar — porque os limites WIP não estão sendo respeitados — a tendência é que ele não consiga concluir nenhuma dessas tarefas com qualidade. Como usual em qualquer atividade humana, quando assumimos muitos compromissos, a qualidade de nossas entregas cai muito. Kanban reconhece esse problema e, para que ele não ocorra, cria uma “trava” automática para impedir que os times aceitem trabalhos além da sua capacidade de entrega. Essas “travas”, que são os limites WIP, servem para uso interno do time e, mais importante ainda, para uso externo. Ou seja, elas são o instrumento de que um time dispõe para recusar trabalho extra que está sendo “empurrado” de cima para baixo por gerentes da organização, por exemplo.

2.5.1 Calculando os Limites WIP

Resta-nos agora explicar como os limites WIP são definidos. Existe mais de uma alternativa, mas vamos adotar uma adaptação de um algoritmo proposto por Eric Brechner — um engenheiro da Microsoft — em seu livro sobre o uso de Kanban no desenvolvimento de software ([link](#)). O algoritmo é descrito a seguir.

Primeiro, temos que estimar quanto tempo em média uma tarefa vai ficar em cada passo do Quadro Kanban. Esse tempo é chamado de **lead time (LT)**. No nosso exemplo, vamos supor os seguintes valores:

- LT(especificação) = 5 dias
- LT(implementação) = 12 dias
- LT(revisão) = 6 dias

Veja que essa estimativa considera uma tarefa média, pois sabemos que vão existir tarefas mais complexas e mais simples. Veja ainda que o *lead time* inclui o tempo em fila, isto é, o tempo que a tarefa vai ficar na 2a sub-coluna dos passos do Quadro Kanban aguardando ser puxada para o passo seguinte.

Em seguida, deve-se estimar o **throughput (TP)** do passo com maior *lead time* do Quadro Kanban, isto é, o número de tarefas produzidas por dia nesse passo.

No nosso exemplo, e na maioria dos projetos de desenvolvimento de software, esse passo é a Implementação. Assim, suponha que o time seja capaz de sustentar a implementação de 8 tarefas por mês. O **throughput** desse passo é então: $8 / 21 = 0.38$ tarefas/dia. Veja que consideramos que um mês tem 21 dias úteis.

Por fim, o WIP de cada passo é assim definido:

$$\mathbf{WIP(\text{passo}) = TP * LT(\text{passo})}.$$

onde throughput refere-se ao throughput do passo mais lento, conforme calculado no item anterior.

Logo, teremos os seguintes resultados:

- $\mathbf{WIP(\text{especificação}) = 0.38 * 5 = 1.9}$
- $\mathbf{WIP(\text{implementação}) = 0.38 * 12 = 4.57}$
- $\mathbf{WIP(\text{revisão}) = 0.38 * 6 = 2.29}$

Arredondando para cima, os resultados finais ficam assim:

- $\mathbf{WIP(\text{especificação}) = 2}$
- $\mathbf{WIP(\text{implementação}) = 5}$
- $\mathbf{WIP(\text{revisão}) = 3}$

No algoritmo proposto por Eric Brechner, sugere-se ainda adicionar uma margem de erro de 50% nos WIPs calculados, para acomodar variações no tamanho das tarefas, tarefas bloqueadas devido a fatores externos, etc. No entanto, como nosso exemplo é ilustrativo, não vamos ajustar os WIPs que foram calculados acima.

Conforme afirmado, limites WIPs são o recurso oferecido por Kanban para garantir um ritmo de trabalho sustentável e a entrega de sistemas de software com qualidade. O papel desses limites é contribuir para que os desenvolvedores não fiquem sobrecarregados de tarefas e, consequentemente, propensos a baixar a qualidade de seu trabalho. Na verdade, todo método de desenvolvimento de software tende a oferecer esse tipo de recurso. Por exemplo, em Scrum existe o conceito de sprints com time-boxes definidos, cujo objetivo é evitar que times aceitem trabalhar em histórias que ultrapassam a sua velocidade de entrega. Adicionalmente, uma vez iniciado, o objetivo de um sprint não pode ser alterado, de forma a blindar o time de mudanças diárias de prioridade. No caso de métodos Waterfall, o recurso para garantir um fluxo de trabalho sustentável e de qualidade é a existência de uma fase detalhada de especificação de requisitos. Com essa fase, a intenção era oferecer aos desenvolvedores uma ideia clara do sistema que eles deveriam implementar.

2.5.2 Lei de Little

O procedimento para cálculo de WIPs explicado anteriormente é uma aplicação direta da **Lei de Little**, um dos resultados mais importantes da Teoria de Filas ([link](#)). A Lei de Little diz que o número de itens em um sistema de filas é igual à taxa de chegada desses itens multiplicado pelo tempo que cada item fica no sistema. Traduzindo para o nosso contexto, o sistema é um passo de um processo Kanban e os itens são tarefas. Assim, temos também que:

- WIP: número de tarefas em um dado passo de um processo Kanban
- Throughput (TP): taxa de chegada dessas tarefas nesse passo
- Lead Time (LT): tempo que cada tarefa fica nesse passo

Ou seja, de acordo com a Lei de Little: **WIP = TP * LT**. Visualmente, podemos representar a Lei de Little da seguinte forma:

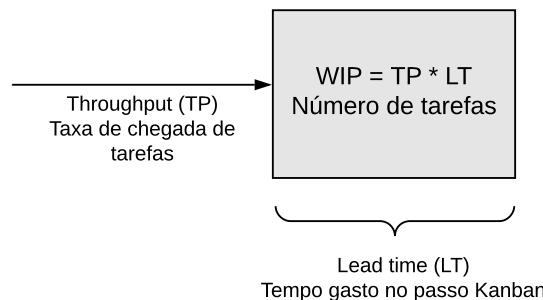


Figura 2.7: Lei de Little: $WIP = TP * LT$

2.5.3 Perguntas Frequentes

Antes de concluir, vamos responder algumas perguntas sobre Kanban:

Quais são os papéis que existem em Kanban? Ao contrário de Scrum, Kanban não define uma lista fixa de papéis. Cabe ao time e à organização definir os papéis que existirão no processo de desenvolvimento, tais como Dono do Produto, Testadores, etc.

Como as histórias dos usuários são priorizadas? Kanban é um método de desenvolvimento mais leve que Scrum e mesmo que XP. Um dos motivos é que ele não define critérios de priorização de histórias. Como respondido na pergunta

anterior, não se define, por exemplo, que o time tem que possuir um Dono do Produto, responsável por essa priorização. Veja que essa é uma possibilidade, isto é, pode existir um Dono do Produto em times Kanban. Mas outras soluções também são possíveis, como priorização externa, por um gerente de produto.

Times Kanban podem realizar eventos típicos de Scrum, como reuniões diárias, revisões e retrospectivas? Sim, apesar de Kanban não prescrever a realização desses eventos. Porém, também não há um voto explícito aos mesmos. Cabe ao time decidir quais eventos são importantes, quando eles devem ser realizados, com que duração, etc.

Em vez de um Quadro Kanban físico, com adesivos em uma parede ou em um quadro branco, pode-se usar um software para gerenciamento de projetos? Kanban não proíbe o uso de software de gerenciamento de projetos. Porém, recomenda-se o uso de um quadro físico, pois um dos princípios mais importantes de Kanban é a visualização do trabalho pelo time, de forma que seus membros possam a qualquer momento tomar conhecimento do trabalho em andamento e de possíveis problemas e gargalos que estejam ocorrendo. Em alguns casos, chega-se a recomendar a adoção de ambas soluções: um quadro físico, mas com um backup em um software de gerenciamento de projetos, que possa ser acessado pelos gerentes e executivos da organização.

2.6 Quando não Usar Métodos Ágeis?

Apesar de métodos ágeis — como aqueles estudados nas seções anteriores — terem alcançado um sucesso inquestionável, é bom lembrar que não existe bala-de-prata em Engenharia de Software, conforme comentamos no Capítulo 1. Assim, neste capítulo vamos discutir cenários e domínios nos quais práticas ágeis podem não ser adequadas.

No entanto, a pergunta que abre essa seção não admite uma resposta simples, como, por exemplo, sistemas das áreas X, Y e Z não devem usar métodos ágeis; e os demais devem usar. Em outras palavras, sistemas de qualquer área podem se beneficiar de pelo menos algumas das práticas propostas por métodos ágeis. Por outro lado, existem práticas que não são recomendadas para determinados tipos de sistemas, organizações e contextos. Assim, vamos responder a pergunta proposta em uma granularidade mais fina. Isto é, vamos comentar a seguir sobre quando **não** usar determinadas práticas de desenvolvimento ágil.

- **Design Incremental.** Esse tipo de design faz sentido quando o time tem uma primeira visão do design do sistema. Se o time não tem essa visão, ou o domínio do sistema é novo e complexo, ou o custo de mudanças futuras é

muito alto, recomenda-se adotar uma fase de design e análise inicial, antes de partir para iterações que requeiram implementações de funcionalidades.

- **Histórias do Usuário.** Histórias são um método leve para especificação de requisitos, que depois são clarificados com o envolvimento cotidiano de um representante dos clientes no projeto. Porém, em certos casos, pode ser importante ter uma especificação detalhada de requisitos no início do projeto, principalmente se ele for um projeto de uma área totalmente nova para o time de desenvolvedores.
- **Envolvimento do Cliente.** Se os requisitos do sistema são estáveis e de pleno conhecimento do time de desenvolvedores, não faz sentido ter um Representante dos Clientes ou Dono do Produto integrado ao time. Por exemplo, esse papel não é importante no desenvolvimento de um compilador para uma linguagem conhecida, com uma gramática e semântica consolidadas.
- **Documentação Leve e Simplificada.** Em certos domínios, documentações detalhadas de requisitos e de projeto são mandatórias. Por exemplo, sistemas cujas falhas podem causar a morte de seres humanos, como aqueles das áreas médicas e de transporte, costumam demandar certificação por uma entidade externa, que pode exigir uma documentação detalhada, além do código fonte.
- **Times Auto-organizáveis.** Times ágeis são autônomos e empoderados para trabalhar sem interferências durante o time-box de uma iteração. Consequentemente, eles não precisam prestar contas diárias para os gerentes e executivos da organização. No entanto, essa característica pode ser incompatível com os valores e cultura de certas organizações, principalmente aquelas com uma tradição de níveis hierárquicos e de controle rígidos.
- **Contratos com Escopo Aberto.** Em contratos com escopo aberto, a remuneração é por hora trabalhada. Assim, a empresa que assina o contrato não tem — no momento da assinatura — uma ideia precisa de quais funcionalidades serão implementadas e nem do prazo e custo do sistema. Algumas organizações podem não se sentir seguras para assinar esse tipo de contrato, principalmente quando elas não têm uma experiência prévia com desenvolvimento ágil ou referências confiáveis sobre a empresa contratada.

Para concluir, é importante mencionar que duas práticas ágeis são atualmente adotadas na grande maioria de projetos de software:

- Times pequenos, pois o esforço de sincronização cresce muito quando os times são compostos por dezenas de membros.

- Iterações (ou sprints), mesmo que com duração maior do que aquela típica de métodos ágeis. Por exemplo, iterações com duração de dois ou três meses, em vez de iterações com menos de 30 dias. Na verdade, entre o surgimento de Waterfall e de métodos ágeis, alguns métodos iterativos foram propostos, isto é, métodos com pontos de validação ao longo do desenvolvimento. Na próxima seção, iremos estudar dois destes métodos.

2.7 Outros Métodos Iterativos

A transição entre Waterfall — dominante nas décadas de 70 e 80 — e métodos ágeis — que começaram a surgir na década de 90, mas que só ganharam popularidade no final dos anos 2000 — foi gradativa. Como em métodos ágeis, esses métodos surgidos nesse período de transição possuem o conceito de iterações. Ou seja, eles não são estritamente sequenciais, como em Waterfall. Porém, as iterações tem duração maior do que aquela usual em desenvolvimento ágil. Em vez de poucas semanas, elas duram alguns meses. Por outro lado, eles preservam características relevantes de Waterfall, como ênfase em documentação e em uma fase inicial de levantamento de requisitos e depois de design.

Um exemplo de proposta de processo surgida nesta época é o **Modelo em Espiral**, proposto por Barry Boehm, em 1986 ([link](#)). Nesse modelo, um sistema é desenvolvido na forma de uma espiral de iterações. Cada iteração, ou “volta completa” na espiral, inclui quatro etapas (veja também a próxima figura):

- Definição de objetivos e restrições, tais como custos, cronogramas, etc.
- Avaliação de alternativas e análise de riscos. Por exemplo, pode-se chegar à conclusão de que é mais interessante comprar um sistema pronto, do que desenvolver o sistema internamente.
- Desenvolvimento e testes, por exemplo, usando Waterfall. Ao final dessa etapa, deve-se gerar um protótipo que possa ser demonstrado aos usuários do sistema.
- Planejamento da próxima iteração ou então tomar a decisão de parar, pois o que já foi construído é suficiente para atender às necessidades da organização.

Assim, o Modelo em Espiral produz, a cada iteração, versões mais completas de um sistema, começando da versão gerada no centro da espiral. Porém, cada iteração, somando-se as quatro fases, pode levar de 6 a 24 meses. Portanto, mais do que em XP e Scrum. Uma outra característica importante é a existência de uma fase explícita de análise de riscos, da qual devem resultar medidas concretas para mitigar os riscos identificados no projeto.

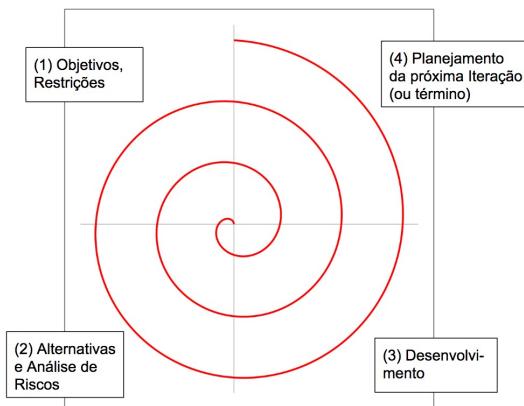


Figura 2.8: Modelo Espiral. Cada iteração é dividida em quatro etapas.

O Processo Unificado (UP), proposto no final da década de 90, é outro exemplo de método iterativo de desenvolvimento. UP foi proposto por profissionais ligados a uma empresa de consultoria e de ferramentas de apoio ao desenvolvimento de software chamada Rational, que em 2003 seria comprada pela IBM. Por isso, o método é também chamado de **Rational Unified Process (RUP)**.

Devido a suas origens, UP é vinculado a duas tecnologias específicas:

- UP é baseado na linguagem de modelagem UML. Todos os seus resultados são documentados e representados usando-se diagramas gráficos de UML. No Capítulo 4, iremos estudar UML com mais calma. Por enquanto, vamos ressaltar que a proposta era ter uma linguagem de modelagem “unificada” (UML) e também um processo “unificado” (UP), ambos propostos pelo mesmo grupo de profissionais.
- UP é associado a ferramentas de apoio ao projeto e análise de software, conhecidas como **ferramentas CASE** (*Computer-Aided Software Engineering*). O nome é uma analogia com ferramentas CAD (*Computer Aided-Design*), usadas em projetos de Engenharia Civil, Engenharia Mecânica, Arquitetura, etc. A ideia era que o projeto e análise de um sistema deveriam ser integralmente baseados em diagramas UML. Mas esses diagramas não seriam desenhados em papel e sim usando-se ferramentas computacionais (veja um exemplo de tela de uma ferramenta CASE na próxima figura). A Rational, além de propor o método UP, também desenvolvia e vendia licenças de uso de ferramentas CASE.

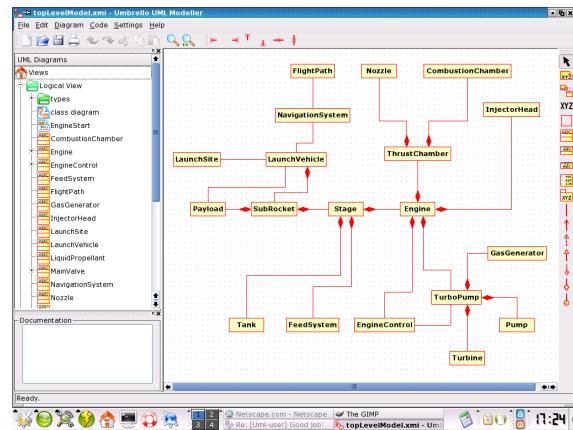


Figura 2.9: Exemplo de diagrama UML desenhado usando uma ferramenta CASE. Imagem extraída de [link](#).

UP propõe que o desenvolvimento seja decomposto nas seguintes fases:

- **Inception** (às vezes, traduzida como iniciação ou concepção): que inclui análise de viabilidade, definição de orçamentos, análise de riscos e definição de escopo do sistema. Ao final dessa fase, o caso de negócios (*business case*) do sistema deve estar bem claro. Pode-se inclusive decidir que não vale a pena desenvolver o sistema, mas sim comprar um sistema pronto.
- **Elaboração**: que incluiu especificação de requisitos (via casos de uso de UML), definição da arquitetura do sistema, bem como de um plano para o seu desenvolvimento. Ao final dessa fase, todos os riscos identificados na fase anterior devem estar devidamente controlados e mitigados.
- **Construção**: na qual se realiza o projeto de mais baixo nível, implementação e testes do sistema. Ao final dessa fase, deve ser disponibilizado um sistema funcional, incluindo documentação e manuais, que possam ser validados pelos usuários.
- **Transição**: na qual ocorre a disponibilização do sistema para produção, incluindo a definição de todas as rotinas de implantação, como políticas de backup, migração de dados de sistemas legados, treinamento da equipe de operação, etc.

Assim como no Modelo Espiral, pode-se repetir várias vezes o processo; ou seja, o desenvolvimento é incremental, com novas funcionalidades sendo entregues a cada iteração. Adicionalmente, pode-se repetir cada uma das fases. Por exemplo,

construção — em uma dada iteração — pode ser dividida em duas sub-fases, cada uma construindo uma parte do produto. A próxima figura ilustra o modelo de iterações de UP.

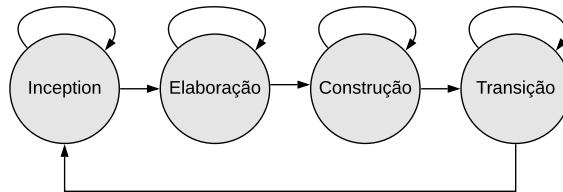


Figura 2.10: Fases e iterações do Processo Unificado (UP). Repetições são possíveis em cada fase (auto-laços). E também pode-se repetir todo o fluxo (laço externo), para gerar mais um incremento de produto.

UP define também um conjunto de disciplinas de engenharia que incluem por exemplo: modelagem de negócios, definição de requisitos, análise e design, implementação, testes e implantação. Essas disciplinas — ou fluxos de trabalho — podem ocorrer em qualquer fase. Porém, espera-se que algumas disciplinas sejam mais intensas em determinadas fases, como mostra a próxima figura. No projeto ilustrado, tarefas de modelagem de negócio estão concentradas nas fases iniciais do projeto (inception e elaboração) e quase não ocorrem nas fases seguintes. Por outro lado, implementação está concentrada na fase de Construção.

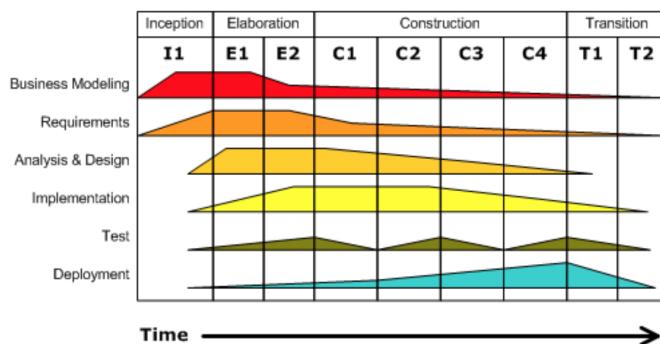


Figura 2.11: Fases (na horizontal) e disciplinas (na vertical) de um projeto desenvolvido usando UP. A área da curva mostra a intensidade da disciplina durante cada fase (imagem da [Wikipedia](#), licença: domínio público).

Bibliografia

Kent Beck, Cynthia Andres. Extreme Programming Explained: Embrace Change. Addison-Wesley, 2nd edition, 2004.

Kent Beck. Embracing Change with Extreme Programming. IEEE Computer, vol. 32, issue 10, p. 70-77, 1999.

Kent Beck, Martin Fowler. Planning Extreme Programming. Addison-Wesley, 2000.

Ken Schwaber, Jeff Sutherland. The Scrum Guide, 2017.

Kenneth Rubin. Essential Scrum: A Practical Guide to the Most Popular Agile Process. Addison-Wesley, 2012

Eric Brechner. Agile Project Management with Kanban. Microsoft Press, 2015.

David Anderson. Kanban. Blue Hole Press, 2013.

Ian Sommerville. Engenharia de Software. Pearson, 10a edição, 2019.

Hans van Vliet. Software Engineering: Principles and Practice. 3rd edition. Wiley, 2008.

Armando Fox, David Patterson. Construindo Software como Serviço: Uma Abordagem Ágil Usando Computação em Nuvem. Strawberry Canyon LLC. 1a edição, 2014.

Exercícios de Fixação

1. Como XP preconiza que devem ser os contratos de desenvolvimento de software.
2. Quais as diferenças entre XP e Scrum?
3. Times Scrum são ditos cross-funcionais e auto-organizáveis. Por que? Defina esses termos.
4. Qual a diferença entre as histórias do “topo” e do “fundo” do Backlog do Produto, em Scrum?
5. O que são e para que servem story points?
6. Em Scrum, qual a diferença entre uma sprint review e uma retrospectiva?

7. Um sprint pode ser cancelado? Se sim, por quem e por qual motivo? Para responder a essa questão, consulte antes o Scrum Guide ([link](#)), que é o guia que documenta a versão oficial de Scrum.
8. Procure pensar em um sistema de uma área da qual tenha algum conhecimento. (a) Escreva então uma história para esse sistema (veja que histórias são especificações resumidas de funcionalidades, com 2-3 sentenças). (b) Em seguida, quebre a história que definiu em algumas tarefas (de forma semelhante ao que fizemos no sistema similar ao Stack Overflow, usado como exemplo na seção sobre XP). (c) Existem dependências entre essas tarefas? Ou elas podem ser implementadas em qualquer ordem?
9. Suponha dois times, A e B, atuando em projetos diferentes, contratados por empresas distintas, sem conexões entre eles. Porém, ambos os times adotam sprints de 15 dias e ambos possuem 5 desenvolvedores. Nos seus projetos, o time A considera que sua velocidade é de 24 pontos. Já o time B assume uma velocidade de 16 pontos. Pode-se afirmar que A é 50% mais produtivo que B? Justifique sua resposta.
10. Quais são as principais diferenças entre Scrum e Kanban?
11. Quais são as diferenças entre um Quadro Scrum e um Quadro Kanban?
12. Qual o erro existe no seguinte Quadro Kanban?

Backlog	Especificação (2)	Implementação (5)	Validação (3)	
X X	X	X X X X	X	X
X X			X	X
X X			X	X

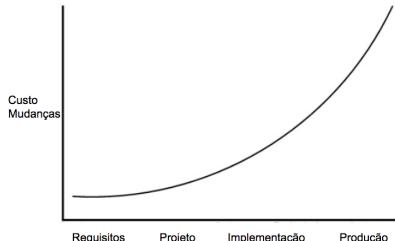
13. Suponha o seguinte quadro Kanban. Neste momento, o time não consegue trabalhar na especificação de novas histórias, pois o WIP do passo Especificação está sendo totalmente preenchido por itens esperando movimentação para o passo seguinte (Implementação). O que seria mais recomendado neste momento: (a) desrespeitar o WIP e já puxar uma nova história do Backlog para Especificação; ou (b) ajudar o time nas três tarefas em Validação, de forma a desbloquear o fluxo do processo.

Backlog	Especificação (2)	Implementação (5)	Validação (3)	
X		X X X	X	X
X X		X X X	X	X
X			X	X

14. Por que se recomenda que os limites WIP calculados usando a Lei de Little sejam incrementados, por exemplo em 50%, de forma a admitir uma margem de erro? Em outras palavras, quais eventos podem originar esses erros na estimativa dos WIPs?
15. Seja um processo Kanban, dividido em quatro passos. A tabela abaixo informa o *lead time* de cada um deles e o throughput do passo C, que é o passo mais lento. Com base nesses valores, calcule o WIP de cada passo (última coluna da tabela).

Passo	Lead Time (médio, dias)	Throughput (tarefas/dia)	WIP
A	4	-	-
B	3	-	-
C	10	0.5	-
D	5	-	-

16. Descreva os principais recursos oferecidos por Waterfall, Scrum e Kanban para controlar riscos e garantir um fluxo de trabalho sustentável e que propicie o desenvolvimento de software com qualidade.
17. Seja o seguinte gráfico, que mostra — para um determinado sistema — como os custos de mudanças variam conforme a fase do desenvolvimento em que elas são realizadas. (a) Qual método de desenvolvimento você recomendaria para esse sistema? Justifique sua resposta. (b) Sistemas de quais domínios podem ter uma curva de custos de mudanças semelhante a essa?



18. O artigo “*Development and Deployment at Facebook*” ([link](#)) apresenta os métodos e práticas de desenvolvimento de software usados no Facebook. Na sua primeira seção (páginas 2-3; figura 2), os autores fazem uma distinção entre alguns métodos de desenvolvimento, baseando-se na frequência com que versões de um sistema são liberadas para uso quando se adota cada um deles.

Complete a seguinte tabela informando a frequência de releases mencionada no artigo para alguns métodos e políticas de liberação de software.

Método	Frequência de novas releases
Waterfall	
Evolucionário	
Ágil	
Facebook	
Deployment Contínuo	

19. Por que métodos como o Processo Unificado (UP) e Espiral não são considerados ágeis? E qual a diferença deles para o Modelo Waterfall?

Capítulo 3

Modelos

All models are wrong, but some models are useful. So the question you need to ask is not “Is the model true?” (it never is) but “Is the model good enough for this particular application?” – George Box

Este capítulo inicia com uma apresentação genérica sobre modelos de software (Seção 4.1). Em seguida, apresentamos uma visão geral sobre UML, que é a notação gráfica mais utilizada para construção de modelos de software (Seção 4.2). Também deixamos claro que vamos estudar UML visando a criação de esboços de software (*sketches*) e não desenhos técnicos detalhados (*blueprints*). Nas seções seguintes, apresentamos quatro diagramas UML com um maior nível de detalhes: Diagramas de Classes (Seção 4.3), Diagramas de Pacotes (Seção 4.4), Diagramas de Sequência (Seção 4.5) e Diagramas de Atividades (Seção 4.6).

3.1 Modelos de Software

Como vimos no capítulo anterior, requisitos documentam “o que” um sistema deve fazer, valendo-se de um nível de abstração próximo do problema e de seus usuários. Por outro lado, o código fonte é uma representação concreta, de baixo nível e executável do comportamento de um sistema. Portanto, existe uma lacuna entre esses dois mundos: requisitos e código fonte. Para preencher essa lacuna, desde a fundação da área, Engenheiros de Software investem na criação de **modelos**, os quais são criados para ajudar no entendimento e análise de um sistema. Para cumprir essa missão, os modelos usados em Engenharia de Software são mais

detalhados do que requisitos, mas ainda menos complexos do que o código fonte de um sistema.

Modelos são largamente usados também em outras Engenharias. Por exemplo, uma engenheira civil pode decidir criar uma maquete para mostrar como será a ponte que ela foi contratada para construir. Em seguida, ela pode criar um modelo matemático e físico da ponte e usá-lo para simular e provar propriedades da mesma, tais como carga máxima, resistência a ventos, ondas, terremotos, etc.

Infelizmente, modelos de software — pelo menos até hoje — são menos efetivos do que os modelos matemáticos e físicos usados em outras Engenharias. O motivo é que ao abstrair detalhes eles também descartam parte da complexidade que é essencial aos sistemas modelados. Frederick Brooks comenta sobre essa questão em seu ensaio "*Não Existe Bala de Prata ...*" ([link](#)):

“A complexidade de um software é uma propriedade essencial e não acidental. Portanto, representações de uma entidade de software que abstraem sua complexidade normalmente também abstraem sua essência. Por três séculos, matemáticos e físicos obtiveram grandes avanços construindo modelos simplificados de um fenômeno complexo, derivando propriedades de tais modelos e verificando tais propriedade por meio de experimentos. Esse paradigma funcionou porque as complexidades ignoradas não são propriedades essenciais do fenômeno sob estudo. Porém, essa abordagem não funciona quando as complexidades são essenciais.”

A frase que abre esse capítulo, do estatístico britânico George Box, também remete a uma reflexão sobre o uso prático de modelos. Apesar de a frase se referir a modelos matemáticos, ela se aplica a outros modelos, inclusive modelos de software. Segundo Box, todos os modelos são “errados”, pois são simplificações ou aproximações da realidade. Por isso, a questão principal consiste em avaliar se, apesar dessas simplificações, um modelo continua sendo uma abstração útil para o estudo de alguma propriedade do objeto ou fenômeno que ele modela.

Nesta introdução, estamos procurando calibrar as expectativas associadas ao estudo de modelos de software. Por um lado, como afirmamos, eles não têm a mesma efetividade de modelos em outras Engenharias. Além disso, via de regra, modelos de software não são formalismos matemáticos, mas sim representações gráficas de determinadas dimensões de um sistema de software. Por outro lado, isso não implica em dizer que modelos de software são inúteis, a ponto de não merecer um capítulo em um livro sobre práticas de Engenharia de Software Moderna. Se

não criarmos expectativas irrealistas, eles podem ter um papel importante no desenvolvimento de sistemas de software, tal como veremos na próxima seção.

Se pensarmos em termos de atividades de desenvolvimento de software, a criação de modelos é considerada uma atividade de projeto (*design*). Durante o levantamento de requisitos, as atenções estão voltadas para a definição do problema que será resolvido pelo sistema. Quando se avança para atividades de design, o problema já deve estar devidamente entendido e as atenções se voltam para a concepção de uma solução capaz de resolvê-lo. Após essa solução ser projetada, ela deve ser implementada, usando-se linguagens de programação, bibliotecas, frameworks, bancos de dados, etc.

Especificamente, neste capítulo, iremos estudar um subconjunto dos diagramas propostos pela UML (*Unified Modelling Language*). Vamos começar descrevendo a história e o contexto que levou à criação da UML. Em seguida, vamos estudar alguns dos principais diagramas UML com um pouco mais de detalhe.

Aprofundamento: Desde a década de 70, pesquisadores têm investigado o uso de modelos matemáticos em Engenharia de Software, por meio do que se chama de **Métodos Formais**. Esses métodos valem-se de uma notação matemática — baseada em lógica, teoria de conjuntos ou Redes de Petri, por exemplo — para derivar **especificações formais** para sistemas de software. Além de serem precisas e não-ambíguas, especificações formais podem ser usadas para provar propriedades de um sistema mesmo antes de sua implementação. Por exemplo, em tese, poderia-se provar que um sistema concorrente não possui deadlocks ou condições de corrida. Pode parecer ambicioso, mas isso ocorre em outras Engenharias. Retomando o exemplo do início da seção, engenheiros civis usam há séculos modelos matemáticos para provar, por exemplo, que uma ponte — antes de ser construída — vai suportar determinada carga e certas condições climáticas. No entanto, o uso de formalismos e especificações matemáticas em Engenharia de Software não avançou como em outras Engenharias. Por isso, eles são pouco usados atualmente, com exceção talvez de alguns sistemas de missão crítica.

3.2 UML

UML é uma notação gráfica para modelagem de software. A linguagem define um conjunto de diagramas para documentar e ajudar no design de sistemas de software, particularmente sistemas orientados a objetos. As origens de UML datam da

década de 80, quando o paradigma de orientação a objetos estava amadurecendo e vivendo seu auge. Assim, surgiram diversas linguagens orientadas a objetos, como C++, e também algumas notações gráficas para modelagem de software. Lembre-se que os sistemas na década de 80 eram desenvolvidos segundo o Modelo Waterfall, que prescreve uma grande e longa fase de design. A proposta de UML era que nessa fase seriam criados modelos gráficos, que depois seriam repassados para os programadores, para serem convertidos em código fonte.

Na verdade, UML é o resultado de um esforço para unificar as notações gráficas que surgiram no final das décadas de 80 e início da década de 90. Especificamente, a primeira versão de UML foi proposta em 1995, como resultado da unificação de notações que estavam sendo desenvolvidas de forma independente por três Engenheiros de Software conhecidos na época: Grady Booch, Jim Rumbaugh e Ivar Jacobson. Nessa época, surgiram também ferramentas para desenhar diagramas UML, as quais foram chamadas de **ferramentas CASE** (*Computer-Aided Software Engineering*). O nome é inspirado em ferramentas CAD (*Computer Aided Design*), usadas para criar modelos para produtos de Engenharia tradicional, como casas, pontes, automóveis, aviões, etc. Por isso, era importante ter uma padronização de UML, de forma que um diagrama criado em uma ferramenta CASE pudesse ser aberto e editado em uma outra ferramenta, de uma empresa diferente. De fato, em 1997, UML passou a ser um padrão gerenciado pela OMG, que é uma organização de padronização financiada por indústrias de software. Desde o início, o desenvolvimento de UML foi comandado por consultores influentes e por grandes empresas de ferramentas ou consultoria, como a Rational, que depois viria a ser comprada pela IBM.

Como usar UML?

Martin Fowler, em seu livro sobre UML ([link](#)), propõe uma classificação sobre formas de uso dessa linguagem de modelagem. Segundo ele, existem três formas principais de uso de UML: como blueprint, como linguagem de programação ou como esboço. Vamos descrever cada uma delas nos próximos parágrafos.

UML como blueprint corresponde ao uso de UML vislumbrado por seus criadores, ainda na década de 90. Nessa forma de uso, defende-se que, após o levantamento de requisitos, seja produzido um conjunto de modelos — ou plantas técnicas (*blueprints*) — documentando diversos aspectos de um sistema e sempre usando diagramas UML. Esses modelos seriam criados por analistas de sistemas, usando-se ferramentas CASE e, depois, repassados a programadores, para codificação. Logo, UML como *blueprint* é recomendado quando se emprega processos de desenvolvimento do tipo Waterfall ou quando se adota o Processo Unificado

(UP). Na verdade, UP foi proposto por pessoas com forte ligação com UML. No entanto, como já discutimos no Capítulo 2, o uso de UML na construção de modelos detalhados e completos é cada vez mais raro. Por exemplo, com métodos ágeis não existe uma longa fase inicial de design (*big design up front*). Em vez disso, decisões de design são tomadas e refinadas ao longo do desenvolvimento, em cada uma das iterações (ou *sprints*). Por isso, não iremos neste capítulo nos aprofundar no uso de UML como *blueprint*.

UML como linguagem de programação corresponde ao uso de UML vislumbrado pela OMG, após a padronização da linguagem de modelagem. De forma ambiciosa e pelo menos durante um período, vislumbrou-se a geração de código automaticamente a partir de modelos UML. Em outras palavras, não haveria mais uma fase de codificação, pois o código seria gerado diretamente a partir da “compilação” de modelos UML. Essa forma de uso é conhecida como **Desenvolvimento Dirigido por Modelos** (*Model Driven Development ou MDD*). Para que MDD fosse viável, UML foi expandida e ganhou novos recursos e diagramas. Foi a partir desse momento que a linguagem ganhou a reputação de ser pesada e complexa. Porém, mesmo com adição de complexidade extra, o uso de UML para geração de código não se tornou comum, pelo menos na grande maioria dos sistemas.

Resta então o terceiro uso, **UML como esboço**, que corresponde à forma que vamos estudar neste capítulo. Nela, usamos UML para construir diagramas leves e informais de partes de um sistema, vindo daí o nome esboço (*sketch*). Esses diagramas são usados para comunicação entre os desenvolvedores, em duas situações principais:

- **Engenharia Avante** (*Forward Engineering*): quando os desenvolvedores usam modelos UML para discutir e analisar alternativas de design, antes que exista qualquer código. Por exemplo, suponha que uma história tenha sido alocada para o sprint corrente. Antes de implementar a história, os desenvolvedores podem se reunir e fazer um esboço das principais classes que deverão ser criadas no sistema, bem como dos relacionamentos entre elas. O objetivo é validar a proposta de tais classes antes de começar a codificar.
- **Engenharia Reversa** (*Reverse Engineering*): quando os desenvolvedores usam modelos UML para analisar e discutir uma funcionalidade que já se encontra implementada no código fonte. Por exemplo, um desenvolvedor mais experiente pode desenhar alguns diagramas UML para explicar para um desenvolvedor recém-contratado como uma funcionalidade está implementada. Normalmente, é mais fácil conduzir essa explicação usando modelos e diagramas gráficos, do que analisar e explicar cada linha de código. Ou

seja, aplica-se aqui o ditado segundo o qual “uma figura vale mais do que mil palavras”.

Nas duas situações, o objetivo não é gerar modelos completos e detalhados. Por isso, não se considera o uso de ferramentas complexas e caras, como ferramentas CASE. Muito menos se cogita a geração automática de código a partir desses esboços. Muitas vezes, os diagramas são desenhados em um quadro e, depois, fotografados e apagados. Adicionalmente, usa-se apenas um subconjunto dos diagramas UML.

Como os esboços são pequenos e informais, pode-se questionar a necessidade de uma linguagem padronizada nos cenários que mencionamos. No entanto, consideramos que é melhor usar uma notação existente há anos, mesmo que de forma parcial, do que inventar uma notação própria. Especificamente, o emprego de UML como esboço contribui para evitar dois extremos. Por um lado, ele não assume o emprego rígido, detalhado e sistemático de UML. Por outro lado, evita-se o uso de uma notação informal e *ad hoc*, cuja semântica pode não ser clara para todos os desenvolvedores. Além disso, UML costuma ser usada em livros, tutoriais e documentos que explicam o uso de frameworks ou técnicas de programação. Por exemplo, no Capítulo 6, usaremos diagramas UML para ilustrar o funcionamento de alguns padrões de projeto. Caso o leitor não tenha tido contato com UML, pode ser que ele tenha dificuldade para entender o conceito que está sendo explicado.

Sintetizando a descrição que acabamos de fazer, modelos de software, como diagramas UML, são usados para comunicação entre desenvolvedores. Ou seja, eles são escritos por e para desenvolvedores. Trata-se de uma diferença importante para documentos de requisitos, que, conforme vimos no capítulo anterior, são escritos por desenvolvedores, mas de forma que eles possam ser lidos e verificados pelos usuários finais do sistema.

Mundo Real: No segundo semestre de 2013, Sebastian Baltes e Stephan Diehl — ambos pesquisadores da Universidade de Trier, na Alemanha — pediram 394 desenvolvedores para responder um questionário sobre o emprego de esboços (*sketches*) em atividades de projeto de software ([link](#)). Esses desenvolvedores estavam distribuídos por mais de 32 países, embora a maioria fosse da Alemanha (54%). A análise das respostas obtidas revelou resultados interessantes sobre o uso de esboços em atividades de desenvolvimento de software, conforme descrito a seguir:

- 24% dos desenvolvedores que participaram da pesquisa criaram o último

esboço no mesmo dia em que responderam ao questionário e 39% no intervalo de tempo máximo de uma semana, antes da resposta. Portanto, esses percentuais indicam que esboços são criados com freqüência por desenvolvedores de software.

- 58% dos últimos esboços criados pelos participantes foram depois arquivados, seja em papel (6%), digitalmente (42%) ou de ambas as formas (10%). Isso sugere que os desenvolvedores consideram que os esboços carregam informação importante, que talvez seja útil no futuro.
- 40% dos esboços foram feitos em papel, 18% em quadros e 39% em computadores.
- 52% dos esboços foram feitos para ajudar no projeto (*design*) da arquitetura do sistema, 48% para ajudar no projeto de novas funcionalidades, 46% para explicar alguma tarefa para um outro desenvolvedor, 45% para analisar requisitos e 44% para ajudar no entendimento de uma tarefa. A soma dos percentuais ultrapassa 100% porque os participantes podiam marcar mais de uma resposta.
- 48% dos esboços continham algum elemento de UML e 9% eram integralmente baseados em diagramas UML. Portanto, esses percentuais reforçam a importância de estudar UML, não como notação para documentação detalhada de sistemas (*blueprints*), mas para ajudar na construção de modelos informais e parciais.

Diagramas UML

Os diagramas UML são classificados em dois grandes grupos:

- **Diagramas Estáticos (ou Estruturais)** modelam a estrutura e organização de um sistema, incluindo informações sobre classes, atributos, métodos, pacotes, etc. Neste capítulo, vamos estudar dois diagramas estáticos: Diagramas de Classes e Diagramas de Pacotes.
- **Diagramas Dinâmicos (ou Comportamentais)** modelam eventos que ocorrem durante a execução de um sistema. Por exemplo, eles podem modelar uma sequência de chamadas de métodos. Neste capítulo, vamos estudar dois diagramas dinâmicos: Diagramas de Seqüência e Diagramas de Atividades.

Para entender melhor a diferença entre esses grupos de diagramas, diagramas estáticos lidam apenas com informações que estão disponíveis, por exemplo, quando da compilação do código resultante dos modelos. Essa visão é estática porque ela

não muda, a não ser que sejam realizadas mudanças nos modelos. Já os diagramas dinâmicos fornecem uma visão de tempo de execução. Eles são dinâmicos porque é comum ter execuções diferentes de um mesmo programa. Por exemplo, os usuários podem executar o programa com entradas diferentes, selecionar opções e menus diferentes, etc. Em resumo, se estiver interessado em modelar a estrutura de um programa, você deve usar diagramas estáticos. Se seu interesse for modelar o comportamento de um programa — isto é, o que pode acontecer durante sua execução, quais métodos são de fato executados, etc — você deve usar algum diagrama dinâmico da UML.

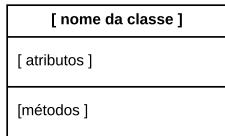
Por fim, gostaríamos de lembrar que já tratamos de Diagramas de Casos de Uso no Capítulo 3, quando apresentamos técnicas para especificação de requisitos.

Aviso: Existem diversas versões de UML. No restante deste capítulo vamos usar a **versão de UML que é adotada na 3a edição do livro UML Distilled, de Martin Fowler** ([link](#)). Esse livro foi o primeiro trabalho a discutir o uso de UML como esboço (*sketches*). Na verdade, vamos estudar um pequeno subconjunto da versão 2.0. Além de tratar de apenas quatro diagramas, não vamos cobrir todos os recursos de cada um deles. O nosso desafio ao escrever este capítulo foi selecionar os 20% (ou menos) dos recursos de UML que são responsáveis por 80% (ou mais) de seu uso prático nos dias de hoje. Para se ter uma ideia do nível de detalhe alcançado por UML, a especificação da versão mais recente da linguagem — versão 2.5.1, quando da escrita deste capítulo — possui 796 páginas. Ela pode ser encontrada no site da OMG ([link](#)).

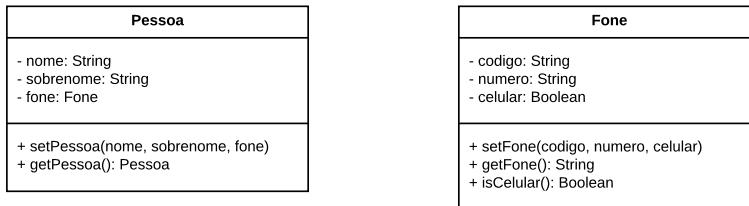
3.3 Diagramas de Classes

Diagramas de classes são os diagramas mais usados da UML. Eles oferecem uma representação gráfica para um conjunto de classes, provendo informações sobre atributos, métodos e relacionamentos que existem entre as classes modeladas.

Um diagrama de classes é desenhado usando-se retângulos e setas. Cada uma das classes é representada por meio de um retângulo com três compartimentos, conforme mostra a figura a seguir. Esses compartimentos contém o nome da classe (normalmente, em negrito), seus atributos e métodos.



Mostra-se a seguir um diagrama com duas classes: `Pessoa` e `Fone`.

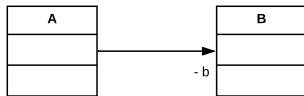


Nesse diagrama, pode-se conferir que a classe `Pessoa` tem três atributos — `nome`, `sobrenome` e `fone` — e dois métodos — `setPessoa` e `getPessoa`. Os três atributos são privados, conforme indicado pelo sinal “-” antes de cada um. Informa-se também o tipo de cada atributo. Por sua vez, os dois métodos são públicos, conforme indicado pelo sinal “+”. O diagrama possui ainda uma segunda classe, chamada `Fone`, com três atributos privados — `codigo`, `numero` e `celular` — e três métodos públicos — `setFone`, `getFone` e `isCelular`. No caso dos métodos, informamos também o nome de seus parâmetros e o tipo de retorno.

Porém, se fosse somente isso, os diagramas dariam a impressão de que as classes de um sistema são “ilhas” sem comunicação entre si. No entanto, um dos principais objetivos de diagramas de classe é mostrar visualmente os relacionamentos que existem entre as classes de um sistema. Por isso, eles incluem também linhas e setas, as quais são usadas para representar três tipos de relacionamentos: **associação**, **herança** e **dependência**. Vamos tratar de cada um deles nos próximos parágrafos.

3.3.1 Associações

Quando uma classe A possui um atributo b de um tipo B, dizemos que existe uma associação de A para B, a qual é representada por meio de uma seta, também de A para B. Na extremidade da seta, informa-se o nome do atributo de A responsável pela associação — no nosso caso, b. Veja o exemplo abaixo (nele, só mostramos as informações que nos interessam; por isso, o compartimento de atributos e métodos está vazio):



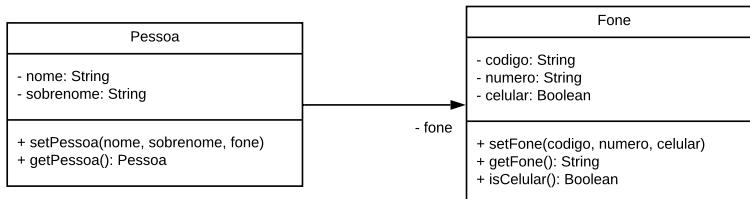
Para ficar ainda mais claro, vamos mostrar como seria o código das classes A e B:

```

class A {
    ...
    private B b;
    ...
}

class B {
    ...
}
  
```

Portanto, usando associações, podemos transformar o primeiro diagrama que mostramos nesta seção, com as classes `Pessoa` e `Fone`, no seguinte diagrama:

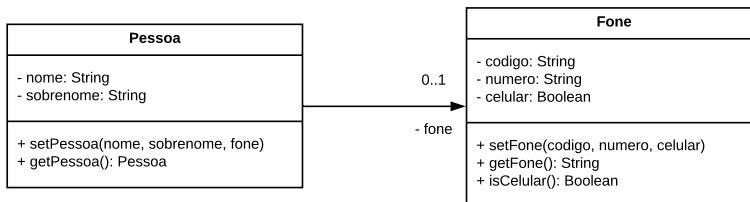


As duas versões do diagrama são semanticamente idênticas. A diferença é que na primeira versão as classes aparecem “ilhadas”. Já na segunda versão, mostrada acima, fica visualmente claro que existe uma associação de `Pessoa` para `Fone`. Reforçando, em ambos diagramas, `Pessoa` tem um atributo `fone` do tipo `Fone`. Porém, na primeira versão, esse atributo é mostrado dentro do compartimento de atributos da classe `Pessoa`. Já na segunda versão, ele é apresentado “fora” desse compartimento. Mais especificamente, na extremidade da seta que liga `Pessoa` a `Fone`. O objetivo é deixar claro que o atributo pertence a `Pessoa`, mas ele “aponta” para um objeto do tipo `Fone`.

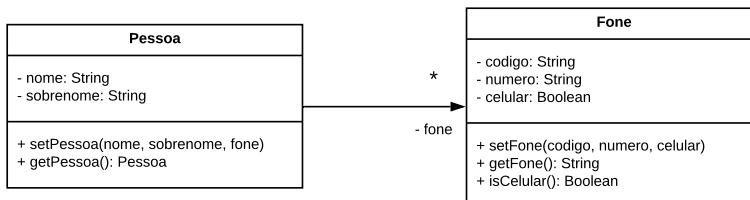
Frequentemente, associações incluem informações de **multiplicidade**, que indicam quantos objetos podem estar associados ao atributo responsável pela associação.

As informações de multiplicidade mais comuns são as seguintes: 1 (exatamente um objeto), 0..1 (zero ou um objeto) e * (zero ou mais objetos).

No próximo exemplo, incluímos informação sobre a multiplicidade da associação entre **Pessoa** e **Fone**, que no caso definimos como sendo 0..1. Essa informação consta acima do nome do atributo responsável pela associação, no caso, fone. E ela explica que uma **Pessoa** pode ter zero ou um único telefone. Usando termos de programação, o atributo fone de **Pessoa** pode ter o valor null, isto é, a **Pessoa** em questão não tem **Fone** associado. Ou então ela pode se associar a um único objeto do tipo **Fone**.



No próximo exemplo, a semântica já é diferente. Nesse caso, uma **Pessoa** pode estar associada a múltiplos objetos do tipo **Fone**, inclusive nenhum. Essa multiplicidade é representada pelo * que adicionamos logo acima da seta da associação.



Neste segundo exemplo com informações de multiplicidade, o tipo do atributo `fone` deve ser um vetor de **Fone**. Para que fique claro, mostramos o código das classes:

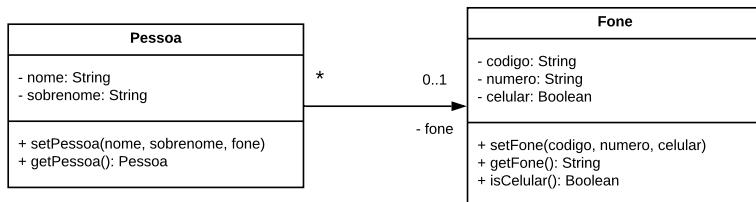
```

class Pessoa {
    private Fone[] fone;
    ...
}

class Fone {
    ...
}
  
```

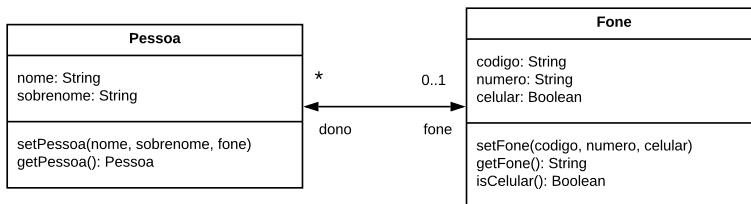
Talvez o leitor possa se perguntar o que é mais correto: uma **Pessoa** ter no máximo um **Fone** (isto é, 0..1) ou uma **Pessoa** ter qualquer quantidade de **Fone** (isto é, *)? A resposta é simples: depende dos requisitos do sistema. Ou seja, quem deve responder a essa pergunta são os usuários do sistema que estamos modelando. Para nós, o que importa é que Diagramas de Classe são capazes de modelar qualquer um dos dois cenários.

Em alguns casos, informações de multiplicidade são também mostradas na extremidade contrária da seta, como no exemplo abaixo:



Nesse diagrama, a multiplicidade 0..1 — da extremidade com a seta — indica que uma **Pessoa** pode ter zero ou um único **Fone**. Mas o mais importante é explicar a multiplicidade que foi adicionada na extremidade oposta da seta, isto é, a multiplicidade *. Ela indica que um **Fone** pode estar associado a mais de uma **Pessoa**. Em outras palavras, duas pessoas, distintas, podem compartilhar o mesmo objeto do tipo **Fone**. No entanto, a associação continua sendo unidirecional, isto é, **Pessoa** tem um atributo **fone** que representa o seu **Fone**. Porém, **Fone** não possui um atributo para armazenar as diversas pessoas a que ele pode estar associado. Tentando ser mais claro, dada uma **Pessoa** pode-se recuperar o seu **Fone**. Para isso, basta acessar o atributo **fone**. Mas dado um **Fone** não é possível saber, pelo menos via atributos, a quais objetos do tipo **Pessoa** ele está associado.

Para concluir, suponha que seja importante “navegar” nos dois sentidos da associação, isto é, de **Pessoa** para **Fone** e também de **Fone** para **Pessoa**. A solução para essa exigência é simples: basta tornar a **associação bidirecional**, isto é, adicionar uma seta em cada extremidade da linha que conecta as classes, como mostrado no próximo diagrama.



Para não deixar dúvidas sobre a semântica de uma associação bidirecional, mostramos também o código das duas classes:

```

class Pessoa {
    ...
    private Fone fone;
    ...
}

class Fone {
    ...
    private Pessoa[] dono;
    ...
}
  
```

Nesse código, `Pessoa` possui um atributo privado `fone` do tipo `Fone`, que pode ser null; com isso, satisfazemos a extremidade 0..1 da associação bidirecional. Por outro lado, `Fone` possui um vetor privado, de nome `dono`, que vai referenciar objetos do tipo `Pessoa`; assim, satisfazemos a extremidade * da mesma associação.

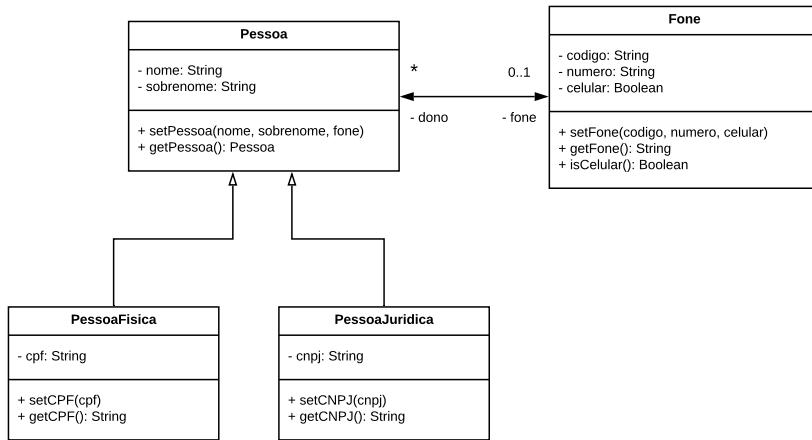
No último diagrama de classes, omitimos todos os símbolos de visibilidade, tanto pública (+) como privada (-). Isso foi feito, de forma deliberada, para destacar que estamos tratando do uso de UML para criação de esboços, quando os diagramas são criados para discutir e ilustrar uma ideia de design. Logo, nesse contexto, não faz sentido exigir que os diagramas sejam sintaticamente perfeitos. Por isso, pequenos erros ou omissões são tolerados, principalmente quando não há prejuízo para o propósito que levou à criação do diagrama.

Aprofundamento: UML — dependendo da versão que está sendo usada — admite notações diferentes para associações. Por exemplo, algumas vezes, informa-se um nome para a associação, o qual é mostrado logo acima e ao longo da seta que une as duas classes. Outras vezes, no caso de associações

bidirecionais, as duas setas são omitidas — pois a padronização de UML define o seguinte: “uma associação em que nenhuma das extremidades é marcada com uma seta de navegabilidade é navegável em ambas as direções”. No entanto, essas notações alternativas tendem a ser confusas ou mesmo ambíguas. Por exemplo, Gonzalo Génova e mais dois pesquisadores da Universidade de Madrid, na Espanha, fazem a seguinte observação sobre o uso de associações bidirecionais sem setas: “infelizmente, isso pode introduzir ambiguidade na notação gráfica, porque não conseguimos mais distinguir entre associações bidirecionais e associações sem especificação de navegabilidade” em uma de suas extremidades. ([link](#), Seção 3, quarto parágrafo). Existem ainda dois conceitos frequentemente mencionados quando estudamos associações em UML: composição e agregação. Composição é uma relação na qual a classe de destino *não* pode existir de forma independente da classe de origem. Por outro lado, quando as duas classes têm ciclos de vida independentes, temos uma relação de agregação. No entanto, na prática, esses conceitos também geram confusão e, por isso, resolvemos não incluí-los na explicação sobre diagramas de classes. A mesma opinião é compartilhada por outros autores. Por exemplo, Fowler afirma que “agregação é algo estritamente sem sentido; portanto, eu recomendo que você ignore esse conceito em seus diagramas” ([link](#), página 68).

3.3.2 Herança

Em diagramas de classes, relações de herança são representadas por meio de setas com a extremidade não preenchida. Essas setas são usadas para conectar subclasses à sua classe base. No próximo exemplo, elas indicam que `PessoaFisica` e `PessoaJuridica` são subclasses de `Pessoa`. Como usual em orientação a objetos, subclasses herdiam todos os atributos e métodos da classe base, mas também podem adicionar novos membros. Por exemplo, apenas `PessoaFisica` tem `cpf` e apenas `PessoaJuridica` tem `cnpj`.



Dependências

Existe uma dependência de uma classe A para uma classe B, representada por uma seta com uma linha tracejada de A para B, quando a classe A usa a classe B, porém esse uso não ocorre por meio de associação (isto é, A não tem um atributo do tipo B) ou herança (isto é, A não é uma subclasse de B). Dependências ocorrem, por exemplo, quando um método de A declara um parâmetro ou variável local do tipo B ou quando um método de A lança uma exceção do tipo B. Uma dependência é considerada uma modalidade menos forte de relacionamento entre classes do que relacionamentos que ocorrem por meio de associação e herança.

Para ilustrar o uso de dependências em diagramas de classes, considere o seguinte trecho de código:

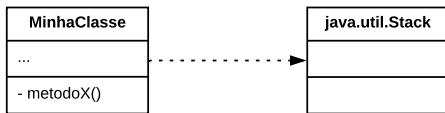
```

import java.util.Stack;

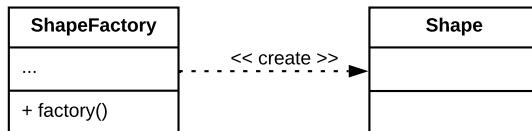
class MinhaClasse {
    ...
    private void metodoX() {
        Stack stack = new Stack();
        ...
    }
    ...
}
  
```

Observe que o `metodoX` de `MinhaClasse` possui uma variável local do tipo

`java.util.Stack`. Nesse caso, dizemos que existe uma dependência de `MinhaClasse` para `java.util.Stack`, a qual é modelada da seguinte forma:



Algumas vezes, logo acima e ao longo da seta tracejada, informa-se o tipo da dependência, usando-se palavras como `create` (para indicar que a classe de origem instancia objetos da classe de destino da dependência) ou `call` (para indicar que a classe de origem chama métodos da classe de destino). Essas palavras são escritas entre sinais de menor («) e maior (»). No diagrama a seguir, por exemplo, fica claro o tipo de dependência que `ShapeFactory` estabelece com a classe `Shape`.



Uma classe pode ter dependências para um grande número de classes. No entanto, não se costuma representar todas elas em diagramas de classes, mas apenas as mais importantes e que estão diretamente relacionadas com a funcionalidade ou propriedade do sistema que pretendemos esboçar.

3.4 Diagramas de Pacotes

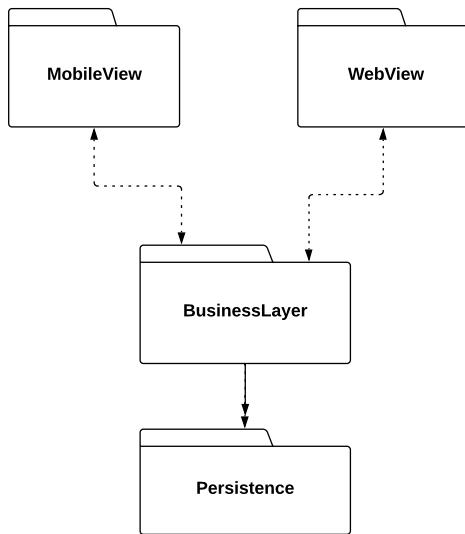
Diagrama de pacotes são recomendados quando se pretende oferecer um modelo de mais alto nível de um sistema, que mostre apenas grupos de classes — isto é, pacotes — e as dependências entre eles. Para isso, UML define um retângulo especial para representar pacotes, mostrado abaixo:



Ao contrário dos retângulos de classes, o retângulo de pacotes inclui apenas o

nome do pacote (em negrito). Ele possui ainda um detalhe na parte de cima, na forma de um trapézio, para melhor diferenciá-lo dos retângulos de classe.

A próxima figura mostra um exemplo de diagrama de pacotes:



Nesse diagrama, podemos ver que o sistema possui quatro pacotes principais: **MobileView**, **WebView**, **BusinessLayer** e **Persistence**. Podemos ver ainda as dependências — setas tracejadas — que existem entre eles. Ambos os pacotes **View** usam classes do **BusinessLayer**. Por outro lado, as classes do **BusinessLayer** também usam classes da **View**, por exemplo, para notificá-las da ocorrência de algum evento. Por isso, as setas que ligam os pacotes de **View** a **BusinessLayer** são bidirecionais. Por fim, apenas classes do pacote **BusinessLayer** usam classes do pacote **Persistence**.

Para concluir, gostaríamos de acrescentar duas observações:

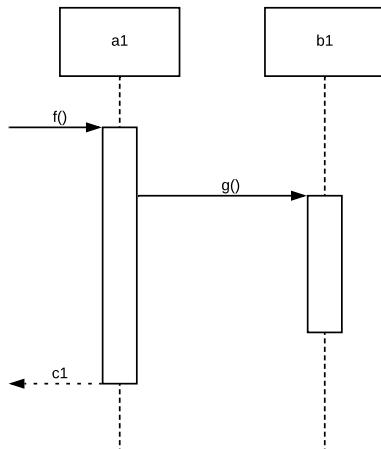
- Dependências não incluem informações sobre quantas classes do pacote de origem dependem de classes do pacote de destino. Por exemplo, suponha dois pacotes P1 e P2, ambos com 100 classes. Suponha ainda que uma única classe de P1 use uma única classe de P2. Mesmo nesse caso, dizemos que existe uma dependência de P1 para P2.
- Em diagramas de pacotes, temos um único tipo de seta, sempre tracejada, que representa qualquer tipo de relacionamento, seja ele por meio de associação, herança ou dependência simples. Essa semântica é diferente daquela que apresentamos para setas tracejadas em diagramas de classes. Nesses

últimos, relações de associação e herança são representadas por meio de setas contínuas. Apenas as demais dependências são representadas por meio de setas tracejadas.

3.5 Diagramas de Sequência

Diagramas de sequência são diagramas dinâmicos, também chamados de comportamentais. Por isso, em vez de classes, eles modelam objetos de um sistema. Adicionalmente, eles incluem informações sobre quais métodos desses objetos são executados em um determinado cenário de uso de um programa. Logo, eles são usados quando se pretende explicar o comportamento de um sistema, em um determinado cenário. Por exemplo, no final desta seção, vamos apresentar um diagrama de sequência que ilustra os métodos que são chamados quando um cliente chega em um caixa eletrônico e solicita uma operação de retirada de valores.

Antes disso, para iniciar a apresentação de diagramas de sequência, vamos usar o seguinte diagrama:



Apesar de simples, esse diagrama serve para mostrar a dinâmica e a notação usada por diagramas de sequência. Como já dissemos, diagramas de sequência modelam objetos, os quais são representados por meio de retângulos, com o nome dos objetos modelados. Esses retângulos ficam dispostos logo na primeira linha do diagrama. Portanto, dois objetos são representados no diagrama anterior, de nomes **a1** e **b1**. Abaixo de cada objeto, desenha-se uma linha vertical, a qual pode assumir duas formas: (1) quando ela é desenhada de forma tracejada, o objeto está inativo, isto é, nenhum de seus métodos está sendo executado; (2) quando a

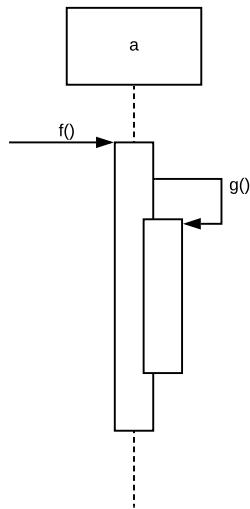
linha fica cheia, ganhando um formato retangular, um dos métodos do objeto foi chamado e encontra-se em execução. Quando essa execução termina, a linha volta a ficar tracejada. Além disso, o início da chamada é indicado por uma seta na horizontal, com o nome do método chamado. O retorno da chamada é indicado por uma seta tracejada, com o nome do objeto retornado. No entanto, às vezes a seta de retorno é omitida, como no caso da chamada do método `g`. Existem dois motivos para essa omissão: (1) o tipo de retorno é `void`; ou (2) o objeto de retorno não é relevante, a ponto de merecer ser representado no diagrama.

No diagrama de sequência mostrado acima representamos apenas dois objetos (`a1` e `b1`). Mas um diagrama de sequência pode ter mais objetos. No entanto, esse número não pode crescer tanto, pois o diagrama acaba ficando complexo e de difícil entendimento. Por exemplo, pode não ser possível representá-lo em uma única folha de papel ou em uma tela de computador.

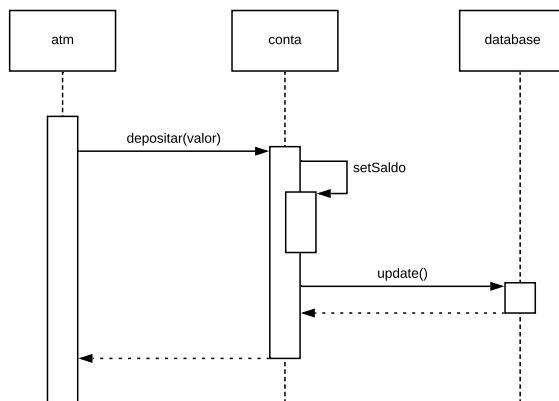
Um objeto pode ficar ativo e inativo diversas vezes em um mesmo diagrama. Ou seja, ele pode executar um método; ficar inativo; executar um novo método; ficar inativo, etc. Existe ainda um caso especial, quando um objeto chama um método dele mesmo, isto é, quando ele chama um método usando `this`. Para ilustrar esse caso, suponha o seguinte programa.

```
class A {  
    void g() {  
        ...  
    }  
  
    void f() {  
        ...  
        g();  
        ...  
    }  
  
    main() {  
        A a = new A();  
        a.f();  
    }  
}
```

A execução desse programa é representada pelo diagrama de sequência a seguir. Observe como a chamada de `g()` feita por `f()` é representada por meio de um novo retângulo, que “sai” do retângulo que representa a ativação de `f()`.



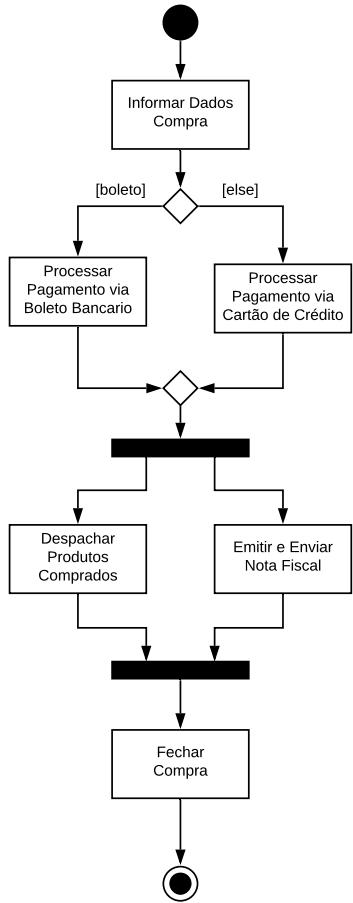
Para concluir, o próximo diagrama mostra um cenário mais real, que ilustra os métodos chamados quando o cliente de uma caixa eletrônico solicita um depósito de certo valor em sua conta.



3.5.1 Diagramas de Atividades

Diagramas de atividades são usados para representar, em alto nível, um processo ou fluxo de execução. Os principais elementos desses diagramas são **ações** representadas por quadrados. Existem ainda elementos de **controle**, que definem a ordem de execução das ações. A próxima figura mostra um diagrama de atividades

que modela o processo seguido após um usuário fechar uma compra em um loja virtual. Para isso, assume-se que os produtos comprados já estão no carrinho de compra.

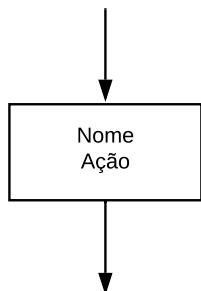


Para entender o funcionamento de um diagrama de atividades (como aquele mostrado acima), devemos assumir que existe uma ficha (*token*) imaginária que caminha pelos nodos do diagrama. A seguir, explicamos o comportamento de cada nodo de um diagrama de atividades, assumindo a existência dessa ficha.

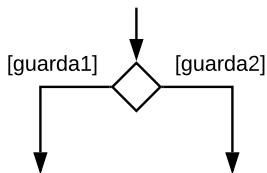
Nodo Inicial: Cria uma ficha para dar início à execução do processo. Feito isso, repassa a ficha para seu único fluxo de saída. Por definição, o nodo inicial não possui fluxo de entrada.



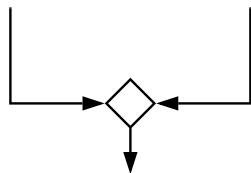
Ações: Possuem um único fluxo de entrada e um único fluxo de saída. Para um ação ser executada uma ficha precisa chegar no seu fluxo de entrada. Após a execução, repassa-se a ficha para o fluxo de saída.



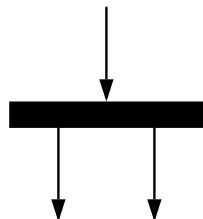
Decisões: Possuem um único fluxo de entrada e dois ou mais fluxos de saída. Cada fluxo de saída possui uma variável booleana associada, chamada de guarda. Para se tomar uma decisão, precisa-se receber uma ficha no fluxo de entrada. Quando isso acontece, a ficha é repassada apenas para o fluxo de saída cuja condição é verdadeira.



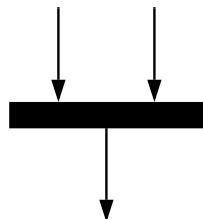
Merges: Podem possuir vários fluxos de entrada, mas um único fluxo de saída. Quando uma ficha chega em um dos fluxos de entrada, fazem seu repasse para o fluxo de saída. São usados para unir os fluxos de nodos de decisão.



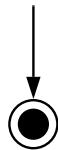
Forks: Possuem um único fluxo de entrada e um ou mais fluxos de saída. Atuam como multiplicadores de ficha: quando recebem uma ficha no fluxo de entrada, criam e repassam fichas idênticas em cada fluxo de saída. Como resultado, passam a existir múltiplos processos em execução de forma paralela.



Joins: Possuem vários fluxos de entrada, mas um único fluxo de saída. Atuam como sorvedouros de fichas: esperam que fichas cheguem em todos os fluxos de entrada. Quando isso acontece, repassam uma única ficha para o fluxo de saída. Logo, são usados para sincronizar processos. Em outras palavras, transformar vários fluxos de execução em um único fluxo.



Nodo Final: Pode possuir mais de um fluxo de entrada; mas não possui fluxos de saída. Quando uma ficha chega em um dos fluxos de entrada, encerra-se a execução do diagrama de atividades.



Aprofundamento: Existem pelo menos três outras alternativas para modelagem de fluxos e processos:

- **Fluxogramas**, os quais foram propostos tão logo se começou a desenvolver os primeiros programas para computadores modernos. Diagramas de atividades são parecidos com fluxogramas; porém, eles incluem suporte a concorrência, por meio de *forks* e *joins*. Por outro lado, fluxogramas modelam processos sequenciais.
- **Redes de Petri** é uma notação gráfica, proposta pelo matemático alemão Carl Adam Petri, em 1962, para modelagem de sistemas concorrentes. Redes de Petri possuem uma representação gráfica e também usam fichas (*tokens*) para marcar o estado corrente do sistema. Elas têm ainda a vantagem de possuir uma definição mais formal, principalmente quando comparada com a definição de diagramas de sequência. Por outro lado, esses últimos tendem a oferecer uma notação mais simples e fácil de entender.
- **BPMN** (*Business Process Model and Notation*) é um esforço mais recente, que teve início nos anos 2000, visando a proposição de uma notação gráfica mais amigável para modelagem de processos de negócio do que aquela oferecida por diagramas de atividades. Um dos objetivos é propiciar que analistas de negócio possam ler, interpretar e validar diagramas BPMN.

Bibliografia

Martin Fowler. UML Distilled: A Brief Guide to the Standard Object Modeling Language. Addison-Wesley, 2003.

Grady Booch, James Rumbaugh, Ivar Jacobson. The Unified Modeling Language User Guide. Addison-Wesley, 2005.

Craig Larman. Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development. Prentice Hall, 2004.

Exercícios de Fixação

1. Explique e discuta os três usos possíveis de UML:
 - (a) Como blueprint (ou plantas técnicas detalhadas)
 - (b) Como sketches (esboços)
 - (c) Como linguagem de programação.
2. Descreva cenários de uso de diagramas de classes UML como instrumento de:
 - (a) Engenharia Reversa
 - (b) Engenharia Avante (*Forward Engineering*).
3. Modele os cenários descritos a seguir usando Diagramas de Classe UML. Veja que as classes são grafadas em uma fonte diferente.
 - (a) `ContaBancaria` possui exatamente um `Cliente`. Um `Cliente`, por sua vez, pode ter várias `ContaBancaria`. Existe naveabilidade em ambos os sentidos.
 - (b) `ContaPoupanca` e `ContaSalario` são subclasses de `ContaBancaria`.
 - (c) No código de `ContaBancaria` declara-se uma variável local do tipo `BancoDados`.
 - (d) Um `ItemPedido` se refere a um único `Produto` (sem naveabilidade). Um `Produto` pode ter vários `ItemPedido` (com naveabilidade).
 - (e) A classe `Aluno` possui atributos `nome`, `matricula`, `curso` (todos privados); e métodos `getCurso()` e `cancelaMatricula()`, ambos públicos.
4. (ENADE 2014, Tec. e Análise de Sistemas) Construa um diagrama de classes para representar as seguintes classes e associações:
 - Uma revista científica possui título, INSS e periodicidade;
 - Essa revista publica diversas edições com os seguintes atributos: número da edição, volume da edição e data da edição. Importante destacar que cada instância da classe edição relaciona-se única e exclusivamente a uma instância da classe revista científica, não podendo relacionar-se com nenhuma outra;
 - Um artigo possui título e nome do autor. Um artigo é um conteúdo exclusivo de uma edição. E uma edição obrigatoriamente tem que possuir no mínimo 10 e no máximo 15 artigos.

5. Crie diagramas de classes para os seguintes trechos de código:

(a)

```
public class HelloWorldSwing {
    public static void main(String[] args) {
        JFrame frame = new JFrame("Hello world!6");
        frame.setVisible(true);
    }
}
```

(b)

```
class HelloWorldSwing extends JFrame {
    public HelloWorldSwing() {
        super("Hello world!");
    }
    public static void main(String[] args) {
        HelloWorldSwing frame = new HelloWorldSwing();
        frame.setVisible(true);
    }
}
```

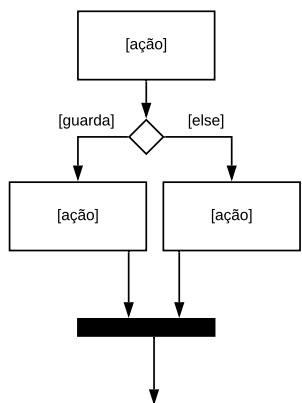
6. Mostre o diagrama de sequência relativo ao seguinte código. O diagrama deve começar com a seguinte chamada *a.m5()*.

```
A a = new A(); // variaveis globais
B b = new B();
C c = new C();

class C {
    void m1() { ... }
}
class B {
    void m2() { ... c.m1(); ... this.m3(); ... }
    void m3() { ... c.m1(); ... }
    void m4() { ... }
}
class A {
    void m5() { ... b.m2(); ... b.m3(); ... b.m4(); ... }
}
```

7. Em diagramas de atividades, explique a diferença entre um nodo de *merge* e um nodo de *join*.

8. Qual é o erro do seguinte diagrama de atividades? Refaça o diagrama de forma a refletir corretamente a intenção do projetista.



Capítulo 4

Arquitetura

Architecture is about the important stuff. Whatever that is. – Ralph Johnson

Este capítulo inicia com uma introdução ao conceito de arquitetura de software. Em seguida, discutimos diversos padrões arquiteturais, incluindo: Arquitetura em Camadas e, especificamente, Arquitetura em Três Camadas (Seção 7.2), Arquitetura MVC (Seção 7.3) e Arquitetura baseada em Microsserviços (Seção 7.4). No caso de microsserviços, procuramos mostrar o contexto que levou ao surgimento desse padrão arquitetural, bem como discutimos seus principais benefícios e desafios. Em seguida, discutimos dois padrões arquiteturais usados para garantir escalabilidade e desacoplamento em sistemas distribuídos: Filas de Mensagens (Seção 7.5) e Publish/Subscribe (Seção 7.6). Terminamos o capítulo discutindo outros padrões arquiteturais (Seção 7.7) e dando um exemplo de anti-padrão arquitetural (Seção 7.8).

4.1 Introdução

Existe mais de uma definição para arquitetura de software. Uma das mais comuns considera que arquitetura preocupa-se com “projeto em mais alto nível”. Ou seja, o foco deixa de ser a organização e interfaces de classes individuais e passa a ser em unidades de maior tamanho, sejam elas pacotes, componentes, módulos, subsistemas, camadas ou serviços — o nome não importa tanto neste primeiro momento. De forma genérica, os termos que acabamos de mencionar devem ser entendidos como conjuntos de classes relacionadas.

Além de possuírem um “maior tamanho”, os componentes arquiteturais devem ser relevantes para que um sistema atenda a seus objetivos. Por exemplo, suponha que você trabalhe em um sistema de informações. Certamente, esse sistema inclui um módulo de persistência, que faz a interface com o banco de dados. Esse módulo é fundamental em sistemas de informações, pois eles têm como objetivo principal automatizar e persistir informações relativas a processos de negócio. Por outro lado, suponha agora que você trabalhe em um sistema que usa técnicas de inteligência artificial para diagnosticar doenças. O sistema também possui um módulo de persistência para armazenar dados das doenças que são diagnosticadas. Porém, esse módulo, além de simples, não é relevante para o objetivo principal do sistema. Logo, ele não faz parte da sua arquitetura.

Existe ainda uma segunda definição para arquitetura de software. Tal como expresso na frase de Ralph Johnson que abre esse capítulo, ela considera que arquitetura de software inclui as decisões de projeto mais importantes em um sistema. Essas decisões são tão importantes que, uma vez tomadas, dificilmente poderão ser revertidas no futuro. Portanto, essa segunda forma de definir arquitetura é mais ampla do que a primeira que apresentamos. Ela considera que arquitetura não é apenas um conjunto de módulos, mas um conjunto de decisões. É verdade que dentre essas decisões, inclui-se a definição dos módulos principais de um sistema. Mas outras decisões também são contempladas, como a escolha da linguagem de programação e do banco de dados que serão usados no desenvolvimento. De fato, uma vez que um sistema é implementado com um determinado banco de dados, dificilmente consegue-se migrar para um outro banco de dados. Prova disso é que ainda hoje temos exemplos de sistemas críticos que funcionam com bancos de dados não-relacionais e que são implementados em linguagens como COBOL.

Padrões arquiteturais propõem uma organização de mais alto nível para sistemas de software, incluindo seus principais módulos e as relações entre eles. Essas relações definem, por exemplo, que um módulo A pode (ou não pode) usar os serviços de um módulo B. Neste capítulo, vamos estudar padrões arquiteturais que dão origem às seguintes arquiteturas: Arquitetura em Camadas (Seção 7.2), Arquitetura Model-View-Controller ou MVC (Seção 7.3), Microsserviços (Seção 7.4), Arquitetura Orientada a Mensagens (Seção 7.5) e Arquitetura Publish/Subscribe (Seção 7.6). Para finalizar, iremos apresentar de forma breve outros padrões arquiteturais, como pipes e filtros (Seção 7.7). Vamos também dar um exemplo de um anti-padrão arquitetural, conhecido como *big ball of mud* (Seção 7.8).

Aprofundamento: Alguns autores — como Taylor et al. ([link](#)) — fazem uma distinção entre padrões e **estilos arquiteturais**. Segundo eles, padrões focam em soluções para problemas específicos de arquitetura; enquanto estilos arquiteturais propõem que os módulos de um sistema devem ser organizados de uma determinado modo, o que não necessariamente ocorre visando resolver um problema específico. Assim, para esses autores, MVC é um padrão arquitetural que resolve o problema de separar apresentação e modelo em sistemas de interfaces gráficas. Por outro lado, Pipes & Filtros constituem um estilo arquitetural. Neste capítulo, porém, não vamos fazer essa distinção. Em vez disso, chamaremos todos eles de padrões arquiteturais.

4.1.1 Debate Tanenbaum-Torvalds

No início de 1992, um debate acalorado sobre a arquitetura de sistemas operacionais tomou conta de um grupo de discussão da Internet. Apesar de vários desenvolvedores e pesquisadores terem participado da discussão, ela ficou conhecida como **Debate Tanenbaum-Torvalds** ([link](#), apêndice A, página 102). Tanenbaum (Andrew) é um pesquisador da área de sistemas operacionais, autor de livros-texto na área e professor da Vrije Universiteit, em Amsterdã, na Holanda. E Torvalds (Linus) na época era estudante de Computação na Universidade de Helsinki, na Finlândia.

A discussão começou quando Tanenbaum postou uma mensagem no grupo com o título “Linux está obsoleto”. O seu principal argumento era que o Linux seguia uma **arquitetura monolítica**, na qual todas as funções do sistema operacional — como gerenciamento de processos, gerenciamento de memória e sistemas de arquivos, por exemplo — são implementadas em um único arquivo executável, que roda em modo supervisor. Desde essa época, Tanenbaum argumentava que a melhor solução para sistemas operacionais era uma **arquitetura microkernel**, na qual o kernel fica responsável apenas pelas funções mais básicas do sistema. As demais funções rodam como processos independentes e fora do kernel. Linus respondeu à mensagem de forma enfática, alegando que pelo menos o Linux já era uma realidade na época, enquanto que o sistema baseado em um microkernel que estava sendo desenvolvido por Tanenbaum apresentava diversos problemas e bugs. A discussão continuou forte e Tanenbaum chegou a declarar que Torvalds tinha sorte por não ter sido seu aluno; se fosse, ele certamente não teria obtido uma boa nota com o projeto monolítico do Linux. Um comentário interessante foi feito em seguida por Ken Thompson, um dos projetistas das primeiras versões do Unix:

“Na minha opinião, é mais fácil implementar um sistema operacional com um kernel monolítico. Mas é também mais fácil que ele se transforme em uma bagunça à medida que o kernel é modificado.”

Na verdade, Thompson previu o futuro, pois em 2009, Linus declarou o seguinte em uma conferência:

“Não somos mais o kernel simples, pequeno e hiper-eficiente que imaginei há 15 anos. Em vez disso, nosso kernel está ficando grande e inchado. E sempre que adicionamos novas funcionalidades, o cenário piora.”

Esse comentário consta de uma página da Wikipedia ([link](#)) e foi objeto de diversas matérias em sites de tecnologia na época. Ele revela que arquitetura não são apenas decisões importantes e difíceis de reverter. Muitas vezes, são também decisões que levam anos para que seus efeitos negativos fiquem mais claros e começem a causar problemas.

4.2 Arquitetura em Camadas

Arquitetura em camadas é um dos padrões arquiteturais mais usados, desde que os primeiros sistemas de software de maior porte foram construídos nas décadas de 60 e 70. Em sistemas que seguem esse padrão, as classes são organizadas em módulos de maior tamanho, chamados de **camadas**. As camadas são dispostas de forma hierárquica, como em um bolo. Assim, uma camada somente pode usar serviços — isto é, chamar métodos, instanciar objetos, estender classes, lançar exceções — da camada imediatamente inferior.

Dentre outras aplicações, arquiteturas em camadas são muito usadas na implementação de protocolos de rede. Por exemplo, HTTP é um protocolo de aplicação, que usa serviços de um protocolo de transporte; por exemplo, TCP. Por sua vez, TCP usa serviços de um protocolo de rede; por exemplo, IP. Finalmente, a camada IP usa serviços de um protocolo de comunicação; por exemplo, Ethernet.

Uma arquitetura em camadas partitiona a complexidade envolvida no desenvolvimento de um sistema em componentes menores (as camadas). Como uma segunda vantagem, ela disciplina as dependências entre essas camadas. Como dissemos, a camada n somente pode usar serviços da camada $n-1$. Isso ajuda no entendimento, manutenção e evolução de um sistema. Por exemplo, torna-se mais fácil trocar uma camada por outra (por exemplo, mudar de TCP para UDP). Fica mais fácil também o reúso de uma camada por mais de uma camada superior. Por exemplo,

a camada de transporte pode ser usada por vários protocolos de aplicação, como HTTP, SMTP, DHCP, etc.

Aprofundamento: Uma das primeiras propostas de arquitetura em camada foi elaborada por Edsger W. Dijkstra, em 1968, para um sistema operacional denominado THE ([link](#)). As camadas propostas por Dijkstra foram as seguintes: multiprogramação (camada 0), alocação de memória (camada 1), comunicação entre processos (camada 2), gerenciamento de entrada/saída (camada 3) e programas dos usuários (camada 4). Dijkstra conclui o artigo destacando que os benefícios de uma estrutura hierárquica são mais importantes ainda em projetos de maior porte.

4.2.1 Arquitetura em Três Camadas

Esse tipo de arquitetura é comum na construção de sistemas de informação corporativos. Até o final da década de 80, aplicações corporativas — como folhas de pagamento, controle de estoque, sistemas financeiros, etc — executavam em **mainframes**, que eram computadores fisicamente grandes e também muito caros. As aplicações eram monolíticas e acessadas por meio de “terminais burros”, isto é, sem qualquer capacidade de processamento e com uma interface totalmente textual. Com o avanço nas tecnologias de rede e de hardware, foi possível migrar esses sistemas de mainframes para outras plataformas. Foi nessa época que arquiteturas em três camadas se tornaram uma alternativa muito comum.

As três camadas dessas arquitetura são as seguintes:

- Interface com o Usuário, também chamada de camada de apresentação, é responsável por toda interação com o usuário. Ela trata tanto da exibição de informação, como da coleta e processamento de entradas e eventos de interfaces, tais como cliques em botões, marcação de texto, etc. A camada de interface pode ser uma aplicação desktop, em Windows ou outro sistema operacional com interface gráfica, como também Web. Por exemplo, um sistema acadêmico deve prover uma interface para os professores lançarem as notas de suas disciplinas. O elemento principal dessa interface pode ser um formulário com duas colunas: nome do aluno e nota. O código que implementa esse formulário deve estar na camada de interface.
- Lógica de Negócio, também conhecida como camada de aplicação, implementa as regras de negócio do sistema. No sistema acadêmico que estamos usando como exemplo, podemos ter a seguinte regra de negócio: as notas são

maiores ou iguais a zero e menores ou iguais ao valor da avaliação. Quando um professor solicitar o lançamento das notas de uma disciplina, cabe à camada de lógica verificar se essa regra é obedecida. Uma outra regra de negócio pode ser a seguinte: após o lançamento de qualquer nota, os alunos devem ser avisados por meio de um e-mail.

- Banco de Dados, que armazena os dados manipulados pelo sistema. Por exemplo, no nosso sistema acadêmico, após lançamento e validação das notas, elas são salvas em um banco de dados.

Normalmente, uma arquitetura em três camadas é uma arquitetura distribuída. Isto é, a camada de interface executa na máquina dos clientes. A camada de negócio executa em um servidor, muitas vezes chamado de servidor de aplicação. E, por fim, temos o banco de dados. A próxima figura mostra um exemplo, que assume que a interface oferecida aos clientes é uma interface gráfica.

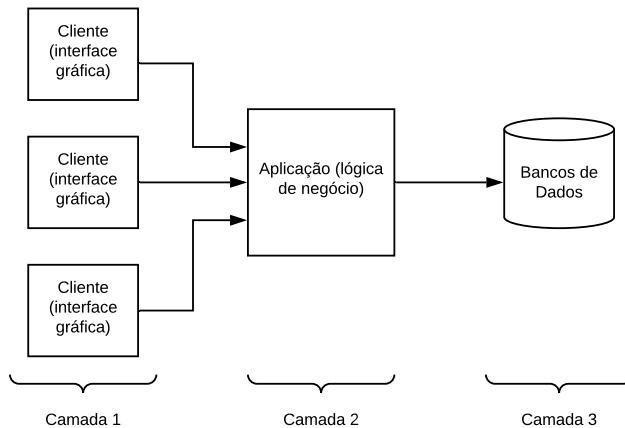


Figura 4.1: Arquitetura em três camadas

Em sistemas três camadas, a camada de aplicação pode ter diversos módulos, incluindo uma fachada, para facilitar o acesso ao sistema pelos clientes, e um módulo de persistência, com a função de isolar o banco de dados dos demais módulos.

Por fim, gostaríamos de mencionar que é possível ter sistemas em **duas camadas**. Nesses casos, as camadas de interface e de aplicação são unidas em uma única camada, que executa no cliente. A camada restante continua sendo o banco de da-

dos. A desvantagem de arquiteturas em duas camadas é que todo o processamento ocorre nos clientes, que, portanto, devem ter um maior poder de computação.

4.3 Arquitetura MVC

O padrão arquitetural MVC (Model-View-Controller) foi proposto no final da década de 70 e, em seguida, usado na implementação de Smalltalk-80, que é considerada uma das primeiras linguagens orientadas a objetos. Além de utilizarem conceitos de orientação a objetos, programas em Smalltalk foram pioneiros no uso de interfaces gráficas, com janelas, botões, scroll bars, mouse, etc. Isso em uma época em que os sistemas operacionais ofereciam apenas interfaces de linha de comando e os programas tinham uma interface textual, isto é, as telas eram uma matriz de caracteres, com, por exemplo, 25 linhas e 80 colunas.

MVC foi o padrão arquitetural escolhido pelos projetistas de Smalltalk para implementação de interfaces gráficas. Especificamente, MVC define que as classes de um sistema devem ser organizadas em três grupos:

- Visão: classes responsáveis pela apresentação da interface gráfica do sistema, incluindo janelas, botões, menus, barras de rolagem, etc.
- Controladoras: classes que tratam e interpretam eventos gerados por dispositivos de entrada, como mouse e teclado. Como resultado de tais eventos, Controladoras podem solicitar uma alteração no estado do Modelo ou da Visão. Suponha, por exemplo, uma Calculadora. Quando o usuário clica em um botão “+”, uma classe Controladora deve capturar esse evento e chamar um método do Modelo. Como um segundo exemplo, quando o usuário clicar no botão “Dark UI”, cabe também a uma classe Controladora solicitar à Visão para mudar as cores da interface gráfica para tons mais escuros.
- Modelo: classes que armazenam os dados manipulados pela aplicação e que têm a ver com o domínio do sistema em construção. Assim, classes de modelo não têm qualquer conhecimento ou dependência para classes de Visão e Controladoras. Além de dados, classes de modelo podem conter métodos que alteram o estado dos objetos de domínio.

Portanto, em uma arquitetura MVC, a interface gráfica é formada por objetos de visão e por controladores. Porém, em muitos sistemas não existe uma distinção clara entre Visão e Controladores. Segundo Fowler ([link](#), página 331), mesmo a maioria das versões de Smalltalk não separa esses dois componentes. Por isso, fica mais fácil entender da seguinte forma:

$$\text{MVC} = (\text{Visão} + \text{Controladores}) + \text{Modelo} = \text{Interface Gráfica} + \text{Modelo}$$

A próxima figura mostra as dependências entre as classes de uma arquitetura MVC. A figura primeiramente reforça que a interface gráfica é composta pela Visão e por Controladores. Podemos observar também que a Interface Gráfica pode depender do Modelo. Porém, classes de Modelo não têm dependências para classes da Interface Gráfica. Na verdade, podemos entender a Interface Gráfica como sendo observadora do Modelo. Quando o estado dos objetos do Modelo é alterado, deve-se atualizar automaticamente a interface do sistema.

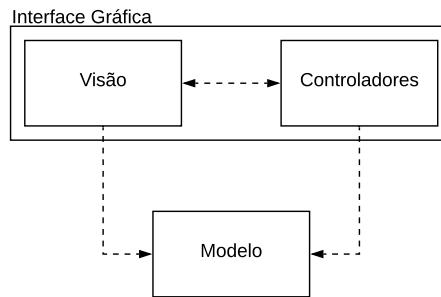


Figura 4.2: Arquitetura MVC

Dentre as vantagens de arquiteturas MVC, podemos citar:

- MVC favorece a especialização do trabalho de desenvolvimento. Por exemplo, pode-se ter desenvolvedores especialistas na implementação de interfaces gráficas, os quais são chamados de desenvolvedores de front-end. Por outro lado, desenvolvedores de classes de Modelo não precisam conhecer e implementar código de interface com usuários.
- MVC permite que classes de Modelo sejam usadas por diferentes Visões, como ilustrado na próxima figura. Neste exemplo, um objeto de Modelo armazena dois valores: hora e minutos. Esses dados são apresentados em duas visões diferentes. Na primeira, como um relógio analógico. Na segunda, como um relógio digital.
- MVC favorece testabilidade. Como veremos no próximo capítulo, é mais fácil testar objetos não-visuais, isto é, não relacionados com a implementação de interfaces gráficas. Por isso, ao separar objetos de apresentação de objetos de Modelo, fica mais fácil testar esses últimos.

Vamos concluir com um resumo sobre MVC, na visão de Fowler e Beck ([link](#), Cap. 12, pág. 370):

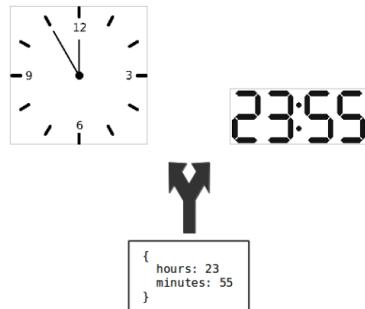


Figura 4.3: Sistema MVC com mais de uma visão (interface gráfica)

“O coração e a parte mais preciosa de MVC está na separação entre código de interface com o usuário (a Visão, também chamada de apresentação) e a lógica do domínio (o Modelo). As classes de apresentação implementam apenas a lógica necessária para lidar com a interface do usuário. Por outro lado, objetos de domínio não incluem código visual, mas apenas lógica de negócios. Isso separa duas partes complexas de sistemas de software em partes que são mais fáceis de se modificar. Também permite várias apresentações da mesma lógica de negócio.”

Pergunta Frequentes: Qual a diferença entre MVC e três camadas? A resposta vai ser um pouco longa e vamos nos basear na evolução histórica dessas arquiteturas:

- Como comentamos, MVC surgiu no final da década de 70, para ajudar na construção de interfaces gráficas. Isto é, aplicações que incluem uma interface com janelas, botões, caixas de texto, etc. Como exemplo, podemos citar um pacote de escritório, com aplicações como Word, Excel e Powerpoint, no caso do sistema operacional Windows.
- Na década de 90, as tecnologias de redes, sistemas distribuídos e bancos de dados se tornaram comuns. Viabilizou-se então a construção de aplicações distribuídas com três camadas. Nesse caso, MVC pode ser usado na implementação da camada de interface, que pode, por exemplo, ser uma aplicação nativa em Windows, implementada usando-se linguagens como Visual Basic ou Java (neste último caso, usando-se frameworks como Swing). Resumindo, a aplicação, como um todo segue, uma arquitetura em três camadas, mas usa MVC na camada de interface com o usuário.

- No início dos anos 2000, a Web se popularizou e a interface das aplicações migrou para HTML e, depois, para HTML e JavaScript. A confusão entre os termos MVC e três camadas surgiu então nessa época, principalmente devido ao aparecimento de frameworks para implementação de sistemas Web que se denominaram frameworks MVC. Como exemplo, podemos citar Spring (para Java), Ruby on Rails, Django (para Python) e CakePHP. Na verdade, esses frameworks expandiram e adaptaram o conceito de MVC para Web. Por exemplo, eles forçam a organização de um sistema Web em três partes (veja também na próxima figura): visão, composta por páginas HTML; controladores, que processam uma solicitação e geram uma nova visão como resposta e modelo, que é a camada que persiste os dados em um banco de dados.

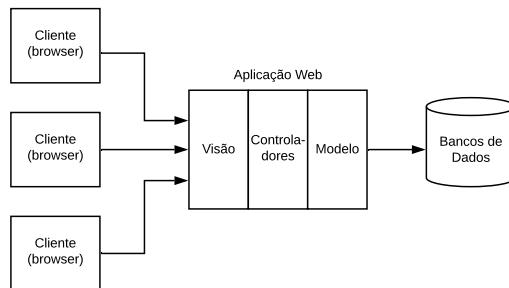


Figura 4.4: Arquitetura MVC Web

Logo, apesar de sistemas Web serem parecidos com sistemas três camadas, os frameworks Web mais populares optaram por usar termos típicos de MVC para nomear seus componentes. Portanto, a melhor maneira de responder à pergunta é afirmar que existem duas vertentes de sistemas MVC: a vertente clássica, que surgiu com Smalltalk-80 e a vertente Web, que se tornou comum na década de 90 e início dos anos 2000. Essa última vertente lembra bastante sistemas três camadas.

4.3.1 Exemplo: Single Page Applications

Em uma aplicação Web tradicional, com formulários, menus, botões, etc, toda vez que o usuário gera um evento — por exemplo, clica em um botão “Gravar” — ocorre uma interação entre o navegador e o servidor Web. Isto é, o navegador envia informações para o servidor Web, que as processa e devolve uma nova página para ser exibida para o usuário. Essas aplicações são então menos interativas e responsivas, devido ao atraso da comunicação entre navegador e servidor Web.

Recentemente, surgiu um novo tipo de sistema Web, chamado de **Single Page Applications (SPAs)**. Essas aplicações são mais parecidas com aplicações desktop do que com aplicações Web tradicionais. Ao se entrar em uma SPA, ela carrega para o navegador todo o código, incluindo páginas HTML e scripts em CSS e JavaScript. Com isso, apesar de usar um navegador, o usuário tem a impressão de que ele está usando uma aplicação local, pois não ocorre mais uma atualização da página do navegador toda vez que ele gera certos eventos. Diversas aplicações modernas são SPAs, sendo o GMail, provavelmente, o exemplo mais conhecido. Evidentemente, continua existindo uma parte da aplicação no servidor, com a qual a SPA comunica-se frequentemente. Por exemplo, quando chega um novo e-mail o GMail atualiza a lista de mensagens na caixa de entrada. Para que isso ocorra de forma automática, a comunicação entre a SPA e o servidor deve ser assíncrona.

Existem diversos frameworks — todos em JavaScript — para implementação de SPAs. A seguir, mostramos um exemplo usando Vue.js.

```
<html>
<script src="https://cdn.jsdelivr.net/npm/vue"></script>

<body>

<h3>Uma Simples SPA</h3>

<div id="ui">
  Temperatura: {{ temperatura }}
  <p><button v-on:click="incTemperatura">Incrementa
  </button></p>
</div>

<script>
var model = new Vue({
  el: '#ui',
  data: {
    temperatura: 60
  },
  methods: {
    incTemperatura: function() {
      this.temperatura++;
    }
  }
})
</script>

</body>
</html>
```

Essa aplicação apresenta uma temperatura na tela do navegador e um botão para incrementá-la (veja figura a seguir).

Uma Simples SPA

Temperatura: 60

Figura 4.5: Interface da Single-Page Application do exemplo

Código Fonte: O código do exemplo está disponível neste [link](#). Se quiser executar a aplicação no seu navegador, basta usar esse [link](#).

O interessante é que SPAs seguem uma arquitetura parecida com MVC. No exemplo mostrado, a interface da SPA, contendo visão e controle, é implementada em HTML, mais precisamente no código delimitado pela tag <div>. O modelo é implementado em JavaScript, usando-se Vue.js. O código do modelo está delimitado pela tag <script>.

Um segundo ponto interessante é que o framework Vue.js se encarrega de propagar para a visão qualquer alteração no modelo. Por exemplo, quando o método incTemperatura é executado, o valor da temperatura é automaticamente atualizado na interface. O processo contrário também é verdadeiro, apesar de não ser exercitado em nosso exemplo, que é bem simples. Esse recurso de frameworks para construção de SPAs é chamado de **two-way data binding**.

4.4 Microsserviços

Como discutimos no Capítulo 2, métodos ágeis preconizam iterações rápidas, com entregas frequentes de novas releases, a fim de obter feedback e, se for preciso, efetuar mudanças de rumo. Porém, mesmo que uma empresa adote um método ágil — como Scrum — ela vai enfrentar um “gargalo” arquitetural quando precisar lançar novas releases de um produto de forma frequente.

Esse gargalo ocorre porque sistemas, via de regra, seguem em tempo de execução uma arquitetura monolítica. Ou seja, mesmo que o desenvolvimento tenha sido particionado em módulos M₁, M₂, M₃, ..., M_n, em tempo de execução esses módulos são executados, pelo sistema operacional, como um processo único. Assim, todos os módulos compartilham o mesmo espaço de endereçamento. Em outras palavras, em tempo de execução o sistema é um grande **monolito**, como ilustra a próxima figura.

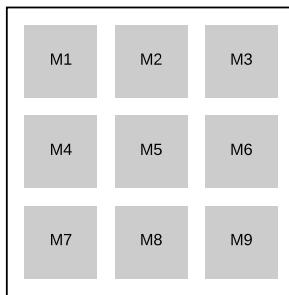


Figura 4.6: Monolito com nove módulos. Em tempo de execução, o sistema executa como um único processo, representado pelo quadrado que delimita os 9 módulos.

Em um monolito, sempre existe o risco de que uma mudança realizada por um time T em um módulo M_i cause um efeito colateral em um módulo M_j . Por exemplo, M_i e M_j podem compartilhar uma variável global ou um atributo estático. Uma mudança nessa variável, realizada em M_i , pode comprometer o funcionamento de M_j . Na realidade, esse risco é maior do que um leitor iniciante em desenvolvimento de sistemas pode imaginar.

Para evitar que os clientes sejam surpreendidos com bugs inesperados em seus sistemas, as empresas que usam arquiteturas monolíticas adotam um processo rigoroso e “burocrático” para o lançamento de novas releases. Esse processo pode incluir até mesmo testes manuais antes da liberação do sistema para produção. Por testes manuais, queremos dizer um testador usar as funcionalidades mais críticas do sistema, de forma a simular uma sessão de uso por um cliente final.

Para resolver esse gargalo — o desenvolvimento passou a ser ágil, mas a entrada em produção permanece burocrática —, recentemente, algumas empresas passaram a migrar os seus monolitos para uma arquitetura baseada em **microsserviços**. A ideia é simples: certos grupos de módulos são executados em processos independentes, sem compartilhamento de memória. Ou seja, o sistema é decomposto em módulos não apenas em tempo de desenvolvimento, mas também em tempo de execução. Com isso, as chances de que mudanças em um módulo causem problemas em outros módulos são menores.

Quando os módulos são separados em processos distintos não há mais possibilidade de que um módulo acesse um recurso interno de outro módulo, como uma variável global, um atributo estático ou uma interface interna. Em vez disso, por construção,

toda comunicação tem que ocorrer via interfaces públicas dos módulos. Assim, microsserviços são um instrumento para garantir que os times de desenvolvimento somente usem interfaces públicas de outros sistemas. A obediência à essa regra é garantida pelo sistema operacional.

A próxima figura mostra uma versão baseada em microsserviços do nosso exemplo. Nela, continuam existindo nove módulos. Mas eles são executados por seis processos independentes, representados pelos quadrados ou retângulos em volta dos módulos. Os módulos M1, M2, M3 e M6 são executados, cada um deles, em um processo independente. Os módulos M4 e M5 são executados em um quinto processo. Por fim, os módulos M7, M8 e M9 são executados, em conjunto, em um sexto processo.

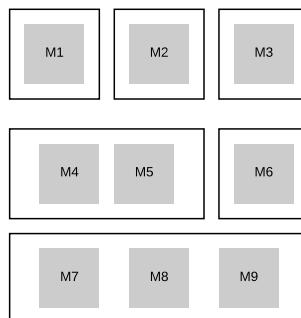


Figura 4.7: Servidor com seis microsserviços: M1, M2, M3, M4-M5, M6, M7-M8-M9. Cada microsserviço executa como um processo autônomo.

Até esse ponto da explicação, usamos o termo processo, mas o nome do padrão refere-se a eles como **serviços**. Ainda, os serviços são micro porque não implementam funcionalidades complexas. Lembre-se que eles são desenvolvidos por times ágeis, que como dissemos no Capítulo 2, são times pequenos, com cerca de cinco desenvolvedores, por exemplo. Consequentemente, times pequenos não têm capacidade para implementar grandes serviços.

Uma segunda vantagem de microsserviços é **escalabilidade**. Quando um monolito enfrenta problemas de performance, uma solução consiste em disponibilizar instâncias do sistema em máquinas diferentes, como mostra a próxima figura. Essa solução é chamada de **escalabilidade horizontal**. Por exemplo, ela permite dividir os clientes do sistema entre as duas instâncias mostradas na figura. Como se trata de um monolito, as duas instâncias são idênticas, isto é, possuem os mesmos módulos.

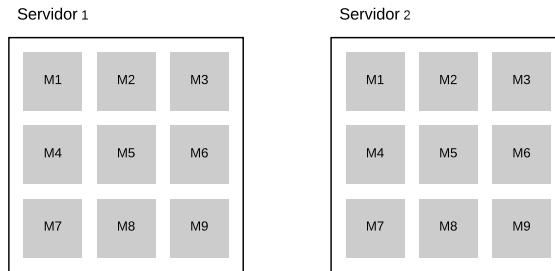


Figura 4.8: Servidor 1, executando o monolito em um único processo. E Servidor 2, executando uma réplica do monolito.

Porém, os problemas de performance podem ser causados por serviços específicos; por exemplo, apenas pelo serviço de autenticação de usuários. Então, microsserviços permitem replicar apenas os componentes diretamente relacionados com os problemas de performance. A próxima figura mostra uma nova instalação do nosso sistema baseado em microsserviços. O segundo servidor que foi disponibilizado inclui apenas instâncias do serviço M1. A suposição é que M1 é responsável pela maior parte dos problemas de performance da instalação inicial. Na primeira instalação, tínhamos uma única instância de M1. Agora, temos seis instâncias, todas elas em um novo servidor.

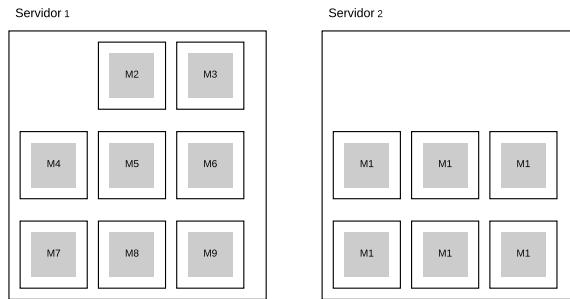


Figura 4.9: Servidor 1, com todos os microsserviços, exceto M1. Servidor 2, executando seis processos, todos eles relativos a M1.

Até o momento, listamos duas vantagens de microsserviços: (1) eles permitem a evolução mais rápida e independente de um sistema, permitindo que cada time

tenha seu próprio regime de liberação de novas releases; (2) eles permitem escalar um sistema em um nível de granularidade mais fino do que é possível com monolitos. Mas existem pelo menos mais duas vantagens:

- Como os microsserviços são autônomos e independentes eles podem ser implementados em tecnologias diferentes, incluindo linguagens de programação, frameworks e bancos de dados. O microsserviço de cadastro de clientes em um sistema de comércio eletrônico, por exemplo, pode ser implementado em Java com um banco de dados relacional. Já o microsserviço de recomendação de novas compras pode ser implementado em Python com um banco de dados NoSQL.
- Quando se usa um monolito, falhas são totais. Se o banco de dados cair, todos os serviços ficam fora do ar. Por outro lado, em arquiteturas baseadas em microsserviços podemos ter **fallas parciais**. Por exemplo, suponha que o microsserviço de recomendação de compras do exemplo do item anterior fique fora do ar. Os clientes ainda conseguirão pesquisar por produtos, fazer compras, etc. Mas eles irão receber uma mensagem na área de recomendação da página dizendo que as recomendações estão desabilitadas; ou então, essa área ficará vazia; ou ainda, ela não será apresentada para os usuários durante o período em que o microsserviço de recomendações estiver fora do ar.

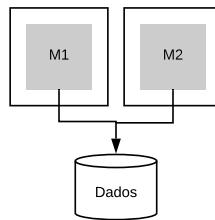
Arquiteturas baseadas em microsserviços tornaram-se possíveis devido ao aparecimento de plataformas de **computação em nuvem**. Com essas plataformas, empresas não precisam mais comprar e manter hardware e software básico, como sistema operacional, bancos de dados e servidores Web. Em vez disso, elas podem alugar uma máquina virtual em uma plataforma de computação em nuvem e pagar por hora de utilização da máquina. Com isso, fica mais fácil escalar um microsserviço horizontalmente, acrescentando novas máquinas virtuais.

Aprofundamento: Microsserviços constituem um exemplo de aplicação da **Lei de Conway**. Formulada em 1968 por Melvin Conway, ela é uma das leis empíricas sobre desenvolvimento de software, assim como a Lei de Brooks, que estudamos no Capítulo 1. A Lei de Conway afirma o seguinte: empresas tendem a adotar arquiteturas de software que são cópias de suas estruturas organizacionais. Em outras palavras, a arquitetura dos sistemas de uma empresa tende a espelhar seu organograma. Por isso, não é coincidência que microsserviços sejam usados, principalmente, por grandes empresas de Internet que possuem centenas de times de desenvolvimento distribuídos em

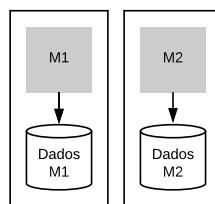
diversos países. Além de descentralizados, esses times são autônomos e sempre incentivados a produzir inovações.

4.4.1 Gerenciamento de Dados

Pelo menos na sua forma pura, microsserviços devem ser autônomos também do ponto de vista de dados. Isto é, eles devem gerenciar os dados de que precisam para prover o seu serviço. Logo, o cenário ilustrado pela figura a seguir — no qual dois microsserviços compartilham o mesmo banco de dados — não é recomendável em uma arquitetura baseada em microsserviços.



O ideal é que M1 e M2 sejam independentes também do ponto de vista de bancos de dados, com mostrado na próxima figura. O principal motivo é que quando se têm um único banco de dados ele também pode se transformar em um gargalo à evolução do sistema.



Por exemplo, equipes e arquiteturas tradicionais de desenvolvimento costumam ter um administrador de dados, a quem cabe cuidar das tabelas do banco de dados. Qualquer mudança no banco de dados — como a criação de uma coluna em uma tabela — precisa de aprovação do administrador de dados. Logo, essa autoridade central tem que conciliar os interesses, muitas vezes conflitantes, das diversas equipes de desenvolvimento. Por isso, suas decisões podem se tornar lentas e burocráticas, atrasando a evolução do sistema.

4.4.2 Quando não usar microsserviços?

Até esse momento, apresentamos as vantagens e benefícios de microsserviços. Mas é importante dizer que essa arquitetura é mais complexa do que uma arquitetura monolítica. O motivo é que microsserviços são processos independentes, isto é, eles por construção dão origem a sistemas distribuídos. Logo, ao usar microsserviços, temos que enfrentar todos os desafios que aparecem quando se implementa um sistema distribuído. Dentre eles, podemos citar:

- Complexidade: quando dois módulos executam em um mesmo processo, a comunicação entre eles é por meio de chamadas de métodos. Quando esses módulos estão em máquinas diferentes, a comunicação entre eles deve usar algum protocolo de comunicação, como **HTTP/REST**. Ou seja, os desenvolvedores terão que dominar e usar um conjunto de tecnologias para comunicação em redes.
- Latência: a comunicação entre microsserviços também envolve um atraso maior, que chamamos de **latência**. Quando um cliente chama um método em um sistema monolítico, a latência é mínima. Por exemplo, é raro um desenvolvedor deixar de usar uma chamada de método apenas para melhorar o desempenho de seu sistema. Porém, esse cenário muda quando o serviço chamado está em uma outra máquina, talvez do outro lado do planeta no caso de uma empresa global. Nessas situações, existe um custo de comunicação que não é desprezível. Qualquer que seja o protocolo de comunicação usado, essa chamada terá que passar pelo cabo da rede — ou pelo ar e pela fibra ótica — até chegar à máquina de destino.
- Transações Distribuídas: Como vimos, microsserviços devem ser autônomos também do ponto de vista de dados. Isso torna mais complexo garantir que operações que operam em dois ou mais bancos de dados sejam atômicas, isto é, ou elas executam com sucesso em todos os bancos ou então falham. Suponha, por exemplo, dois microsserviços de pagamento via cartão de crédito, que vamos chamar de X e Y. Suponha que uma loja virtual permita dividir o valor da compra entre os dois cartões. Por exemplo, uma compra de R\$ 2.000,00 pode ser paga debitando-se R\$ 1.500,00 no cartão X e R\$ 500,00 no cartão Y. Porém, essas transações devem ser atômicas: ou os dois cartões são debitados ou nenhum deles é debitado. Por isso, em arquiteturas baseadas em microsserviços, protocolos de transações distribuídas, como **two-phase commit**, podem ser necessários para garantir uma semântica de transações em operações que escrevem em mais de um banco de dados.

4.5 Arquiteturas Orientadas a Mensagens

Neste tipo de arquitetura, a comunicação entre clientes e servidores é mediada por um terceiro serviço que têm a única função de prover uma **fila de mensagens**, como mostra a próxima figura.

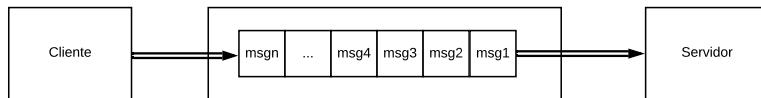


Figura 4.10: Arquitetura Orientada a Mensagens

Os clientes atuam como produtores de informações, isto é, eles inserem mensagens na fila. E os servidores atuam como consumidores de mensagens, isto é, eles retiram mensagens da fila e processam a informação contida nelas. Uma mensagem é um registro (ou um objeto) com um conjunto de dados. E uma fila de mensagens é uma estrutura do tipo **FIFO** (*first in, first out*), isto é, a primeira mensagem a entrar na fila é a primeira a ser consumida pelo servidor.

Com o uso de filas de mensagens, a comunicação pelo lado do cliente torna-se **assíncrona**, pois uma vez que a informação seja colocada na fila, o cliente está liberado para continuar seu processamento. Por isso, é importante que o serviço de mensagens seja instalado em uma máquina estável e com alto poder de processamento. Também é importante que a fila de mensagens seja persistente. Se o servidor que gerencia a fila cair, os dados não podem ser perdidos. Como filas de mensagens são muito usadas na construção de sistemas distribuídos, existem soluções prontas no mercado. Ou seja, provavelmente você não vai implementar sua própria fila de mensagens, mas usar soluções de empresas conhecidas ou então mantidas por fundações de desenvolvimento de sistemas de código aberto. Algumas vezes, filas de mensagens são chamadas também de **brokers** de mensagens.

Além de permitirem comunicação assíncrona entre clientes e servidores, filas de mensagens viabilizam duas formas de desacoplamento entre os componentes de uma aplicação distribuída:

- **Desacoplamento no espaço:** clientes não precisam conhecer os servidores e vice-versa. Em outras palavras, o cliente é exclusivamente um produtor de informações. Mas ele não precisa saber quem vai consumir essa informação. O raciocínio inverso vale para servidores.
- **Desacoplamento no tempo:** clientes e servidores não precisam estar

simultaneamente disponíveis para se comunicarem. Se o servidor estiver fora do ar, os clientes podem continuar produzindo mensagens e colocando-as na fila. Quando o servidor voltar a funcionar, ele irá processar as mensagens.

Desacoplamento no espaço torna soluções baseadas em filas de mensagens bastante flexíveis. Os times de desenvolvimento — tanto do sistema cliente, como do sistema servidor — podem trabalhar e evoluir seus sistemas com autonomia. Atrasos de um time não travam a evolução de outros times, por exemplo. Para isso, basta que o formato das mensagens permaneça estável ao longo do tempo. Já desacoplamento no tempo torna a solução robusta a falhas. Por exemplo, quedas do servidor não têm impacto nos clientes. No entanto, é importante que o broker de mensagens seja estável e capaz de armazenar uma grande quantidade de mensagens. Para garantir a disponibilidade desses brokers, eles costumam ser gerenciados pelos times de infraestrutura básica das empresas.

Filas de mensagens permitem também escalar mais facilmente um sistema distribuído. Para isso, basta configurar múltiplos servidores consumindo mensagens da mesma fila, como mostra a próxima figura.

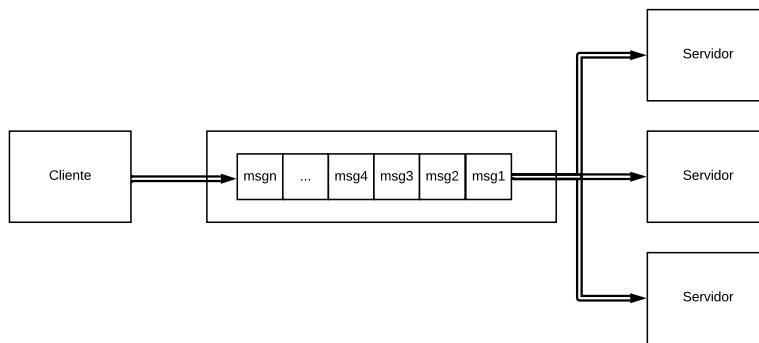


Figura 4.11: Fila de Mensagens com vários consumidores.

4.5.1 Exemplo: Empresa de Telecomunicações

Suponha que uma empresa de telecomunicações tenha dois sistemas principais: vendas e engenharia. O sistema de vendas é responsável pela interação com os clientes da empresa, incluindo vendas de pacotes de voz, dados e TV a cabo. Por outro lado, o sistema de engenharia é responsável por ativar e configurar os serviços contratados, interagindo para isso com os equipamentos da empresa, tais como centrais telefônicas, estações de celulares, roteadores, etc. Portanto, quando

um serviço é contratado no sistema de vendas, ele tem que ser provisionado no sistema de engenharia.

Essa empresa de telecomunicações pode usar uma fila de mensagens para mediar a comunicação entre o sistema de vendas e o sistema de engenharia. Ao se vender um novo pacote de serviços, o sistema de vendas irá depositar uma mensagem na fila de mensagens, com as informações do pacote. Cabe ao sistema de engenharia ler essa mensagem e ativar o serviço que foi vendido.

Ao optar por uma fila de mensagens, a integração entre os sistemas de vendas e de engenharia pode não ocorrer em tempo real. Por exemplo, se o serviço de engenharia estiver ocupado, com várias ativações complexas de serviços pendentes, pode demorar um pouco até que um determinado serviço seja ativado. Por outro lado, a solução com fila de mensagens permite ativar os serviços mais rapidamente do que por meio de uma solução **batch**. Nesse tipo de solução, o sistema de vendas geraria ao final de cada dia um arquivo com todos os pacotes vendidos. Esse arquivo seria processado durante a noite pelo sistema de engenharia. Portanto, um cliente poderia ter que esperar quase 24 horas para ter seu serviço ativado.

4.6 Arquiteturas Publish/Subscribe

Em arquiteturas publish/subscribe, as mensagens são chamadas de **eventos**. Os componentes da arquitetura são chamados de **publicadores** (*publishers*) e **assinantes** (*subscribers*) de eventos. Publicadores produzem eventos e os publicam no serviço de publish/subscribe, que normalmente executa em uma máquina separada. Assinantes devem previamente assinar eventos de seu interesse. Quando um evento é publicado, os seus assinantes são notificados, conforme mostra a próxima figura.

Assim como ocorre quando se usa filas de mensagens, arquiteturas publish/subscribe também oferecem desacoplamento no espaço e no tempo. No entanto, existem duas diferenças principais entre publish/subscribe e filas de mensagens:

- Em publish/subscribe, um evento gera notificações em todos os seus assinantes. Por outro lado, em filas de mensagens, as mensagens são sempre consumidas — isto é, retiradas da fila — por um único servidor. Portanto, em publish/subscribe temos um estilo de comunicação de 1 para n , também conhecida como **comunicação em grupo**. Já em filas de mensagens, a comunicação é 1 para 1, também chamada de **comunicação ponto-a-ponto**.
- Em publish/subscribe, os assinantes são notificados assincronamente. Primeiro, eles assinam certos eventos e, então, continuam seu processamento.

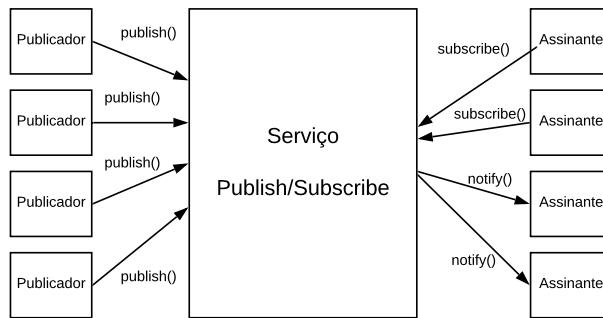


Figura 4.12: Arquitetura Publish/Subscribe

Quando o evento de interesse ocorre, eles são notificados por meio da execução de um determinado método. Por outro lado, quando se usa uma fila de mensagens, os servidores — isto é, os consumidores das mensagens — têm que “puxar” (*pull*) as mensagens da fila.

Em alguns sistemas publish/subscribe, eventos são organizados em **tópicos**, que funcionam como categorias de eventos. Quando um publicador produz um evento, ele deve informar seu tópico. Assim, clientes não precisam assinar todos eventos que ocorrem no sistema, mas apenas eventos de um certo tópico.

Arquiteturas publish/subscribe são, às vezes, chamadas de **arquiteturas orientadas a eventos**. O serviço de publish/subscribe, às vezes, é chamado também de **broker de eventos**, pois ele funciona como um barramento por onde devem trafegar todos os eventos. É importante mencionar ainda que um sistema publish/subscribe lembra o padrão de projeto **Observador**, que estudamos no Capítulo 6. No entanto, publish/subscribe é uma solução para implementação de sistemas distribuídos. Ou seja, produtores e assinantes são processos distintos e, na maioria das vezes, distribuídos. Por outro lado, o padrão de projeto Observador não foi proposto no contexto de arquiteturas distribuídas.

4.6.1 Exemplo: Companhia Aérea

Vamos agora usar os sistemas de uma companhia aérea para ilustrar uma arquitetura publish/subscribe. Suponha que essa companhia tenha um sistema de vendas, que é usado pelos clientes para comprar passagens aéreas. Após efetuar uma venda, esse sistema pode gerar um evento, com todos os dados da venda (data,

horário, número do vôo, dados do passageiro, etc). A figura a seguir ilustra a arquitetura proposta para o sistema.

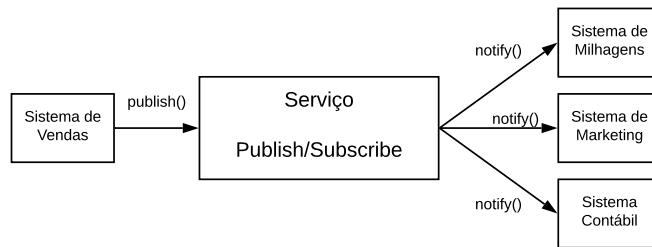


Figura 4.13: Arquitetura Pub/Sub em uma companhia aérea

O evento “venda” será então assinado por três sistemas da companhia aérea: (1) sistema de milhagens, pois as milhas relativas à passagem devem ser creditadas na conta do passageiro; (2) sistema de marketing, que pode usar os dados da venda para fazer ofertas para o cliente, como aluguel de carros, promoção para classe executiva, etc; (3) Sistema de contabilidade, pois a venda que foi realizada precisa ser incluída na contabilidade da empresa.

Essa arquitetura tem as seguintes características interessantes: (1) comunicação em grupo, pois o mesmo evento é assinado por três sistemas; (2) desacoplamento no espaço, pois o sistema de vendas não tem conhecimento dos sistemas interessados nos eventos que ele publica; (3) desacoplamento no tempo, pois o sistema de publish/subscribe reenvia os eventos caso os sistemas assinantes estejam fora do ar; (4) notificação assíncrona, pois os assinantes são notificados assim que um evento ocorre; isto é, eles não precisam consultar periodicamente o sistema publish/subscribe sobre a ocorrência dos eventos de interesse.

4.7 Outros Padrões Arquiteturais

Pipes e Filtros é um tipo de arquitetura orientada a dados, na qual os programas — chamados de **filtros** — têm como função processar os dados recebidos na entrada e gerar uma nova saída. Os filtros são conectados por meio de **pipes**, que agem como buffers. Isto é, pipes são usados para armazenar a saída de um filtro, enquanto ela não é lida pelo próximo filtro da sequência. Com isso, os filtros não precisam conhecer seus antecessores e sucessores, o que torna esse tipo de arquitetura bastante flexível, permitindo as mais variadas combinações de programas. Além disso, por construção, filtros podem ser executados em paralelo. O exemplo clássico

de arquitetura baseada em pipes e filtros são os comandos de sistemas Unix. Por exemplo, a linha de comando:

```
ls | grep csv | sort
```

especifica a execução de três comandos (filtros) que são conectados por dois pipes (barras verticais). No caso dos comandos Unix, as entradas e saídas são sempre arquivos texto.

Cliente/Servidor é uma arquitetura muito usada na implementação de serviços básicos de rede. Clientes e servidores são os dois únicos módulos desse tipo de arquitetura e eles se comunicam por meio de uma rede. Os clientes solicitam serviços ao módulo servidor e aguardam o processamento. Arquiteturas cliente/servidor são usadas para implementar serviços como os seguintes: (1) serviço de impressão, que possibilita que clientes imprimam em uma impressora remota, que não está fisicamente conectada à máquina deles; (2) serviço de arquivos, que possibilita que clientes acessem o sistemas de arquivos (isto é, o disco) de uma máquina servidora; (3) serviço de bancos de dados, que permite que clientes acessem um banco de dados instalado em uma outra máquina; (4) serviço Web, que permite que clientes (navegadores) acessem recursos (páginas HTML) armazenadas e providas por um servidor Web.

Arquiteturas **peer-to-peer** são arquiteturas distribuídas nas quais os módulos da aplicação podem desempenhar tanto o papel de cliente, como o papel de servidor. Em outras palavras, esses módulos — chamados de **pares** — são tanto consumidores como provedores de recursos. Por exemplo, BitTorrent é um protocolo peer-to-peer para compartilhamento de arquivos na Internet. Aplicações que implementam o protocolo podem tanto prover arquivos como realizar o download de arquivos disponíveis na rede.

4.8 Anti-padrões Arquiteturais

Vamos encerrar com a descrição de um **anti-padrão** arquitetural, isto é, uma organização de sistemas que não é recomendada. Talvez, o mais conhecido anti-padrão é chamado de **big ball of mud** (ou “grande bola de lama”). Esse anti-padrão — proposto por Brian Foote e Joseph Yoder ([link](#)) — descreve sistemas nos quais qualquer módulo comunica-se com praticamente qualquer outro módulo, como mostra a próxima figura.

Ou seja, um “big ball of mud” não possui uma arquitetura definida. Em vez disso, o que existe é uma explosão no número de dependências, que dá origem a um

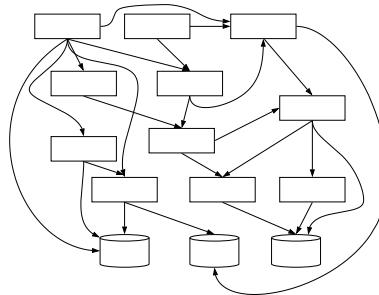


Figura 4.14: Anti-padrão *big ball of mud*

espaguete de código. Consequentemente, a manutenção do sistema torna-se muito difícil e arriscada.

Mundo Real: Em um artigo publicado em 2009 na revista IEEE Software ([link](#)), Santonu Sarkar e mais cinco colegas — na época consultores da empresa Indiana InfoSys — descrevem uma experiência de modularização de um grande sistema bancário. O sistema nasceu no final da década de 90 e desde então aumentou seu tamanho em 10 vezes: passou de 2.5 milhões para mais de 25 milhões de linhas de código! Segundo os autores, os times de desenvolvimento do sistema contavam com “várias centenas de engenheiros”. Apesar de não usarem o termo, o artigo caracteriza a arquitetura desse sistema bancário como uma “big ball of mud”. Por exemplo, eles mencionam que apenas um diretório, chamado “sources”, possuía quase 15 mil arquivos. Em seguida, os autores analisam o problema que era manter esse sistema: (1) o tempo de aprendizado de novos engenheiros só aumentava, passando de três para sete meses, no intervalo de cinco anos; (2) frequentemente, a correção de bugs introduzia novos bugs no sistema; (3) o tempo de implementação de novas funcionalidades, mesmo que simples, também estava aumentando muito.

Pode parecer que sistemas como esse — analisado no artigo da IEEE Software — são exceções. Porém, eles são mais comuns do que se pode imaginar. E a origem do problema reside na transformação do código em uma “big ball of mud”. De forma interessante, os autores citam que o banco tentou contornar o problema adotando práticas como documentação detalhada, revisões de código e programação em pares. Porém, todas se mostraram incapazes de atenuar os problemas causados pela arquitetura em forma de “big ball of mud”.

Bibliografia

James Lewis, Martin Fowler. Microservices: a definition of this new architectural term. Blog post, 2014.

Martin Fowler. Patterns of Enterprise Application Architecture, Addison-Wesley, 2002.

Martin Fowler. Who Needs an Architect, IEEE Software, vol. 20, issue 5, p. 11-13, 2003.

Patrick Eugster et al. The many faces of publish/subscribe. ACM Computing Surveys, vol. 35, issue 2, p. 114-131, 2003.

Glenn Krasner, Stephen Pope. A cookbook for using the model-view controller user interface paradigm in Smalltalk-80. Journal of Object-Oriented Programming, vol. 1, issue 3, p. 26-49, 1988.

Kevlin Henney, Frank Buschmann, Douglas Schmidt. Pattern-Oriented Software Architecture: A Pattern Language for Distributed Computing, Vol. 4, John Wiley & Sons, 2007.

Exercícios de Fixação

1. Dada a sua complexidade, sistemas de bancos de dados são componentes relevantes na arquitetura de qualquer tipo de sistema. Verdadeiro ou falso? Justifique.
2. Qual a diferença entre classes Controladoras em uma Arquitetura MVC tradicional e classes Controladoras de um sistema Web implementado usando um framework MVC como Ruby on Rails?
3. Descreva resumidamente quatro vantagens de microsserviços.
4. Por que microsserviços não são uma bala de prata? Isto é, descreva pelo menos três desvantagens do uso de microsserviços.
5. Explique o que significa desacoplamento no espaço e desacoplamento no tempo. Por que arquiteturas baseadas em filas de mensagens e arquiteturas Publish/Subscribe oferecem essa forma de desacoplamento?
6. Quando uma empresa deve considerar o uso de uma arquitetura baseada em filas de mensagens ou uma arquitetura publish/subscribe?
7. Explique o objetivo do conceito de tópicos em uma arquitetura publish/subscribe.

8. (POSCOMP, 2019, adaptado) Marque V ou F.

() Mesmo que um dado padrão arquitetural ofereça uma solução para o problema sendo resolvido, nem sempre ele é adequado. Fatores como contexto e o sistema de forças que afeta a solução fazem também parte do processo de avaliação e da escolha de padrões adequados.

() Padrão MVC é uma adaptação do padrão arquitetural Camadas. A Camada Visão lida com a apresentação e a manipulação da interface, a Camada Modelo organiza os objetos específicos da aplicação, e a Camada Controle posiciona-se entre estas duas com as regras do negócio.

() O padrão Broker é voltado a problemas de ambientes distribuídos. Sugere uma arquitetura na qual um componente (broker) estabelece uma mediação que permite um desacoplamento entre clientes e servidores.

Capítulo 5

Refactoring

I'm not a great programmer; I'm just a good programmer with great habits.

– Kent Beck

Este capítulo inicia com uma introdução a refactorings, os quais são modificações realizadas em um programa para facilitar o seu entendimento e futura evolução. Na Seção 9.2, apresentamos uma série de operações de refactoring, incluindo exemplos de código, alguns deles de refactorings reais, realizados em sistemas de código aberto. Em seguida, na Seção 9.3, discutimos alguns aspectos sobre a prática de refactoring, incluindo a importância de uma boa suíte de testes de unidade. A Seção 9.4 apresenta os recursos oferecidos por IDEs para realização automatizada de refactorings. Para finalizar, a Seção 9.5 descreve uma lista de code smells, isto é, indicadores de que uma determinada estrutura de código não está “cheirando bem” e que, portanto, poderia ser objeto de uma refatoração.

5.1 Introdução

No capítulo anterior, vimos que software precisa ser testado, como qualquer produto de engenharia. A mesma recomendação vale para atividades de manutenção. Isto é, software também precisa de manutenção. Na verdade, na Introdução deste livro, já comentamos que existem diversos tipos de manutenção que podem ser realizadas em sistemas de software. Quando um bug é detectado, temos que realizar uma **manutenção corretiva**. Quando os usuários ou o dono do produto solicitam uma nova funcionalidade, temos que realizar uma **manutenção evolutiva**. Quando

uma regra de negócio ou alguma tecnologia usada pelo sistema muda, temos que reservar tempo para uma **manutenção adaptativa**.

Além disso, sistemas de software também envelhecem, como ocorre com os seres vivos. Ainda no início da década de 1970, Meir Lehman — então trabalhando na IBM — começou a observar e analisar esse fenômeno e, como resultado, enunciou um conjunto de leis empíricas sobre envelhecimento, qualidade interna e evolução de sistemas de software, que ficaram conhecidas **Leis da Evolução de Software** ou simplesmente **Leis de Lehman**. As duas primeiras leis enunciadas por Lehman foram as seguintes:

- Um sistema de software deve ser continuamente mantido para se adaptar ao seu ambiente. Esse processo deve continuar até o ponto em que se torna mais vantajoso substituí-lo por um sistema completamente novo.
- À medida que um sistema sofre manutenções, sua complexidade interna aumenta e a qualidade de sua estrutura deteriora-se, a não ser que um trabalho seja realizado para estabilizar ou evitar tal fenômeno.

A primeira lei justifica a necessidade de manutenções adaptativas e também evolutivas em sistemas de software. Ela também menciona que sistemas podem “morrer”, isto é, pode chegar a um ponto em que vale mais a pena descontinuar o desenvolvimento de um sistema e substituí-lo por um novo. Já a segunda Lei de Lehman afirma que manutenções tornam o código e a estrutura interna de um sistema mais complexos e difíceis de manter no futuro. Em outras palavras, existe uma deterioração natural da qualidade interna de um sistema, à medida que ele passa por manutenções e evoluções. No entanto, a segunda lei faz uma ressalva: um certo trabalho pode ser realizado para estabilizar ou mesmo evitar esse declínio natural da qualidade interna de sistemas de software. Modernamente, esse trabalho é chamado de **refactoring**.

Refactorings são transformações de código que melhoram a manutenibilidade de um sistema, mas sem afetar o seu funcionamento. Para explicar essa definição, vamos dividi-la em três partes. Primeiro, quando a definição menciona “transformações de código”, ela está se referindo a modificações no código, como dividir uma função em duas, renomear uma variável, mover um função para outra classe, extrair uma interface de uma classe, etc. Em seguida, a definição menciona o objetivo de tais transformações: “melhorar a manutenibilidade” do sistema, isto é, melhorar sua modularidade, melhorar seu projeto ou arquitetura, melhorar sua testabilidade, tornar o código mais legível, mais fácil de entender e modificar, etc. Por fim, coloca-se uma restrição: não adianta melhorar a manutenibilidade do sistema e prejudicar o seu funcionamento. Ou seja, refactoring deve entregar o sistema

funcionando exatamente como antes das transformações. Uma outra maneira de dizer isso é afirmando que refactorings devem preservar o comportamento ou a semântica do sistema.

No entanto, nas décadas de 70 e 80, quando as Leis de Lehman foram formuladas, o termo refactoring ainda não era usado. Um dos primeiros usos do termo em Engenharia de Software ocorreu em 1992, em uma tese de doutorado defendida por William Opdyke, na Universidade de Illinois, EUA ([link](#)). Em seguida, em 1999, refactoring — já com esse nome — foi incluído entre as práticas de programação preconizadas por Extreme Programming. Na primeira edição do livro de XP, recomenda-se que desenvolvedores devem realizar refactorings com o objetivo de “reestruturar seus sistemas, sem mudar o comportamento deles e sim para remover duplicação de código, melhorar a comunicação com outros desenvolvedores, simplificar o código ou torná-lo mais flexível”.

Em 2000, Martin Fowler lançou a primeira edição de um livro dedicado exclusivamente a refactoring, que alcançou grande sucesso e contribuiu para popularizar essa prática de programação. Um dos motivos desse sucesso foi o fato de o livro incluir um catálogo com dezenas de refactorings. De uma forma que lembra um catálogo de padrões de projeto (tal como estudamos no Capítulo 6), a apresentação dos refactorings começa dando um nome para eles, o que contribuiu para criar um vocabulário sobre refactoring. No livro, Fowler também apresenta a mecânica de funcionamento de cada refactoring, inclui exemplos de código e discute os benefícios e desvantagens dos refactorings.

Em seu livro, Fowler cita também a frase de Kent Beck que abre esse capítulo e que ressalta a importância de seguir “bons hábitos de programação”, os quais são fundamentais para preservar a saúde de um sistema, garantindo que ele continuará evoluindo por anos. Portanto, desenvolvedores não devem realizar apenas manutenções corretivas, adaptativas e evolutivas. É importante cuidar também da manutenibilidade do sistema, por meio da realização de refactorings.

Tradução: Decidimos não traduzir refactoring quando usado como substantivo. O motivo é que achamos que o substantivo em inglês já faz parte do vocabulário dos desenvolvedores de software brasileiros. Porém, quando usada como verbo (*to refactor*), traduzimos para refatorar.

Aviso: O termo refactoring tornou-se comum em desenvolvimento de software. Por isso, às vezes ele é usado para indicar a melhoria de outros requisitos não-funcionais, que não estão relacionados com manutenibilidade. Por exemplo, frequentemente, ouvimos desenvolvedores mencionar que vão refatorar o código para melhorar seu desempenho, para introduzir concorrência, para melhorar a usabilidade de sua interface, etc. No entanto, neste livro, vamos usar o termo restrito à sua definição original, isto é, apenas para denotar modificações de código que melhoraram a sua manutenibilidade.

5.2 Catálogo de Refactorings

Nesta seção, vamos apresentar os principais refactorings do catálogo de Fowler. Assim como adotado nesse catálogo, vamos comentar sobre os seguintes tópicos na apresentação de cada refactoring: nome, motivação, mecânica de aplicação e exemplos de uso. Além disso, vamos usar alguns exemplos de refactorings reais, realizados por desenvolvedores de sistemas de código aberto disponíveis no GitHub.

5.2.1 Extração de Método

Extração de Método é um dos principais refactorings, tendo como objetivo extrair um trecho de código de um método `f` e levá-lo para um novo método `g`. Então, o método `f` passa a incluir uma chamada de `g`. O próximo código ilustra o funcionamento desse refactoring. Primeiro, o código antes do refactoring:

```
void f () {  
    ... // A  
    ... // B  
    ... // C  
}
```

E agora o código após a extração do método `g`:

```
void g() { // método extraído  
    ... // B  
}  
  
void f () {  
    ... // A  
    g(); // chama método extraído  
    ... // C  
}
```

Existem também variações na mecânica de funcionamento de uma extração de método. Por exemplo, pode-se extrair de uma vez vários métodos g_1, g_2, \dots, g_n de um método f . Pode-se também extrair o mesmo código g de vários métodos f_1, f_2, \dots, f_n . Nesse caso, a extração é usada para eliminar **duplicação de código**, pois o código de g estava aparecendo em diversos métodos.

Para realizar uma Extração de Método pode ser necessário passar parâmetros para o método extraído. Isso ocorre, por exemplo, se o método precisar acessar variáveis locais do método original. O método extraído pode ter ainda que retornar algumas dessas variáveis, caso elas sejam usados depois pelo método original. Por fim, se existirem variáveis locais que somente são usadas no método extraído, deve-se extrair também a declaração delas do método original. Ou seja, elas não serão mais necessárias no método original.

Extração de Método é conhecido como o “canivete suíço” dos refactorings, pois ele é um dos refactorings com mais aplicações. Por exemplo, Extração de Método pode ser usado para quebrar um método grande em métodos menores. Cada um desses métodos tem uma função específica, que deve ficar clara em seu nome. Assim, fica mais fácil entender o método original, pois ele passará a conter apenas uma sequência de chamadas dos métodos extraídos. Vamos mostrar um exemplo a seguir. Outras aplicações de Extração de Método são discutidas em uma seção “Mundo Real”, após o exemplo.

Exemplo: A seguir, mostramos um exemplo real de Extração de Método, realizada em um sistema Android. Esse sistema possui um método `onCreate`, que usa comandos SQL para criar as tabelas do banco de dados que ele manipula. O código dessa primeira versão de `onCreate` — após algumas edições, simplificações e remoções de comentários — é mostrado a seguir. O código fonte original possui mais de 200 linhas.

```
void onCreate(SQLiteDatabase database) { // antes da extração
    // cria tabela 1
    database.execSQL("CREATE TABLE " +
        CELL_SIGNAL_TABLE + " (" + COLUMN_ID +
        " INTEGER PRIMARY KEY AUTOINCREMENT, " + ...
    database.execSQL("CREATE INDEX cellID_index ON " + ...);
    database.execSQL("CREATE INDEX cellID_timestamp ON " + ...);

    // cria tabela 2
    String SMS_DATABASE_CREATE = "CREATE TABLE " +
        SILENT_SMS_TABLE + " (" + COLUMN_ID +
        " INTEGER PRIMARY KEY AUTOINCREMENT, " + ...
    database.execSQL(SMS_DATABASE_CREATE);
```

```

String ZeroSMS = "INSERT INTO " + SILENT_SMS_TABLE +
    " (Address,Display,Class,ServiceCtr,Message) " +
    "VALUES ('"+ ...
database.execSQL(ZeroSMS);

// cria tabela 3
String LOC_DATABASE_CREATE = "CREATE TABLE " +
    LOCATION_TABLE + " (" + COLUMN_ID +
    " INTEGER PRIMARY KEY AUTOINCREMENT, " + ...
database.execSQL(LOC_DATABASE_CREATE);

// mais 200 linhas, criando outras tabelas
}

```

Para simplificar o entendimento desse método, um dos desenvolvedores do IntelliJ resolveu realizar sete Extrações de Método. Cada método extraído ficou responsável pela criação de uma das tabelas. Apenas olhando para o nome de tais métodos, pode-se ter uma ideia das tabelas que eles criam. Após as extrações, o método `onCreate` chama os métodos extraídos, como mostrado a seguir. O tamanho do método caiu de mais de 200 linhas de código para apenas sete linhas. Observe também que os métodos extraídos têm como um parâmetro que representa o banco de dados no qual as tabelas serão criadas.

```

public void onCreate(SQLiteDatabase database) {
    createCellSignalTable(database);
    createSilentSmsTable(database);
    createLocationTable(database);
    createCellTable(database);
    createOpenCellIDTable(database);
    createDefaultMCCTable(database);
    createEventLogTable(database);
}

```

5.2.2 Motivações para Extração de Métodos

Em 2016, junto com Danilo Silva (na época aluno de doutorado do ASERG/DCC/UFMG) e com o Prof. Nikolaos Tsantalis (da Universidade Concordia, em Montreal, no Canadá), realizamos um estudo com desenvolvedores de sistemas GitHub para descobrir suas reais motivações para realização de refactorings ([link](#)). No estudo, revelamos 11 motivações distintas para Extrações de Métodos, conforme mostrado na tabela a seguir.

Motivações para Extração de Métodos	Ocorrências
Extração de um método para ser reusado em outras partes do código	43
Introduzir uma assinatura alternativa para um método existente	25
Dividir um método em partes menores para melhorar o seu entendimento	21
Remover duplicação de código	15
Extração de um método para facilitar a implementação de uma nova funcionalidade ou a correção de um bug	14
Extração de um método que terá um nome melhor ou menos parâmetros. O método original é depreciado e delega as suas chamadas para o método extraído	6
Permitir o teste do método extraído, isto é, melhorar testabilidade	6
Permitir que subclasses sobrescrevam o método que foi extraído	4
Permitir a implementação de um método recursivo	2
Introduzir um método fábrica, isto é, as chamadas de new são movidas para o novo método	1
Implementar operações assíncronas, isto é, o método é extraído para ser executado em uma thread independente	1

Analizando essa tabela, podemos ver que a principal motivação é extrair um código para permitir o seu reúso. Ou seja, o desenvolvedor precisa de um determinado código e descobre que ele já está implementado, mas dentro de um método `f`. Então, ele extrai esse código para um método `g`. Com isso, ele pode agora reusar `g` — isto é, chamar `g` — no código no qual está trabalhando. Veja como um dos desenvolvedores relatou sua motivação para extrair um método:

“Esses refactorings foram realizados para garantir reusabilidade. Eu precisava usar o mesmo código em um novo método. E eu sempre tento reusar código, porque quando começa a ter muita redundância no código ele se torna mais difícil de ser mantido no futuro, visto que quando alguma coisa muda em código duplicado, essa mudança tem que ser replicada em todas as cópias.”

Como pode ser conferido na tabela anterior, a segunda motivação mais frequente consiste em introduzir uma assinatura alternativa para um método. Um exemplo hipotético é mostrado a seguir, envolvendo um método que faz o logging de uma string em um arquivo.

Primeiro vamos mostrar o código original:

```
void log(String msg) {  
    // salva msg em arquivo  
}
```

O desenvolvedor decidiu então prover uma versão alternativa desse método com um parâmetro booleano, que indica se a string também deve ser mostrada na console. Veja o código após a extração desse segundo método `log`:

```
void log(String msg, boolean console) {  
    if (console)  
        System.out.println(msg);  
    // salva msg em um arquivo  
}  
  
void log(String msg) {  
    log(msg, false);  
}
```

O leitor deve observar que, tecnicamente, a transformação mostrada no exemplo anterior é uma Extração de Método. Primeiro, todo o código de `log(String)` foi extraído para `log(String,boolean)`, que é um novo método. Em seguida, pequenos ajustes foram realizados no código extraído. Especificamente, foi adicionado um comando `if`. Tais ajustes são comuns e não descaracterizam a transformação como uma Extração de Método.

5.2.3 Inline de Método

Esse refactoring funciona no sentido contrário a uma extração de método. Suponha um método pequeno, com uma ou duas linhas de código, e que seja chamado poucas vezes. O benefício proporcionado por esse método — em termos de reúso e incremento de legibilidade do código — é pequeno. Portanto, ele pode ser removido do sistema e seu corpo incorporado nos pontos de chamada. No entanto, é importante ressaltar que Inline de Métodos é uma operação mais rara e menos importante do que Extração.

Exemplo: A seguir, mostramos um exemplo de Inline de Método, realizado no sistema IntelliJ, uma IDE para Java. Primeiro, segue o código antes do inline. Podemos ver que o método `writeContentToFile` tem uma única linha de código e é chamado apenas uma vez, pelo método `write`.

```
private void writeContentToFile(final byte[] revision) {
    getVirtualFile().setBinaryContent(revision);
}

private void write(byte[] revision) {
    VirtualFile virtualFile = getVirtualFile();
    ...
    if (document == null) {
        writeContentToFile(revision); // única chamada
    }
    ...
}
```

Os desenvolvedores do IntelliJ decidiram então remover `writeContentToFile` e expandir o seu corpo no único ponto de chamada. O código após o refactoring é mostrado a seguir.

```
private void write(byte[] revision) {
    VirtualFile virtualFile = getVirtualFile();
    ...
    if (document == null) {
        virtualFile.setBinaryContent(revision); // após inline
    }
    ...
}
```

5.2.4 Movimentação de Método

Não é raro encontrar um método implementado na classe errada. Ou seja, apesar de implementado em uma classe A, um método f pode usar mais serviços de uma classe B. Por exemplo, ele pode ter mais dependências para elementos de B do que de sua classe A. Nesses casos, deve-se avaliar a possibilidade de mover f para a classe B. Esse refactoring pode melhorar a coesão da classe A, diminuir o acoplamento entre A e B e, em última instância, tornar ambas as classes mais fáceis de serem entendidas e modificadas.

Pelas suas características, Movimentação de Métodos é um dos refactorings com maior potencial para melhorar a modularização de um sistema. Como sua atuação

não está restrita a uma única classe, Movimentação de Métodos pode ter um impacto positivo na arquitetura do sistema, garantindo que os métodos estejam nas classes apropriadas, tanto do ponto de vista funcional como arquitetural.

Exemplo: O sistema IntelliJ possui um pacote para execução de testes de unidade, via IDE. Esse pacote, por sua vez, possui uma classe com métodos de uso geral, chamada `PlatformTestUtil`. Conforme mostra o código a seguir, essa classe tinha um método chamado `averageAmongMedians`, que calcula a média das medianas de uma parte de um vetor de inteiros. No entanto, esse método não possui relação com a execução de testes de unidade. Por exemplo, ele não usa métodos e atributos de `PlatformTestUtil`. Ou seja, ele é independente do resto da classe.

```
class PlatformTestUtil {  
    ...  
    public static long averageAmongMedians(long[] time,  
                                            int part) {  
        int n = time.length;  
        Arrays.sort(time);  
        long total = 0;  
        for (int i = n/2-n/part/2; i < n/2+n/part/2; i++) {  
            total += time[i];  
        }  
        int middlePartLength = n/part;  
        return middlePartLength == 0 ? 0 : total/middlePartLength;  
    }  
    ...  
}
```

Pelos motivos expostos no parágrafo anterior, um dos desenvolvedores do IntelliJ decidiu mover `averageAmongMedians` para uma classe chamada `ArrayUtil`, que tem como objetivo disponibilizar funções utilitárias para manipulação de vetores. Logo, trata-se de uma classe mais relacionada com a funcionalidade provida por `averageAmongMedians`.

Após a movimentação para a nova classe, as chamadas de `averageAmongMedians` tiveram que ser atualizadas, como mostrado na figura a seguir. No entanto, como o leitor pode observar nesse diff, isso não foi difícil, pois `averageAmongMedians` é um método estático. Logo, apenas o nome de sua classe teve que ser atualizado em cada ponto de chamada.

Em outros casos, no entanto, pode não ser tão simples atualizar as chamadas de um método após ele ser movido para uma nova classe. Isso acontece quando nos pontos de chamada não existem referências para objetos da nova classe do método.

```
- System.out.println("Average among the N/3 median times: " + PlatformTestUtil.averageAmongMedians(time, 3) + "ms");
+ System.out.println("Average among the N/3 median times: " + ArrayUtil.averageAmongMedians(time, 3) + "ms");

//System.out.println("JobLauncher.COUNT    = " + JobLauncher.COUNT);
//System.out.println("JobLauncher.TINY     = " + JobLauncher.TINY_COUNT);

@@ -205,7 +205,7 @@ public int compare(HighlightInfo o1, HighlightInfo o2) {
}

FileEditorManagerEx.getInstanceEx(getProject()).closeAllFiles();

- System.out.println("Average among the N/3 median times: " + PlatformTestUtil.averageAmongMedians(time, 3) + "ms");
+ System.out.println("Average among the N/3 median times: " + ArrayUtil.averageAmongMedians(time, 3) + "ms");
```

Figura 5.1: Atualizando chamadas de método estático após movimentação. `averageAmongMedians` foi movido de `PlatformTestUtil` para `ArrayUtil`.

Uma solução consiste então em deixar uma implementação simples do método na classe de origem. Essa implementação apenas delega as chamadas para a nova classe do método. Consequentemente, nenhum cliente precisará ser alterado.

Um exemplo é mostrado a seguir. Primeiro, o código original:

```
class A {
    B b = new B();
    void f { ... }
}

class B {
    ...
}

class Cliente {
    A a = new A();
    void g() {
        ...
        a.g();
        ...
    }
}
```

E agora o código após a refatoração:

```

class A {
    B b = new B();
    void f {
        b.g(); // apenas delega chamada para B
    }
}

class B { // f foi movido de A para B
    void f { ... }
}

class Cliente {
    A a = new A();
    void g() {
        ...
        a.g(); // não precisa mudar
        ...
    }
}

```

Observe que o método `f` foi movido da classe `A` para a classe `B`. Porém, em `A` ficou uma versão do método que apenas repassa (ou delega) a chamada para `B`. Por isso, o código da classe `Cliente` não precisou ser alterado.

Quando ocorre em uma mesma hierarquia de classes, Movimentação de Métodos ganha nomes especiais. Por exemplo, quando o refactoring move um método de subclasses para uma superclasse, ele é chamado de **Pull Up Method**. Para ilustrar, suponha um mesmo método `f` implementado em duas subclasses `B1` e `B2`. Para evitar **duplicação de código**, pode-se então “subir” com ambas implementações para a superclasse `A`, como mostra o seguinte diagrama de classes:

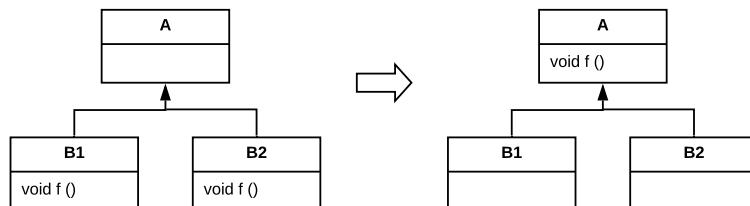


Figura 5.2: Pull Up Method

Por outro lado, quando um método é movido para baixo na hierarquia de classes, isto é, de uma superclasse para uma subclasse, dizemos que foi realizado um **Push Down Method**. Por exemplo, apesar de implementado na superclasse **A**, um método **f** pode ser do interesse de uma única subclasse, digamos que **B1**. Logo, podemos “descer” com sua implementação para **B1**, como mostrado a seguir:

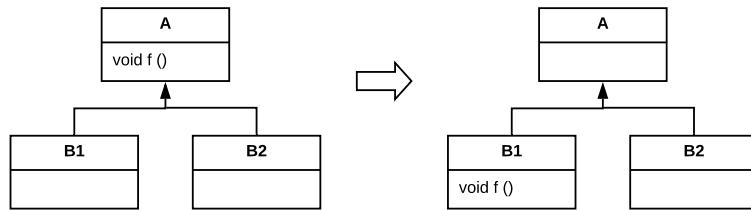


Figura 5.3: Push Down Method

Para concluir, operações de refactoring podem ser feitas em sequência. Por exemplo, suponha a seguinte classe **A** com um método **f**:

```
class A {  
    B b = new B();  
  
    void f(){  
        S1;  
        S2;  
    }  
  
}  
  
class B {  
    ...  
}
```

Primeiro, vamos extrair um método g com o comando s2 de f:

```
class A {  
    B b = new B();  
  
    void g() { // novo método extraído de f()  
        S2;  
    }  
  
    void f(){  
        S1;  
        g();  
    }  
}  
  
class B {  
    ...  
}
```

Na sequência, vamos mover g para uma classe B, como ilustrado a seguir:

```
class A {  
    B b = new B();  
  
    void f(){  
        S1;  
        b.g();  
    }  
}  
  
class B {  
  
    void g() { // movido de A para B  
        S2;  
    }  
}
```

5.2.5 Extração de Classes

Esse refactoring é recomendado quando um sistema possui uma classe A com muitas responsabilidades e atributos. Alguns desses atributos são relacionados e poderiam ter vida própria. Logo, eles podem ser extraídos para uma nova classe B. Em seguida, declaramos em A um atributo do tipo B.

Exemplo: Seja a classe `Pessoa` mostrada a seguir. Além de outros atributos, que foram omitidos do exemplo, ela armazena dados do telefone fixo e celular da pessoa em questão, incluindo código de área e número.

```
class Pessoa {  
    String codigoArea;  
    String fone;  
    String codigoAreaCelular;  
    String celular;  
    ...  
}
```

Assim, podemos extrair de `Pessoa` uma nova classe, chamada `Fone`, dedicada a armazenar informações sobre telefones, conforme mostrado a seguir. Após o refactoring, `Pessoa` passou a ter dois atributos do novo tipo, destinados a armazenar o telefone fixo e o telefone celular de uma pessoa.

```
class Fone { // classe extraída  
    String codigoArea;  
    String numero;  
}  
  
class Pessoa {  
    Fone foneFixo;  
    Fone foneCelular  
    ...  
}
```

Um refactoring similar a **Extração de Classes** é chamado de **Extração de Interfaces**. Por exemplo, suponha que uma biblioteca de estruturas de dados implemente as classes `LinkedList` e `ArrayList`. Desses duas classes podemos extrair uma interface `List`, com os métodos comuns a elas. Assim, os clientes de tais classes podem começar a praticar o princípio de projeto “Prefira Interfaces a Classes Concretas”, que estudamos no Capítulo 5.

5.2.6 Renomeação

Existe uma frase provocativa, atribuída a Phil Karlton, que afirma que “existem apenas duas coisas difíceis em Ciência da Computação: invalidação de cache e dar nomes às coisas.” Como dar nomes é difícil, frequentemente temos que renomear um elemento de código, seja ele uma variável, função, método, parâmetro, atributo, classe, etc. Isso pode ocorrer porque o nome dado ao elemento não foi uma boa

escolha. Um outro motivo é que a responsabilidade desse elemento pode ter mudado com o tempo e assim seu nome ficou desatualizado. Em ambos os casos, recomenda-se realizar um dos refactorings mais populares que existe: **renomeação**. Isto é, dar um nome mais adequado e significativo para o elemento de código.

Quando esse refactoring é aplicado, a parte mais complexa não é renomear o elemento, mas atualizar os pontos do código em que ele é referenciado. Por exemplo, se um método `f` é renomeado para `g`, todas as chamadas de `f` devem ser atualizadas. Na verdade, se `f` for muito usado, pode ser interessante extraí-lo para um novo método, com o novo nome, e manter o nome antigo, mas depreciado.

Para mostrar um exemplo, suponha o seguinte método `f`:

```
void f () {  
    // A  
}
```

Segue agora o código após o refactoring. Veja que extraímos um método `g`, com o código antigo de `f`, que foi então depreciado.

```
void g() {    // novo nome do método  
    // A  
}  
  
@deprecated  
void f() {    // mantém nome antigo  
    g();        // mas delega chamada para novo nome  
}
```

Depreciação é um mecanismo oferecido por linguagens de programação para indicar que um elemento de código está desatualizado e, portanto, desencorajar seu uso. Quando o compilador descobre que um trecho de código está usando um elemento depreciado, ele emite um *warning*. No exemplo anterior, o método `f` não foi simplesmente renomeado para `g`. Em vez disso, primeiro criou-se um método `g` com o código original de `f`. Em seguida, `f` foi depreciado e seu código modificado para apenas chamar `g`.

A estratégia baseada em depreciação torna a renomeação mais segura, pois ela não obriga a atualização de uma só vez de todas as chamadas de `f`, que podem ser muitas e espalhadas em diversos programas. Ou seja, ela dá tempo para que os clientes se adaptem à mudança e passem a usar o novo nome. Na verdade, refactorings — mesmo os mais simples, como uma renomeação — devem ser

realizado em pequenos passos, ou em **baby steps**, para ter certeza que eles não vão prejudicar o correto funcionamento de um sistema.

5.2.7 Outros Refactorings

Os refactorings apresentados anteriormente têm maior potencial para melhorar o projeto de um sistema, pois eles envolvem operações com um escopo global, como Movimentação de Métodos ou Extração de Classes. Porém, existem refactorings com escopo local, que melhoraram, por exemplo, a implementação interna de um único método. Vamos a seguir descrever resumidamente alguns desses refactorings.

Extração de Variáveis é usado para simplificar expressões e torná-las mais fáceis de ler e entender. Seja o seguinte código de exemplo:

```
x1 = (-b + sqrt(b*b-4*a*c)) / (2*a);
```

Esse código pode ser refatorado para:

```
delta = b*b-4*a*c; // variável extraída
x1 = (-b + sqrt(delta)) / (2*a);
```

Veja que uma variável `delta` foi criada e inicializada com uma parte de uma expressão maior. Com isso, o código da expressão, após a refatoração, ficou menor e mais fácil de entender.

Remoção de Flags é um refactoring que sugere usar comandos como `break` ou `return`, em vez de variáveis de controle, também chamadas de flags. Seja o seguinte código de exemplo:

```
boolean search(int x, int []a) {
    boolean achou = false;
    i = 0;
    while (i < a.length) && (!achou) {
        if (a[i] == x);
            achou = true;
        i++;
    }
    return achou;
}
```

Esse código pode ser refatorado da seguinte forma:

```
boolean search(int x, int[] a) {
    for (i = 0; i < a.length; i++)
        if (a[i] == x)
            return true;
    return false;
}
```

Observe que a função após a refatoração ficou menor e com uma lógica mais clara, graças ao uso de um comando `return` para retornar imediatamente assim que um determinado valor tenha sido encontrado em um vetor.

Existe também refactorings que tratam da simplificação de comandos condicionais. Um deles é chamado de **Substituição de Condisional por Polimorfismo**. Para entendê-lo, suponha um comando `switch` que retorna o valor da bolsa de pesquisa de um aluno, dependendo do seu tipo:

```
switch (aluno.type) {
    case "graduacao":
        bolsa = 500;
        break;
    case "mestrado":
        bolsa = 1500;
        break;
    case "doutorado":
        bolsa = 2500;
        break;
}
```

Em uma linguagem orientada a objetos, esse comando pode ser refatorado para apenas uma linha de código:

```
bolsa = aluno.getBolsa();
```

Na versão refatorada, o atributo `type` de `Aluno` não é mais necessário e, portanto, pode ser removido. Além disso, temos que implementar nas subclasses de `Aluno` — por exemplo, `AlunoGraduacao`, `AlunoMestrado` e `AlunoDoutorado` — um método `getBolsa()`. Por fim, na superclasse `Aluno` esse método deve ser abstrato.

Remoção de Código Morto recomenda deletar métodos, classes, variáveis, atributos, etc que não estão sendo mais usados. Por exemplo, no caso de um método, pode não existir mais chamadas para ele. No caso de uma classe, ela pode não ser mais instanciada ou herdada por outras classes. No caso de um atributo, ele pode não ser usado no corpo da classe, nem em subclasses ou em

outras classes. Pode parecer que Remoção de Código Morto é um refactoring raro, mas em grandes sistemas, desenvolvidos ao longo de anos por programadores diferentes, costuma existir uma quantidade considerável de código que não é mais chamado.

5.3 Prática de Refactoring

Tendo apresentado diversos refactorings na seção anterior, vamos agora discutir como a prática de refactoring pode ser adotada em projetos de software.

Primeiro, a realização bem sucedida de refactorings depende da existência de bons testes, principalmente testes de unidade. Ou seja, sem testes fica arriscado realizar mudanças em um sistema, ainda mais quando elas não agregam novas funcionalidades ou corrigem bugs, como é o caso de refactorings. John Ousterhout tem o seguinte comentário sobre a importância de testes durante atividades de refactoring ([link](#)):

“Testes, particularmente testes de unidade, desempenham um papel importante no projeto de software porque eles facilitam a realização de refactorings. Sem uma suíte de testes, torna-se arriscado realizar mudanças estruturais em um sistema. Como não há uma maneira fácil de encontrar bugs, é mais provável que eles fiquem escondidos até que o código entre em produção, quando é mais caro detectá-los e corrigi-los. Como resultado, desenvolvedores tendem a evitar refatorações em sistemas sem boas suítes de teste. Em vez disso, eles reduzem as modificações no código àquelas necessárias para implementar novas funcionalidades ou corrigir bugs. Isso implica que a complexidade vai se acumulando e erros de projeto não são corrigidos.”

Uma segunda questão importante diz respeito ao momento em que o código deve ser refatorado. Existem dois modos principais de realizar refactorings: de forma oportunista ou de forma estratégica.

Refactorings oportunistas são realizados no meio de uma tarefa de programação, quando descobre-se que um trecho de código não está bem implementado e que, portanto, pode ser melhorado. Isso pode acontecer quando se está corrigindo um bug ou implementando uma nova funcionalidade. Por exemplo, no meio dessas tarefas, pode-se perceber que o nome de um método não está claro, que um método está muito grande e difícil de entender, que um comando condicional está muito complexo, que um determinado código não é mais usado, etc. Assim, se o desen-

volvedor descobrir problemas na implementação de um trecho de código, ele deve refatorá-lo imediatamente. Tentando explicar de um modo mais claro, suponha que um desenvolvedor trabalhe por uma hora na implementação de uma nova funcionalidade. É compreensível e desejável que parte desse tempo — talvez 20% ou mais — seja investido em refactorings. Kent Beck tem um frase interessante sobre refactorings oportunistas:

“Para cada mudança que você tiver que realizar em um sistema, primeiro torne essa mudança fácil (aviso: isso pode ser difícil), então realize a mudança facilmente.”

A ideia de fundo dessa recomendação é que um desenvolvedor pode estar enfrentando dificuldades para implementar uma mudança exatamente porque o código não está preparado para acomodá-la. Assim, primeiro ele deve dar um “passo atrás”, isto é, refatorar o código de forma oportunista, para tornar a mudança em questão “fácil”. Feito isso, ele terá aberto caminho para dar dois passos a frente e implementar a mudança que ficou sob sua responsabilidade.

Na maior parte das vezes, os refactorings são oportunistas. No entanto, é possível ter também **refactorings planejados**. Normalmente, eles são mudanças mais profundas, demoradas e complexas, que não valem a pena encaixar no meio de uma outra tarefa de desenvolvimento. Em vez disso, eles devem ser realizados em sessões planejadas e dedicadas. Por exemplo, esses refactorings podem envolver a quebra de um pacote em dois ou mais subpacotes, o que pode exigir atualizações em diversas partes de um sistema. Como um segundo exemplo, pode ser que o time de desenvolvimento negligenciou a prática de refactorings por muito tempo. Então, como existem muitos refactorings pendentes, é melhor planejá-los para um período de tempo específico. No entanto, como afirma Fowler “tais episódios de refactorings planejados devem ser raros. A maior parte do esforço de refactoring deve ser do tipo normal e oportunista.”

5.4 Refactorings Automatizados

Diversas IDEs oferecem suporte para automatizar a realização de refactorings, da seguinte forma: o usuário seleciona o trecho de código que pretende refatorar e a operação de refactoring que deseja realizar. Então, a IDE realiza essa operação automaticamente. Para deixar o conceito mais claro, as próximas figuras ilustram um renomeação automática de método via uma IDE. Primeiro, o usuário marca o nome do método `m1`, como mostra a figura:

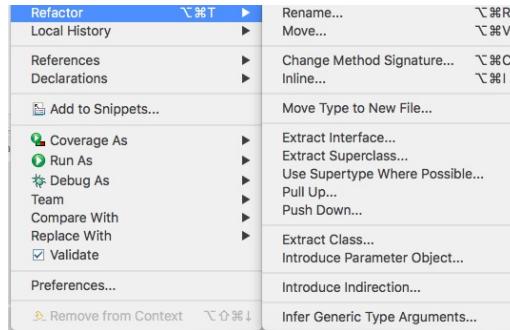
```

class A {
    void m1() {}
    void m2() {}
    void m3() { m1(); }
}

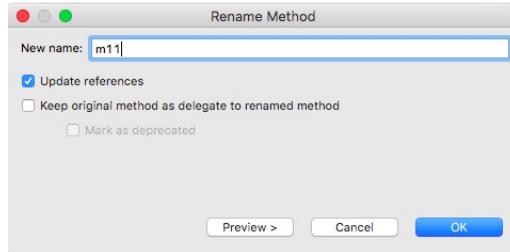
class B {
    void m4() { new A().m1(); }
}

```

Em seguida, ele escolhe as opções de `Refactor` e `Rename` da sua IDE:



Então a IDE pergunta o novo nome que ele pretende dar ao método (veja figura a seguir). Nessa mesma caixa de diálogo, o desenvolvedor informa que deseja atualizar as referências para esse método, para que elas usem o novo nome.



Feito isso, a IDE realiza o refactoring automaticamente:

```

class A {
    void m11() {}
    void m2() {}
    void m3() { m11(); }
}

class B {
    void m4() { new A().m11(); }
}

```

Primeiro, o nome do método na classe `A` foi renomeado para `m11`. Além disso, as chamadas feitas em `m3` e `m4` foram atualizadas para usar o novo nome.

Apesar de chamado de refactoring automatizado, o exemplo que mostramos deixa claro que cabe ao usuário indicar o trecho de código que deseja refatorar e o refactoring que deseja realizar. Ele também deve dar informações sobre esse refactoring. Por exemplo, o novo nome do identificador, no caso de uma renomeação. A partir daí é que o refactoring torna-se automatizado.

Antes de aplicar o refactoring, a IDE verifica se as suas **pré-condições** são verdadeiras, isto é, se a execução do refactoring — conforme requisitado pelo usuário — não vai causar um erro de compilação ou então mudar o comportamento do programa. No exemplo anterior, se o usuário pedir para renomear `m1` para `m2`, a IDE vai informar que esse refactoring não pode ser realizado, pois a classe já tem um método chamado `m2`.

5.4.1 Verificação de Pré-condições de Refactorings

A verificação das pré-condições de refactorings não é uma tarefa trivial. Para ilustrar a complexidade dessa tarefa, vamos reusar um pequeno programa em Java, proposto por Friedrich Steimann and Andreas Thies ([link](#)). Conforme mostrado a seguir, esse programa inclui duas classes, `A` e `B`, implementadas em arquivos distintos, mas pertencentes a um mesmo pacote `P1`. Assim, a chamada de `m("abc")` no primeiro arquivo irá resultar na execução do método `m(String)` da classe `B`.

```
// arquivo A.java
package P1;

public class A {
    void n() {
        new B().m("abc"); // executa m(String) de B
    }
}
```

```
// arquivo B.java
package P1;

public class B {
    public void m(Object o) {...}
    void m(String s) {...}
}
```

No entanto, suponha que a classe `B` seja movida para um novo pacote; por exemplo, para um pacote `P2`:

```
// arquivo B.java
package P2; // novo pacote de B

public class B {
    public void m(Object o) {...}
    void m(String s) {...}
}
```

Apesar de parecer inofensiva, essa Movimentação de Classe muda o comportamento do programa. Na nova versão, a chamada de `m("abc")` resulta na execução de `m(Object)` e não mais de `m(String)`. O motivo é que a classe `B` não está mais no mesmo pacote da classe `A`. Por isso, `m(String)` deixou de ser visível para `A`, pois ele não é um método público. Para leitores que não conhecem Java, um método público de uma classe pública, como `m(Object)`, pode ser chamado em qualquer parte do código. Mas métodos sem modificador de visibilidade, como `m(String)`, somente podem ser chamados por código do mesmo pacote.

Em resumo, o exemplo de Movimentação de Classes que apresentamos não é um refactoring, pois ele não preserva o comportamento do programa. Se realizado com o apoio de uma IDE, caberia a ela detectar tal fato e avisar ao usuário que o refactoring não poderá ser realizado, pois ele muda o comportamento do programa.

5.5 Code Smells

Code Smells — também conhecidos como **bad smells** — são indicadores de código de baixa qualidade, isto é, código difícil de manter, entender, modificar ou testar. Em resumo, código que não está “cheirando bem” e que portanto talvez possa ser refatorado. No entanto, nessa definição, o termo “indicadores” significa que não devemos considerar que todo code smell deve ser imediatamente refatorado. Essa decisão depende de outros fatores, como a importância do trecho de código e a frequência com que ele precisará ser mantido. Feita essa ressalva, vamos apresentar, no restante desta seção, alguns dos principais code smells.

5.5.1 Código Duplicado

Duplicação de código é o principal code smell e aquele com o maior potencial para prejudicar a evolução de um sistema. Código duplicado aumenta o esforço de manutenção, pois alterações têm que ser replicadas em mais de uma parte do código. Consequentemente, corre-se o risco de alterar uma parte e esquecer de uma outra. Código duplicado também torna a base de código mais complexa, pois dados e comandos que poderiam estar modularizados em métodos ou classes ficam espalhados pelo sistema.

Para eliminar duplicação de código, pode-se usar os seguintes refactorings: Extração de Método (recomendado quando o código duplicado está dentro de dois ou mais métodos), Extração de Classe (recomendado quando o código duplicado refere-se a atributos e métodos que aparecem em mais de uma classe) e Pull Up Method (recomendado quando o código duplicado é um método presente em duas ou mais subclasses).

Trechos de código que possuem código idêntico são chamados de **clones**. No entanto, diferentes critérios podem ser usados para definir quando dois trechos de código A e B são, de fato, idênticos. Esses critérios dão origem a quatro tipos de clones, conforme descrito a seguir:

- Clone do Tipo 1: quando A e B têm o mesmo código, com diferenças apenas em comentários e espaços.
- Clone do Tipo 2: quando A e B são clones do Tipo 1, porém as variáveis usadas em A e B podem ter nomes diferentes.
- Clone do Tipo 3: quando A e B são clones do Tipo 2, porém com pequenas mudanças em comandos.
- Clone do Tipo 4: quando A e B são semanticamente equivalentes, mas suas implementações são baseadas em algoritmos diferentes.

Exemplo: Para ilustrar esses tipos de clones, vamos usar a seguinte função:

```
int fatorial(int n) {
    fat = 1;
    for (i = 1; i <= n; i++)
        fat = fat * i;
    return fat;
}
```

A seguir, mostramos quatro clones dessa função.

- Clone Tipo 1: insere um comentário e remove espaços entre os operadores.

```
int fatorial(int n) {
    fat=1;
    for (i=1; i<=n; i++)
        fat=fat*i;
    return fat; // retorna fatorial
}
```

- Clone Tipo 2: renomeia algumas variáveis.

```
int fatorial(int n) {  
    f = 1;  
    for (j = 1; j <= n; j++)  
        f = f * j;  
    return f;  
}
```

- Clone Tipo 3: insere um comando simples, para imprimir o valor do fatorial.

```
int fatorial(int n) {  
    fat = 1;  
    for (j = 1; j <= n; j++)  
        fat = fat * j;  
    System.out.println(fat); // novo comando  
    return fat;  
}
```

- Clone Tipo 4: implementa uma versão recursiva da função.

```
int fatorial(int n) {  
    if (n == 0)  
        return 1;  
    else return n*fatorial(n-1);  
}
```

Em todos esses casos, não precisaríamos ter duas ou mais funções `fatorial` no sistema. Apenas uma delas poderia ficar no código e a outra seria removida.

Mundo Real: Em 2013, Auki Yamashita e Leon Moonen, dois pesquisadores de um laboratório de pesquisa na Noruega, publicaram os resultados de um estudo exploratório sobre code smells envolvendo 85 desenvolvedores de software ([link](#)). Quando esses desenvolvedores foram questionados sobre os code smells com os quais eles tinham mais preocupação, a resposta mais comum foi Código Duplicado, com quase 20 pontos na escala usada pelos pesquisadores para ordenar as respostas. Em segundo lugar, com metade dos pontos, ficou Métodos Longos, que é o code smell que estudaremos a seguir.

5.5.2 Métodos Longos

Em qualquer sistema, métodos devem ser pequenos, com nomes auto-explicativos e poucas linhas de código. **Métodos Longos** são considerados um code smell, pois eles tornam o código mais difícil de entender e manter. Quando nos deparamos com um método longo, devemos considerar a possibilidade de usar uma Extração de Método para quebrá-lo em métodos menores. No entanto, não existe um limite máximo de linhas de código que possa ser usado arbitrariamente para classificar métodos longos, pois isso depende da linguagem de programação, da relevância do método, do domínio do sistema, etc. Porém, modernamente, existe uma tendência em escrever métodos pequenos, com menos de 20 linhas de código, por exemplo.

5.5.3 Classes Grandes

Assim como métodos, classes não devem assumir muitas responsabilidades e prover serviços que não são coesos. Por isso, **Classes Grandes** (*Large Class*) é considerado um code smell, pois, assim como métodos longos, elas tornam o código mais difícil de entender e manter. Normalmente, é mais difícil também reusar essas classes em outro pacote ou sistema. Classes grandes são caracterizadas por um grande número de atributos, com baixa coesão entre eles. A solução para esse smell consiste em usar Extração de Classe para extrair uma classe menor A' a partir de uma classe grande A. Em seguida, a classe A passa a ter um atributo do tipo para A'.

Aprofundamento: Quando uma classe cresce tanto que ela passa a monopolizar grande parte da “inteligência” de um sistema, ela é chamada de **God Class** — ou então de **Blob**. Classes com nomes muito genéricos, como **Manager**, **System** ou **Subsystem** podem representar instâncias desse smell.

5.5.4 Feature Envy

Esse smell designa um método que parece “invejar” os dados e métodos de uma outra classe. Dizendo de outro modo, ele acessa mais atributos e métodos de uma classe B do que de sua atual classe A. Portanto, deve-se analisar a possibilidade de usar Movimentação de Método para migrá-lo para a classe “invejada”.

O método `fireAreaInvalidated2`, mostrado no código a seguir, é um exemplo de Feature Envy. Podemos observar que ele realiza três chamadas de métodos, mas todas têm como alvo um mesmo objeto `abt` do tipo `AbstractTool`. Por outro lado,

ele não acessa nenhum atributo ou chama qualquer método da sua classe atual. Logo, deve-se analisar a conveniência de mover esse método para `AbstractTool`.

```
public class DrawingEditorProxy
    extends AbstractBean implements DrawingEditor {
    ...
    void fireAreaInvalidated2 (AbstractTool abt , Double r ){
        Point p1 = abt.getView().drawingToView (...);
        Point p2 = abt.getView().drawingToView (...);
        Rectangle r = new Rectangle(p1.x,p1.y,p2.x-p1.x p2.y-p1.y)
        abt.fireAreaInvalidated (r);
    }
    ...
}
```

5.5.5 Métodos com Muitos Parâmetros

Além de pequenos, métodos, na medida do possível, devem ter poucos parâmetros. Isto é, **Métodos com Muitos Parâmetros** é um smell, que pode ser eliminado de duas formas principais. Primeiro, deve-se verificar se um dos parâmetros pode ser obtido diretamente pelo método chamado, como mostrado a seguir:

```
p2 = p1.f();
g(p1, p2);
```

Nesse caso, `p2` é desnecessário, pois ele pode ser obtido logo no início de `g`, da seguinte forma:

```
void g(p1) {
    p2 = p1.g();
}
```

Uma outra possibilidade é criar um tipo que agrupe alguns dos parâmetros de um método. Por exemplo, suponha o seguinte método:

```
void f(Date inicio, Date fim) {
    ...
}
```

Pode-se criar uma classe `DateRange` para representar uma faixa de datas. O código refatorado ficaria assim:

```
class DateRange {  
    Date inicio;  
    Date fim;  
}  
  
void f(DateRange range) {  
    ...  
}
```

5.5.6 Variáveis Globais

Conforme estudamos no capítulo sobre princípios de projeto, variáveis globais devem ser evitadas, pois elas dão origem a um tipo de **acoplamento ruim**. Por isso, elas também constituem um code smell. O principal motivo é que variáveis globais dificultam o entendimento de um módulo de forma independente dos demais módulos de um sistema. Suponha a seguinte função:

```
void f(...) {  
    // computa um determinado valor x  
    return x + g; // onde g é uma variável global.  
}
```

Apenas analisando e estudando esse código, você consegue dizer o valor que `f` retorna? A resposta é negativa, pois não basta entender o código que precede o comando `return` da função. Precisamos conhecer também o valor de `g`. Porém, como `g` é uma variável global, seu valor pode ser alterado em qualquer parte do programa. Tal situação pode facilmente introduzir bugs nessa função, pois agora uma única linha de código distante e não relacionada com `f` pode influir no seu resultado. Para isso, basta que essa linha altere o valor de `g`. Antes de concluir, em linguagens como Java, atributos estáticos de classes funcionam exatamente como variáveis globais. Logo, eles também representam um code smell.

5.5.7 Obsessão por Tipos Primitivos

Este code smell ocorre quando tipos primitivos (isto é, `int`, `float`, `String`, etc) são usados no lugar de classes. Por exemplo, suponha que precisamos declarar uma variável para armazenar o CEP de um endereço. Na pressa para usar rapidamente a variável, podemos declará-la como sendo do tipo `String`, em vez de criar uma classe dedicada — por exemplo, `CEP` — para esse fim. A principal vantagem é que uma classe pode oferecer métodos para manipular os valores que a variável vai armazenar. Por exemplo, a construtora da classe pode verificar se o CEP informado é válido antes de inicializar o objeto. Dessa forma, a classe assume essa

responsabilidade e, consequentemente, evita que ela seja uma preocupação de seus clientes. Em resumo, não devemos ficar “obcecados” com tipos primitivos. Em vez disso, devemos analisar a possibilidade de criar classes que encapsulem valores primitivos e que ofereçam operações para manipulá-los. No próximo code smell, iremos complementar essa recomendação e sugerir que tais objetos, sempre que possível, devem ser também imutáveis.

5.5.8 Objetos Mutáveis

Na segunda versão de seu livro, Fowler considera que **Objetos Mutáveis** são um code smell. Um objeto mutável é aquele cujo estado pode ser modificado. Por outro lado, um objeto imutável, uma vez criado, não pode mais ser alterado. Para viabilizar a criação de objetos imutáveis, classes devem declarar todos os seus atributos como privados e final (um atributo final somente pode ser usado para leitura). A classe também deve ser declarada final, para proibir a criação de subclasses. Se precisarmos alterar um objeto imutável, a única alternativa consiste em criar uma nova instância do objeto com o estado desejado.

Por exemplo, objetos do tipo `String` em Java são imutáveis, como ilustra o seguinte programa.

```
class Main {
    public static void main(String[] args) {
        String s1 = "Hello World";
        String s2 = s1.toUpperCase();
        System.out.println(s1);
        System.out.println(s2);
    }
}
```

Esse programa vai imprimir “Hello World” e depois “HELLO WORLD”. O motivo é que o método `toUpperCase` não altera a string `s1`, mas apenas retorna uma cópia dela com as letras em maiúsculo.

Sempre que possível devemos criar objetos imutáveis, pois eles podem ser compartilhados de forma livre e segura com outras funções. Por exemplo, em Java, você pode passar uma string como parâmetro de uma função e ter certeza de que essa função não vai mudar o seu conteúdo. Isso não ocorreria se strings fossem mutáveis, pois sempre haveria o risco de a função chamada alterar a string recebida como parâmetro. Como uma segunda vantagem, objetos imutáveis são por construção thread-safe, isto é, não é necessário sincronizar o acesso de threads aos seus métodos. O motivo é também simples: os problemas clássicos de sistemas

concorrentes, como condições de corrida, ocorrem apenas quando múltiplas threads alteram o estado de um objeto. Se esse estado for imutável, tais problemas ficam automaticamente eliminados.

No entanto, precisamos entender esse code smell no contexto da linguagem de programação usada em um sistema. Por exemplo, em linguagens funcionais, objetos são imutáveis por definição, ou seja, esse code smell nunca vai ocorrer. Por outro lado, em linguagens imperativas, é normal ter um certo número de objetos mutáveis. Ou seja, nessas linguagens, o que temos que fazer é minimizar o número de tais objetos, sem, no entanto, imaginar que vamos eliminá-los por completo. Por exemplo, devemos considerar a possibilidade de tornar imutáveis objetos simples e pequenos, como aqueles das classes `CEP`, `Moeda`, `Endereco`, `Data`, `Hora`, `Fone`, `Cor`, `Email`, etc. Para ilustrar, mostramos a seguir a implementação de uma classe `Data` imutável:

```
final public class Data { // final => não pode ter subclasses
    final private int dia; // final => inicializado uma única vez
    final private int mes;
    final private int ano;

    private void check(int dia, int mes, int ano)
        throws InvalidDateException {
        // verifica se data valida
        // se nao for, lanca InvalidDateException
    }

    public Data(int dia, int mes, int ano)
        throws InvalidDateException {
        check(dia, mes, ano);
        this.dia = dia;
        this.mes = mes;
        this.ano = ano;
    }

    // outros metodos
}
```

5.5.9 Classes de Dados

Tais classes possuem apenas atributos e nenhum método. No máximo, elas possuem getters e setters. Porém, como recorrente com code smells, não devemos considerar que **Classes de Dados** são sempre um erro de projeto. Em vez disso, o importante é analisar o código e verificar a possibilidade de mover comportamento para essas

classes. Isto é, criar métodos nessas classes, para realizar operações que já estão sendo realizadas, mas de forma espalhada em outras classes.

5.5.10 Comentários

Pode soar estranho ver comentários incluídos em uma lista de code smells. Por exemplo, em cursos de Introdução a Programação, os alunos são incentivados a comentar todo o código produzido, com o objetivo de ensinar a importância de documentação de código. No livro *Elements of Programming Style*, Brian Kerninghan — um dos criadores das primeiras versões do sistema operacional Unix e da linguagem de programação C — e P. J. Plauger, dão uma recomendação que ajuda, de forma certeira, a esclarecer essa dúvida. Eles recomendam o seguinte:

“Não comente código ruim, reescreva-o”.

A ideia é que comentários não devem ser usados para explicar código ruim. Em vez disso, deve-se refatorar o código e, com isso, melhorar sua qualidade e legibilidade. Feito isso, existe uma boa chance de que o comentário não seja mais necessário. Um exemplo são métodos longos como aquele do programa a seguir.

```
void f() {  
    // task1  
    ...  
    // task2  
    ...  
    // taskn  
    ...  
}
```

Se usarmos Extração de Método para extrair o código comentado, teremos o código a seguir de melhor qualidade:

```
void task1 { ... }  
void task2 { ... }  
void taskn { ... }  
  
void f {  
    task1();  
    task2();  
    ...  
    taskn();  
}
```

Observe que, no método `f`, após a refatoração, comentários não são mais necessários, pois os nomes dos métodos chamados já revelam muito do que eles fazem.

Bibliografia

Martin Fowler. Improving the Design of Existing Code, Addison-Wesley, 2000.

Martin Fowler. Improving the Design of Existing Code, Addison-Wesley, 2018.

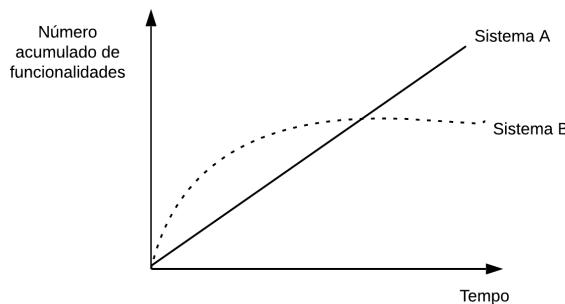
Danilo Silva, Nikolaos Tsantalis, Marco Túlio Valente. Why We Refactor? Confessions of GitHub Contributors. Foundations of Software Engineering (FSE), 2016.

Exercícios de Fixação

1. Marque a alternativa FALSA:

- (a) refactoring melhora o projeto de um sistema de software.
- (b) refactoring torna o código de um sistema mais fácil de ser entendido.
- (c) refactoring facilita a localização e a correção de bugs futuros.
- (d) refactoring acelera a implementação de novas funcionalidades.
- (e) refactoring melhora o desempenho de um sistema.

2. O gráfico a seguir mostra o total acumulado de novas funcionalidades implementadas em dois sistemas (A e B), de domínios semelhantes, desenvolvidos por times semelhantes, usando as mesmas tecnologias. Em qual dos dois sistemas você acha que refactorings foram realizados de forma sistemática? Justifique.



3. Descreva as diferenças entre refactorings oportunistas e refactorings planejados. Qual dessas formas de refactoring deve ser mais comum?

4. Escreva o nome de refactorings A e B que se executados em sequência não produzem impacto no código de um sistema. Ou seja, o refactoring B reverte as transformações realizadas pelo refactoring A.

(a) Cite um refactoring A e um refactoring B.

(b) Cite um refactoring A e um refactoring B.

5. Em cada um dos exemplos a seguir, extraia o código em negrito (e apenas ele) para um método g.

(a)

```
class A {
    void f() {
        int x = 10
        x++;
        print x;
    }
}
```

(b)

```
class A {
    void f() {
        int x = 10
        x++;
        print x;
    }
}
```

(c)

```
class A {
    void f() {
        int x = 10
        x++;
        print x;
        int y = x+1;
        ...
    }
}
```

(d)

```
class A {
    void f() {
        int x = 10
        int y;
        y = h()*2;
        print y;
        int z = y+1;
        ...
    }
}
```

6. A seguinte função calcula o n-ésimo termo da sequência de Fibonacci. O primeiro termo dessa sequência é 0; o segundo termo é 1; e a partir daí o n-ésimo termo é a soma dos dois termos anteriores.

```
int fib(int n) {
    if (n == 1)
        return 0;
    if (n == 2)
        return 1;
    return fib(n-1) + fib(n-2);
}
```

Descreva clones dos Tipos 1, 2, 3 e 4 para essa função. Não precisa implementar o código do clone, mas seja bem preciso na sua resposta ao indicar as diferenças entre cada clone e o código acima.

7. Seja o seguinte trecho de código de uma classe `Moeda`, que vai armazenar um determinado valor em reais.

```
class Moeda {
    ...
    private double valor = 0.0;
    void add(Moeda m) {
        this.valor = this.valor + m.getValor();
    }
    ...
}
```

- (a) Descreva porque objetos da classe `Moeda` são mutáveis.

(b) Reimplemente esse trecho da classe `Moeda` de forma a assegurar que os seus objetos sejam imutáveis (como vimos no capítulo, objetos mutáveis tendem a ser um code smell, principalmente no caso de objetos pequenos e simples, como provavelmente é o caso de objetos da classe em questão).

8. Como discutido no final da Seção 9.5, comentários que são usados para explicar “código ruim” são considerados um code smell. Nessas situações, o ideal é tornar o código mais claro e, então, remover os comentários. A seguir, mostramos mais um caso de comentário que pode ser deletado. Explique porque esses comentários são desnecessários.

```
// classe Aluno
class Aluno {

    // matrícula do aluno
    int matricula;

    // data de nascimento do aluno
    Date dataNascimento;

    // endereço do aluno
    Endereco endereco;

    // construtor default da classe Aluno
    Aluno() {
        ...
    }
    ...
}
```

9. Use uma IDE, como o Eclipse ou IntelliJ para Java, para realizar um refactoring simples, em um de seus programas. Por exemplo, realize uma renomeação de método. Quais as vantagens de se realizar refactorings com o suporte de IDEs?

10. Use uma IDE para testar o exemplo de Movimentação de Classe discutido no parágrafo de aprofundamento da Seção 9.4.1 (isto é, o exemplo com classes A e B e pacotes P1 e P2). Se realizar a Movimentação de Classe discutida nesse exemplo via IDE, ocorrerá algum erro? Se sim, descreva o erro detectado pela IDE.

Capítulo 6

DevOps

Imagine a world where product owners, development, QA, IT Operations, and Infosec work together, not only to help each other, but also to ensure that the overall organization succeeds. – G. Kim, J. Humble, P. Debois, J. Willes

6.1 Introdução

Até agora, neste livro, estudamos um conjunto de práticas para desenvolvimento de software com qualidade e agilidade. Por meio de métodos ágeis — como Scrum, XP ou Kanban —, vimos que o cliente deve participar desde o primeiro dia da construção de um sistema. Também estudamos práticas importantes para produção de software com qualidade, como testes de unidade e refactoring. Estudamos ainda princípios e padrões de projeto e também padrões arquiteturais.

Logo, após aplicar o que vimos, o sistema — ou um incremento dele, resultante de um sprint — está pronto para entrar em produção. Essa tarefa é conhecida pelos nomes de **implantação (deploy)**, **liberação (release)** ou **entrega (delivery)** do sistema. Independentemente do nome, ela não é tão simples e rápida como pode parecer.

Historicamente, em organizações tradicionais, a área de Tecnologia da Informação costumava ser dividida em dois departamentos:

- Departamento de Sistemas (ou Desenvolvimento), formado por desenvolvedores, programadores, analistas, arquitetos, etc.

- Departamento de Suporte (ou Operações), no qual ficavam alocados os administradores de rede, administradores de bancos de dados, técnicos de suporte, etc.

Hoje em dia, é fácil imaginar os problemas causados por essa divisão. Na maioria das vezes, a área de suporte tomava conhecimento de um sistema na véspera da sua implantação. Consequentemente, a implantação poderia atrasar por meses, devido a uma variedade de problemas que não foram identificados. Dentre eles, podemos citar a falta de hardware para executar o novo sistema ou a nova funcionalidade, problemas de desempenho, incompatibilidades com o banco de dados de produção, vulnerabilidades de segurança, etc. No limite, esses problemas poderiam resultar no cancelamento da implantação e no abandono do sistema.

Resumindo, nesse modelo tradicional, existia um stakeholder importante — os administradores de sistemas ou *sysadmins* — que tomava conhecimento das características e requisitos não-funcionais de um novo software na véspera da implantação. Esse problema era agravado pelo fato de os sistemas serem grandes monolitos, cuja implantação gerava todo tipo de preocupação, como mencionado no final do parágrafo anterior.

Então, para facilitar a implantação e entrega de sistemas, foi proposto o conceito de **DevOps**. Por ser um termo recente, ele ainda não possui uma definição consolidada. Mas seus proponentes gostam de descrever DevOps como um movimento que visa unificar as culturas de desenvolvimento (Dev) e de operação (Ops), visando permitir a implantação mais rápida e ágil de um sistema. Esse objetivo está refletido na frase que abre esse capítulo, de autoria de Gene Kim, Jez Humble, Patrick Debois e John Willes, todos eles membros de um grupo de desenvolvedores que ajudou a difundir os princípios de DevOps. Segundo eles, DevOps implica na seguinte disruptão na cultura tradicional de implantação de sistemas ([link](#)):

“Em vez de iniciar as implantações à meia-noite de sexta-feira e passar todo o fim de semana trabalhando para concluir-las, as implantações ocorrem em qualquer dia útil, quando todos estão na empresa e sem que os clientes percebam — exceto quando encontram novas funcionalidades e correções de bugs.”

No entanto, DevOps não advoga a criação de um profissional novo, que fique responsável tanto pelo desenvolvimento como pela implantação de sistemas. Em vez disso, defende-se uma aproximação entre o pessoal de desenvolvimento e o pessoal de operações e vice-versa, visando fazer com que a implantação de sistemas

seja mais ágil e menos traumática. Tentando explicar com outras palavras, a ideia é evitar dois silos independentes: desenvolvedores e operadores, com pouca ou nenhuma interação entre eles, como ilustrado na figura a seguir.

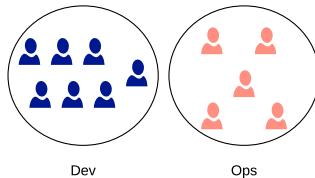


Figura 6.1: Organização que **não** é baseada em DevOps. Existe pouca comunicação entre Dev e Ops.

Em vez disso, defende-se que esses profissionais atuem em conjunto desde os primeiros sprints de um projeto, como na figura a seguir. Para o cliente final, o benefício deve ser a entrada em produção mais cedo do sistema que ele contratou.

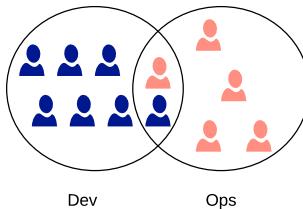


Figura 6.2: Organização baseada em DevOps. Frequentemente, alguns Dev e alguns Ops sentam juntos para discutir questões sobre a entrega do sistema.

Quando migra-se para uma cultura de DevOps, os times ágeis podem incluir um profissional de operações, que participe dos trabalhos do time em tempo parcial ou mesmo em tempo integral. Sempre em função da demanda, esse profissional pode também participar de mais de um time. A ideia é que ele antecipe problemas de desempenho, segurança, incompatibilidades com outros sistemas, etc. Ele pode também, enquanto o código está sendo implementado, começar a trabalhar nos scripts de instalação, administração e monitoramento do sistema em produção.

De forma não menos importante, DevOps advoga ainda a automatização de todos os passos necessários para colocar um sistema em produção e monitorar o seu correto funcionamento. Isso implica na adoção de práticas que já vimos neste livro,

notadamente testes automatizados. Mas também implica no emprego de novas práticas e ferramentas, tais como Integração Contínua (*Continuous Integration*) e Entrega Contínua (*Continuous Deployment*), que iremos estudar neste capítulo.

Para finalizar, vamos discutir um conjunto de princípios para entrega de software, enunciados por Jez Humble e David Harley ([link](#)). Apesar de propostos antes da ideia de DevOps ganhar tração, eles são completamente alinhados com essa ideia. Alguns desses princípios são os seguintes:

- **Crie um processo repetível e confiável para entrega de software.** Esse princípio é o mais importante deles. A ideia é que a entrega de software não pode ser um evento traumático, com passos manuais e sujeitos a surpresas. Em vez disso, colocar um software em produção deve ser tão simples como apertar um botão.
- **Automatize tudo que for possível.** Na verdade, esse princípio é um pré-requisito indispensável para atender ao princípio anterior. Advoga-se que todos os passos para entrega de um software devem ser automáticos, incluindo seu *build*, a execução dos testes, a configuração e ativação dos servidores e da rede, a carga do banco de dados, etc. De novo, idealmente, queremos apertar um botão e, em seguida, ver o sistema em produção.
- **Mantenha tudo em um sistema de controle de versões.** “Tudo” no enunciado do princípio refere-se não apenas a todo o código fonte, mas também arquivos e scripts de administração do sistema, documentação, páginas Web, arquivos de dados, etc. Consequentemente, deve ser simples restaurar e voltar o sistema para um estado anterior. Neste capítulo, iniciaremos estudando alguns conceitos básicos de **controle de versões**, na Seção 10.2. Além disso, no Apêndice A apresentamos e ilustramos o uso dos principais comandos do sistema Git, que é o sistema de controle de versões mais usado atualmente.
- **Se um passo causa dor, execute-o com mais frequência e o quanto antes.** Esse princípio não tem uma inspiração masoquista. Em vez disso, a ideia é antecipar os problemas, antes que eles se acumulem e as soluções fiquem complicadas. O exemplo clássico é o de **integração contínua**. Se um desenvolvedor passa muito tempo trabalhando de forma isolada, ele e o seu time podem depois ter uma grande dor de cabeça para integrar o código. Logo, como integração pode causar dor, a recomendação consiste em integrar código novo com mais frequência e o quanto antes, se possível, diariamente. Iremos estudar mais sobre integração contínua na Seção 10.3.

- “**Concluído**” significa pronto para entrega. Com frequência, desenvolvedores dizem que uma nova história está pronta (*done*). Porém, ao serem questionados se ela pode entrar em produção, começam a surgir “pequenas” pendências, tais como: a implementação ainda não foi testada com dados reais, ela ainda não foi documentada, ela ainda não foi integrada com o sistema X, etc. Esse princípio defende então que “concluído”, em projetos de software, deve ter uma semântica clara, isto é: 100% pronto para entrar em produção.
- **Todos são responsáveis pela entrega do software.** Esse último princípio alinha-se perfeitamente com a cultura de DevOps que discutimos no início desta Introdução. Ou seja, não admite-se mais que os times de desenvolvimento e operação trabalham em silos independentes e troquem informações apenas na véspera de uma implantação.

Mundo Real: O termo DevOps começou a ser usado no final dos anos 2000 por desenvolvedores frustrados com os atritos constantes entre as equipes de desenvolvimento e operações. Então, eles convenceram-se de que uma solução seria a adoção de princípios ágeis não apenas na fase de desenvolvimento, mas também de implantação. Para citar uma data precisa, em Novembro de 2009 foi realizada, na Bélgica, a primeira conferência da indústria sobre o tema, chamada DevOpsDay. Considera-se que foi nesta conferência, organizada por Patrick Dubois, que a palavra DevOps foi cunhada ([link](#)).

6.2 Controle de Versões

Como mencionamos algumas vezes neste livro, software é desenvolvido em equipe. Por isso, precisamos de um servidor para armazenar o código fonte do sistema que está sendo implementado por um grupo de desenvolvedores. A existência desse servidor é fundamental para que esses desenvolvedores possam colaborar e para que os operadores saibam precisamente qual versão do sistema deve ser colocada em produção. Além disso, sempre é útil manter o histórico das versões mais importantes de cada arquivo. Isso permite, se necessário, realizar uma espécie de “undo” no tempo, isto é, recuperar o código de um arquivo como ele estava há anos atrás, por exemplo.

Um sistema de controle de versões (VCS, na sigla em inglês) oferece os dois serviços mencionados no parágrafo anterior. Primeiro, ele oferece um **repositório** para armazenar a versão mais recente do código fonte de um sistema, bem como

de arquivos relacionados, como arquivos de documentação, configuração, páginas Web, manuais, etc. segundo lugar, ele permite que se recupere versões mais antigas de qualquer arquivo, caso isso seja necessário. Como enunciamos na Introdução, modernamente é inconcebível desenvolver qualquer sistema, mesmo que simples, sem um VCS.

Os primeiros sistemas de controle de versões surgiram no início da década de 70, como o sistema SCCS, desenvolvido para o sistema operacional Unix. Em seguida, surgiram outros sistemas, como o CVS, em meados da década de 80, e depois o sistema Subversion, também conhecido pela sigla svn, no início dos anos 2000. Todos são sistemas centralizados e baseados em uma arquitetura cliente/servidor (veja figura a seguir). Nessa arquitetura, existe um único servidor, que armazena o repositório e o sistema de controle de versões. Os clientes acessam esse servidor para obter a versão mais recente de um arquivo. Feito isso, eles podem modificar o arquivo, por exemplo, para corrigir um bug ou implementar uma nova funcionalidade. Por fim, eles atualizam o arquivo no servidor, realizando uma operação chamada **commit**, a qual torna o arquivo visível para os outros desenvolvedores.

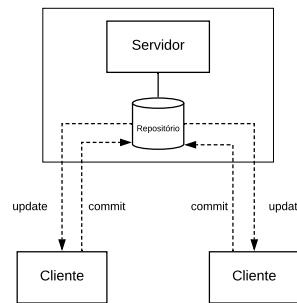


Figura 6.3: VCS Centralizado. Existe um único repositório, no nodo servidor

No início dos anos 2000, começaram a surgir **sistemas de controle de versões distribuídos (DVCS)**. Dentre eles, podemos citar o sistema BitKeeper, cujo primeiro release é de 2000, e os sistemas Mercurial e git, ambos lançados em 2005. Em vez de uma arquitetura cliente/servidor, um DVCS adota uma arquitetura peer-to-peer. Na prática, isso significa que cada desenvolvedor possui em sua máquina um servidor completo de controle de versões, que pode se comunicar com os servidores de outras máquinas, como ilustrado na próxima figura.

Apesar de todos os clientes serem funcionalmente equivalentes, na prática, quando

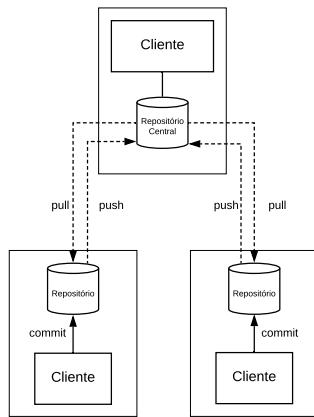


Figura 6.4: VCS Distribuído (DVCS). Cada cliente possui um servidor. Logo, a arquitetura é peer-to-peer.

se usa um DVCS, existe uma máquina principal, que armazena a versão de referência do código fonte. Na nossa figura, chamamos esse repositório de **repositório central**. Cada desenvolvedor pode trabalhar de forma independente e até mesmo offline em sua máquina cliente, realizando commits no seu repositório. De tempos em tempos, ele deve sincronizar esse repositório com o central, por meio de duas operações: **pull** e **push**. Um pull atualiza o repositório local com novos commits disponíveis no repositório central. Por sua vez, um push faz a operação contrária, isto é, envia para o repositório central os commits mais recentes realizados pelo desenvolvedor em seu repositório local. Quando comparado com um VCS centralizado, um DVCS tem as seguintes vantagens:

- Pode-se trabalhar e gerenciar versões de forma offline, sem estar conectado a uma rede, pois os commits são realizados primeiro no repositório local.
- Pode-se realizar commits com mais frequência, incluindo commits com implementações parciais, pois eles não vão chegar imediatamente até o repositório central.
- Commits são executados em menos tempo, isto é, são operações mais rápidas e leves. O motivo é que eles são realizados no repositório local de cada máquina.
- A sincronização não precisa ser sempre com o repositório central. Em vez disso, dois nodos podem também sincronizar os seus repositórios. Por

exemplo, pode-se ter uma estrutura hierárquica dos repositórios. Nesses casos, os commits "nascem" nos repositórios que representam as folhas da hierarquia e vão subindo até chegar ao repositório central.

Git é um sistema de controle de versões distribuído cujo desenvolvimento foi liderado por Linus Torvalds, também responsável pela criação do sistema operacional Linux. Nos anos iniciais, o desenvolvimento do kernel do Linux usava um sistema de controle de versões comercial, chamado BitKeeper, que também possui uma arquitetura distribuída. No entanto, em 2005, a empresa proprietária do BitKeeper resolveu revogar as licenças gratuitas que eram usadas no desenvolvimento do Linux. Os desenvolvedores do sistema operacional, liderados por Torvalds, decidiram então iniciar a implementação de um DVCS próprio, ao qual deram o nome de Git. Assim como o Linux, o Git é um sistema de código aberto, que pode ser gratuitamente instalado em qualquer máquina. O Git é ainda um sistema de linha de comando. Porém, existem também algumas interfaces gráficas, produzidas por outras empresas, que permitem usar o sistema sem ter que digitar comandos.

GitHub é um serviço de hospedagem de código que usa o sistema git para prover controle de versões. O GitHub oferece repositórios públicos e gratuitos, para projetos de código aberto, e repositórios fechados e pagos, para uso por empresas. Assim, em vez de manter internamente um DVCS, uma empresa desenvolvedora de software pode alugar esse serviço do GitHub. Uma comparação pode ser feita com serviços de mail. Em vez de instalar um servidor de mail em uma máquina própria, uma empresa pode contratar esse serviço de terceiros, como do Google, via GMail. Apesar de o GitHub ser o mais popular, existem serviços semelhantes providos por outras empresas, como GitLab e BitBucket.

No Apêndice A, apresentamos e ilustramos os principais comandos do sistema Git. São explicados também os conceitos de forks e pull requests, os quais são específicos do GitHub.

Multirepos vs Monorepos

Um VCS gerencia repositórios. Assim, uma organização precisa decidir os repositórios que vai criar em seu VCS. Uma decisão tradicional consiste em criar um repositório para cada projeto ou sistema da organização. Porém, soluções baseadas em um único repositório estão sendo adotadas com mais frequência, principalmente entre grandes empresas, como Google, Facebook e Microsoft. Essas duas alternativas — chamadas, respectivamente, de **multirepos** e **monorepo** — são ilustradas nas próximas figuras.

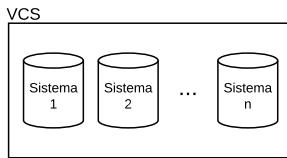


Figura 6.5: Multirepos: VCS gerencia vários repositórios (um repositório por projeto ou sistema)

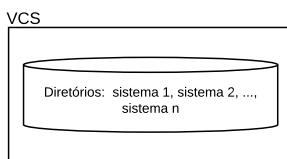


Figura 6.6: Monorepo: VCS gerencia um único repositório. Projetos são diretórios desse repositório.

Se pensarmos em contas do GitHub, podemos exemplificar da seguinte forma:

- Se optar por multirepos, uma organização terá vários repositórios, tais como `aserg-ufmg/sistema1`, `aserg-ufmg/sistema2`, `aserg-ufmg/sistema3`, etc.
- Se optar por monorepos, ela terá um único repositório — digamos, `aserg-ufmg/monorepo`. No diretório raiz desse repositório, teremos subdiretórios como os seguintes: `sistema1`, `sistema2`, `sistema3`, etc.

Dentre as vantagens de monrepos podemos citar:

- Como existe um único repositório, não há dúvida sobre em qual repositório encontra-se a versão mais atualizada de um arquivo. Isto é, com monrepos, existe uma única fonte de “verdade” sobre versões do código fonte.
- Monrepos incentivam o reuso e compartilhamento de código, pois os desenvolvedores têm acesso mais rápido a qualquer arquivo, de qualquer sistema.
- Mudanças são sempre atômicas. Com multirepos, dois commits podem ser necessários para implementar uma única mudança, caso ela afete dois sistemas. Com monrepos, a mesma mudança pode ser realizada por meio de um único commit.

- Facilita a execução de refactorings em larga escala. Por exemplo, suponha a renomeação de uma função de uma API que é usada em todos os sistemas da organização. Com monorepos, essa renomeação pode ser realizada em um único commit.

Por outro lado, monorepos demandam o uso de ferramentas que ajudem a navegar em grandes bases de código. Por exemplo, caso use git, o desenvolvedor terá em seu repositório local todos os arquivos de todos os sistemas da organização. Por isso, os responsáveis pelo monorepo do Google comentam que foram obrigados a implementar internamente um plug-in para a IDE Eclipse, o qual facilita o trabalho com uma base de código muito grande, como a que eles possuem na empresa ([link](#)).

6.3 Integração Contínua

Para explicar o conceito de Integração Contínua (CI), iniciamos com uma subseção de motivação. Em seguida, apresentamos o conceito propriamente dito. Feito isso, discutimos outras práticas que uma organização deve adotar junto com CI. Terminamos com uma breve discussão sobre cenários que podem desmotivar o emprego de CI em uma organização.

Motivação

Antes de definir o que é integração contínua, vamos descrever o problema que levou à proposta dessa prática de integração de código. Tradicionalmente, era comum o uso de branches durante a implementação de novas funcionalidades. Branches podem ser entendidos como um sub-diretório interno e virtual, gerenciado pelo sistema de controle de versões. Nesses sistemas, existe um branch principal, conhecido pelo nome de **master** (quando usa-se Git) ou **trunk** (quando usa-se outros sistemas, como svn). Além do branch principal, os usuários podem criar seus próprios branches.

Por exemplo, antes de implementar uma nova funcionalidade, era comum criar um branch apenas para conter o seu código. Tais branches são chamados de **branches de funcionalidades (feature branches)** e, dependendo da complexidade da funcionalidade, eles podiam levar meses para serem integrados de volta à linha principal de desenvolvimento. Logo, era comum em sistemas maiores e mais complexos existirem dezenas de branches ativos.

Quando a implementação da nova funcionalidade terminava, o código do branch era “copiado” de volta para o master, por meio de um comando do sistema de controle de versões chamado **merge**. Nesse momento, uma variedade de conflitos poderia

ocorrer, os quais são conhecidos como **conflitos de integração** ou **conflitos de merge**.

Para ilustrar esses conflitos, suponha que Alice criou um branch para implementar uma nova funcionalidade X em seu sistema. Como essa funcionalidade era complexa, Alice trabalhou de forma isolada no seu branch por 40 dias, conforme ilustra a figura a seguir (cada nodo desse grafo é um commit). Observe que enquanto Alice trabalhava — realizando commits em seu branch — também ocorriam commits no branch principal.

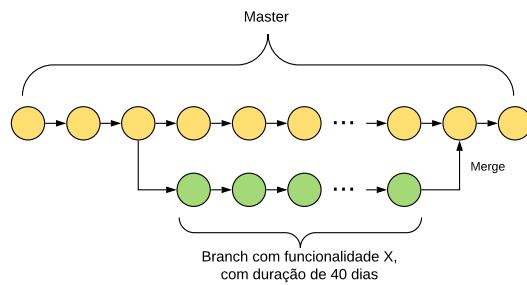


Figura 6.7: Desenvolvimento usando branches de funcionalidades.

Então, após 40 dias, quando Alice integrhou seu código no master, surgiram diversos conflitos. Alguns deles são descritos a seguir:

- Para implementar a funcionalidade X, o novo código desenvolvido por Alice chamava uma função `f1`, que existia no master no momento da criação do branch. Porém, no intervalo de 40 dias, a assinatura dessa função foi modificada no master por outros desenvolvedores. Por exemplo, a função pode ter sido renomeada ou ter ganho um novo parâmetro. Ou ainda, em um cenário mais radical, `f1` pode ter sido removida da linha principal de desenvolvimento.
- Para implementar a funcionalidade X, Alice mudou o comportamento de uma função `f2` do master. Por exemplo, `f2` retornava seu resultado em milhas e Alice alterou o seu código para que o resultado fosse retornado em quilômetros. Evidentemente, Alice atualizou todo o código que chamava `f2` no seu branch, para considerar resultados em quilômetros. Porém, no período de 40 dias, surgiram novas chamadas de `f2`, que foram integradas no master, mas supondo um resultado ainda em milhas.

Em sistemas grandes, com milhares de arquivos, dezenas de desenvolvedores e de

branches funcionais, os problemas causados por conflitos podem assumir proporções consideráveis e atrasar a entrada em produção de novas funcionalidades. Veja que a resolução de conflitos é uma tarefa manual, que requer análise e consenso entre os desenvolvedores envolvidos. Por isso, o termo **integration hell** é frequentemente usado para descrever os problemas que ocorrem durante a integração de branches de funcionalidades.

Adicionalmente, branches de funcionalidades, principalmente aqueles com duração longa, ajudam a criar silos de conhecimento. Isto é, cada nova funcionalidade passa a ter um dono, pois um desenvolvedor ficou dedicado a ela por semanas. Assim, esse desenvolvedor pode sentir-se mais confortável para adotar padrões diferentes do restante do time, incluindo padrões para leiaute do código, para organização de janelas e telas, para acesso a dados, etc.

O que é Integração Contínua?

Integração contínua (*continuous integration* ou CI) é uma prática de desenvolvimento proposta por Extreme Programming (XP), conforme estudamos no Capítulo 2. O princípio motivador da prática já foi enunciado na Introdução deste capítulo: se uma tarefa causa “dor”, não podemos deixar que ela acumule. Em vez disso, devemos quebrá-la em subtarefas que possam ser realizadas de forma frequente. Como essas subtarefas serão pequenas e simples, a “dor” decorrente da sua realização será menor.

Adaptando para o contexto de integração de código, sabemos que grandes integrações são uma fonte de “dor” para os desenvolvedores, pois eles têm que resolver de forma manual diversos conflitos. Assim, CI recomenda integrar o código de forma frequente, isto é, contínua. Como isso, as integrações serão pequenas e irão gerar menos conflitos.

Kent Beck, em seu livro de XP, defende o uso de CI da seguinte forma ([link](#)):

“Você deve integrar e testar o seu código em intervalos menores do que algumas horas. Programação em times não é um problema do tipo dividir-e-conquistar. Na verdade, é um problema que requer dividir, conquistar e integrar. A duração de uma tarefa de integração é algo imprevisível e pode facilmente levar mais tempo do que a tarefa original de codificação. Assim, quanto mais tempo você levar para integrar, maiores e mais imprevisíveis serão os custos.”

Nessa citação, Beck defende várias integrações ao longo de um dia de trabalho de um

desenvolvedor. No entanto, essa recomendação não é consensual. Outros autores, como Fowler, mencionam pelo menos uma integração por dia por desenvolvedor ([link](#)), o que parece ser um limite mínimo para um time argumentar que está usando CI.

Boas Práticas para Uso de CI

Quando usa-se CI, o master é constantemente atualizado com código novo. Para garantir que ele não seja quebrado — isto é, deixe de compilar ou possua bugs —, recomenda-se o uso de algumas práticas junto com CI. Vamos discutir algumas delas a seguir.

Build Automatizado

Build é um nome usado para designar a compilação de todo os arquivos de um sistema, até a geração uma versão executável. Quando se usa CI, o build deve ser automatizado, isto é, não incluir nenhum passo manual. Além disso, é importante que ele seja o mais rápido possível, pois com integração contínua ele será sempre executado. Alguns autores, por exemplo, chegam a recomendar um limite de 10 minutos para execução de um build ([link](#)).

Testes Automatizados

Além de garantir que o sistema compila sem erros após cada novo commit, é importante garantir também que ele continua com o comportamento esperado. Por isso, ao usar CI, deve-se ter uma boa cobertura de testes, principalmente testes de unidade, conforme estudamos no Capítulo 8.

Servidores de Integração Contínua

Por fim, não basta ter builds e testes automatizados. É importante que eles sejam executados com frequência, se possível após cada novo commit realizado no master. Para isso, existem Servidores de CI, que funcionam da seguinte forma (acompanhe também pela próxima figura):

- Após um novo commit, o sistema de controle de versões avisa o servidor de CI, que clona o repositório e executa então um build completo do sistema, bem como roda todos os testes.
- Após a execução do build e dos testes, o servidor de CI notifica o usuário.

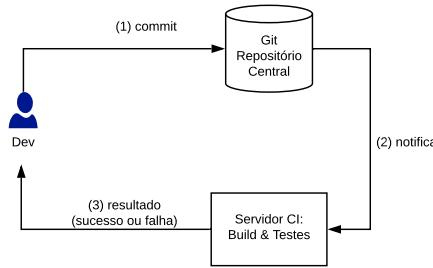


Figura 6.8: Servidor de Integração Contínua

O objetivo principal de um servidor de integração contínua é evitar a integração de código com problemas, sejam eles de build ou de comportamento. Quando o build falha, costuma-se dizer que ele “quebrou”. Com frequência, o build na máquina do desenvolvedor pode ter sido concluído com sucesso. Mas ao ser executado no servidor de CI, ele pode falhar. Isso ocorre, por exemplo, quando o desenvolvedor esquece de realizar o commit de algum arquivo. Dependências incorretas são outro motivo para quebra de builds. Por exemplo, o código foi compilado e testado na máquina local do desenvolvedor usando a versão 2.0 de uma determinada biblioteca, mas o servidor de CI realiza o build usando a versão 1.0.

Se o servidor de CI notificar o desenvolvedor de que seu código não passou nos testes ou quebrou o build, ele deve parar tudo que está fazendo e providenciar a correção. Isso é importante porque um build quebrado impacta o trabalho dos outros desenvolvedores, pois eles não vão conseguir compilar ou executar o código. Costuma-se dizer que nada em uma empresa de software tem maior prioridade do que a correção de um build quebrado. No entanto, a solução pode ser simplesmente reverter o código para a versão anterior ao commit com problemas.

Ainda nesta linha de raciocínio, um desenvolvedor somente deve avançar para uma próxima tarefa de programação após receber o resultado do servidor de CI. Por exemplo, ele não deve começar a escrever código novo, antes de ter certeza de que seu último commit passou pelo serviço de integração contínua. Ele não deve também iniciar outras tarefas importantes, como entrar em uma reunião, sair para almoçar ou ir para a casa, antes do resultado do servidor de CI.

Existem diversos servidores de integração contínua no mercado. Alguns deles são oferecidos como um serviço independente, normalmente gratuito para repositórios de código aberto, mas pago para repositórios privados de empresas. Assim, se

você possui um repositório aberto no GitHub, existe mais de uma opção gratuita para ativar um serviço de CI no mesmo.

Uma dúvida comum é se CI é compatível com o uso de branches. De forma coerente com a definição de CI, a melhor resposta é a seguinte: sim, desde que os branches sejam integrados de forma frequente no master, via de regra, todo dia. Dizendo de outra forma, CI não é incompatível com branches, mas apenas com branches com um tempo de vida elevado. Ainda nessa linha, Martin Fowler tem a seguinte observação sobre o uso de branches, especificamente branches de funcionalidades, junto com CI ([link](#)):

“Na maioria das vezes, branches de funcionalidades constituem uma abordagem incompatível com CI. Um dos princípios de CI é que todos devem enviar commits para a linha de desenvolvimento principal diariamente. Então, a não ser que os branches de funcionalidades durem menos do que um dia, eles são um animal diferente de CI. É comum ouvir desenvolvedores dizendo que eles estão usando CI porque eles rodam builds automáticos, talvez usando um servidor de CI, após cada commit. Isso pode ser chamado de building contínuo e pode ser uma coisa boa. Porém, como não há integração, não podemos chamar essa prática de CI.”

Desenvolvimento Baseado no Trunk

Como vimos, ao adotar CI, os branches devem durar no máximo um dia de trabalho. Logo, o custo/benefício de criá-los pode não compensar. Por isso, quando migram para CI, é comum que as organizações usem também **desenvolvimento baseado no trunk** (*trunk based development* ou TBD). Quando isso ocorre, não existem mais branches para implementação de novas funcionalidades ou para correção de bugs. Em vez disso, todo desenvolvimento ocorre no branch principal, isto é, no trunk ou master.

Mundo Real: TBD é usado por grandes empresas desenvolvedoras de software, incluindo Google e Facebook:

- No Google, “quase todo desenvolvimento ocorre no HEAD do repositório [isto é, no master]. Isso ajuda a identificar problemas de integração mais cedo e minimiza o esforço para realização de merges.” ([link](#))
- No Facebook, “todos engenheiros de front-end trabalham em um único branch que é mantido sempre estável, o que também torna o desen-

volvimento mais rápido, pois não dispende-se esforço na integração de branches de longa duração no trunk.” ([link](#))

Programação Pareada

Programação pareada pode ser entendida como uma forma contínua de revisão de código. Quando adota-se essa prática, qualquer novo trecho de código é revisado por um outro desenvolvedor, que encontra-se sentado ao lado do desenvolvedor líder da sessão de programação. Portanto, assim como builds e testes contínuos, recomenda-se usar programação pareada com CI. Porém, tal uso também não é obrigatório. Por exemplo, o código pode ser revisado após o commit ser realizado no master. No entanto, nesse caso, como o código já foi integrado, os custos de aplicar a revisão podem ser maiores.

Quando não usar CI?

Os proponentes de CI definem um limite rígido para integrações no master: pelo menos uma integração por dia por desenvolvedor. No entanto, dependendo da organização, do domínio do sistema (que pode ser um sistema crítico) e do perfil dos desenvolvedores (que podem ser iniciantes), pode ser difícil aplicar esse limite.

E também é preciso lembrar que esse limite não é uma lei da natureza. Por exemplo, talvez seja mais factível realizar uma integração a cada dois ou três dias. Na verdade, qualquer prática de Engenharia de Software — incluindo integração contínua — não deve ser considerada ao pé da letra, isto é, exatamente como está descrita no manual ou no livro texto. Adaptações justificadas pelo contexto da organização são sempre possíveis e devem ser consideradas. Experimentação com diferentes intervalos de integração pode também ajudar a definir a melhor configuração para uma determinada organização.

CI também não é compatível com projetos de código livre. Na maioria das vezes, os desenvolvedores desses projetos são voluntários e não têm disponibilidade para trabalhar diariamente no seu código. Nesses casos, um modelo baseado em Pull Requests e Forks, conforme usado pelo GitHub, é mais adequado.

6.4 Entrega Contínua

Feature Flags

6.5 Engenharia de Releases

Dentre as vantagens de CD, podemos citar:

- CD reduz o tempo de entrega de novas funcionalidades. Por exemplo, suponha que as funcionalidades F₁, F₂, ..., F_n estão previstas para uma nova release de um sistema. No modo tradicional, todas elas devem ser implementadas e testadas, antes da liberação da nova release. Por outro lado, com CD, as funcionalidades são gradativamente liberadas, à medida que ficam prontas. Ou seja, CD diminui o intervalo entre releases. Passa-se a ter mais releases, mas com um menor número de funcionalidades.
- CD torna novas releases (ou implantações) um “não-evento”. Ou seja, não existe mais um dia D ou um deadline para entrega de novas releases. Deadlines são sempre uma fonte de stress para desenvolvedores e operadores de sistema. A perda de um deadline, por exemplo, pode implicar que uma funcionalidade somente entre em produção meses depois.
- Além de reduzir o stress causado por deadlines, CD ajuda a manter os desenvolvedores motivados, pois eles não ficam meses trabalhando sem receber feedback. Em vez disso, os desenvolvedores rapidamente recebem retorno — proveniente de usuários reais — sobre o sucesso ou não de suas tarefas.
- Em linha com o item anterior, CD favore experimentação e um estilo de desenvolvimento orientado por dados e feedback dos usuários. Novas funcionalidades entram rapidamente em produção. Com isso, recebe-se retorno dos usuários, que podem recomendar mudanças na implementação ou, no limite, o cancelamento de alguma funcionalidade.

Entrega Contínua (vs Deployment Contínuo)

Deployment Contínuo (CD) não é recomendável para certos tipos de sistemas, incluindo sistemas desktop (como uma IDE, um navegador Web, etc), aplicações móveis e aplicações embutidas em hardware. Provavelmente, você não gostaria de ser notificado diariamente de que existe uma nova versão do navegador que usa em seu desktop, ou do sistema de rede social que usa em seu celular ou ainda de que um novo driver está disponível para sua impressora. Esses sistemas envolvem

um processo de instalação que não é transparente para seus usuários, como é a atualização de um sistema Web.

No entanto, mesmo nos sistemas mencionados no parágrafo anterior, pode-se usar um versão mais *fraca* de CD, chamada de **Entrega Contínua (Continuous Delivery)**. A ideia é simples: todo commit *pode* entrar em produção imediatamente. Ou seja, os desenvolvedores devem programar como se isso fosse acontecer. No entanto, existe uma autoridade externa — um gerente de projetos ou de releases, por exemplo — que toma a decisão sobre quando os commits, de fato, serão liberados para os usuários finais. Inclusive forças de mercado ou de estratégia da empresa podem influenciar nessa decisão.

Uma outra maneira de explicar esses conceitos é por meio da seguinte diferença:

- **Deployment to production:** processo de liberar uma nova versão de um sistema para seus usuários.
- **Delivery:** processo de liberar uma nova versão de um sistema para ser objeto de deployment.

Assim, quando adota-se Deployment Contínuo, ambos os processos são automáticos e contínuos. Porém, com Entrega Contínua, a entrega é realizada com frequência, mas o deployment depende de uma autorização manual.

Mundo Real: Vamos citar alguns dados sobre a frequência de deployments em sistemas não-Web. Por exemplo, o Google libera novas releases do navegador Chrome para o público a cada seis semanas. Até 2019, a IDE Eclipse tinha uma única nova release por ano. A partir de 2019, o sistema passou a ter uma nova release a cada 13 semanas. Um dos motivos foi “permitir que os desenvolvedores liberem novas funcionalidades de forma rápida”. Como um terceiro exemplo, a versão para Android do Facebook sofria uma atualização a cada oito semanas. Mais recentemente, o Facebook encurtou esse tempo para uma semana ([link](#)). Ou seja, as empresas estão lançando releases de forma mais rápida, para agradar os usuários, receber feedback, manter os desenvolvedores motivados e continuarem competindo no mercado.

Feature Flags

Nem sempre todo commit está pronto para entrar em produção. Por exemplo, o desenvolvedor pode estar trabalhando em uma funcionalidade X, mas ainda falta implementar parte de seus requisitos. Se novas releases são liberadas quase todo dia,

como garantir que essas implementações parciais não vão chegar até os usuários? Uma solução seria não integrá-las no branch principal de desenvolvimento. Porém, não queremos mais seguir essa prática, pois ela leva ao que chamamos de *integration (ou merge) hell*.

A solução para esse problema é simples: integre continuamente o código parcial de X, mas com a funcionalidade desabilitada, isto é, qualquer código relativo a X estará “guardado” por uma variável booleana (um *flag*) que, enquanto a implementação de X não for concluída, vai avaliar como falso:

```
featureX = false;
...
if (featureX)
    "aqui tem codigo incompleto de X"
...
if (featureX)
    "aqui tem mais codigo incompleto de X"
```

No contexto de deployment contínuo, variáveis booleanas usadas para desativar a execução de implementações parciais de funcionalidades são chamadas de **Feature Flags** ou **Feature Toggles**.

Para mostrar um segundo exemplo, suponha que você está trabalhando em uma nova página de um sistema. Você pode declarar uma feature flag para desabilitar o carregamento da nova página, como mostrado a seguir:

```
nova_pag = false;
...
if (nova_pag)
    "carregue nova pagina"
else
    "carregue pagina antiga"
```

Esse é o código que vai para produção, enquanto a nova página não estiver pronta. Porém, durante a implementação, localmente, você pode habilitar a nova página, fazendo `nova_pag` receber `true`. Assim, durante a implementação da nova página, vai existir duplicação de código entre as duas páginas. Porém, após a nova página ser aprovada, entrar em produção e receber feedback positivo dos clientes, o código da página antiga pode ser removido. Ou seja, a duplicação de código foi uma situação temporária.

Mundo Real: Pesquisadores de duas universidades canadenses, liderados pelos professores Peter Rigby e Bram Adams, realizaram um estudo sobre o uso de feature flags ao longo de 39 releases do navegador Chrome, relativas a cinco anos de desenvolvimento ([link](#)). Neste período, eles encontraram mais de 2.400 feature flags distintas no código do navegador. Na primeira versão analisada, eles catalogaram 263 flags; na última versão, o número aumentou para 2.409 flags. Na média, uma nova release adiciona 73 novos flags e remove 43 flags. Por isso, o crescimento reportado pelos autores.

Bibliografia

- Gene Kim, Jez Humble, John Willis, Patrick Debois. Manual de Devops. Como Obter Agilidade, Confiabilidade e Segurança em Organizações Tecnológicas. Alta Books, 2018.
- Jez Humble, David Farley. Entrega Contínua: Como Entregar Software de Forma Rápida e Confiável. Bookman, 2014.
- Steve Matyas, Andrew Glover, Paul Duvall. Continuous Integration: Improving Software Quality and Reducing Risk. Addison-Wesley, 2007.

Exercícios de Fixação

1. Defina e descreva os objetivos de DevOps.
2. Em sites de oferta de empregos na área de TI, é comum encontrar vagas para “Engenheiro Devops”, requerendo habilidades como as seguintes:
 - Ferramentas de controle de versão (Git, Bitbucket, SVN, etc)
 - Gerenciadores de dependência e build (Maven, Gradle e etc)
 - Ferramentas de integração contínua (Jenkins, Bamboo, VSTS)
 - Administração de servidores em Cloud (AWS e Azure)
 - Sistemas Operacionais (Ubuntu, CentOS e Red Hat)
 - Banco de dados (DynamoDB, Aurora Mysql)
 - Docker e orquestração de docker (Kubernetes, Mesos, Swarm)
 - Desenvolvimento com APIs REST, Java

Considerando a definição de DevOps que respondeu no exercício anterior, você considera adequado que a função de um funcionário seja “Engenheiro DevOps”? Justifique a sua resposta.

3. Defina (e diferencie) os seguintes termos: integração contínua (*continuous integration*); entrega contínua (*continuous delivery*) e implantação contínua (*continuous deployment*).
4. Porque integração contínua, entrega contínua e implantação contínua são práticas importantes em DevOps? Na sua resposta, considere a definição de DevOps que deu no primeiro exercício.
5. Pesquise o significado da expressão Teatro de CI (*CI Theater*) e então descreva esse significado com suas próprias palavras.

Apêndice A

Git

The best way to learn git is probably to first only do very basic things and not even look at some of the things you can do until you are familiar and confident about the basics. – Linus Torvalds

Neste apêndice, apresentamos e discutimos exemplos de uso do sistema Git, que é o sistema de controle de versões mais usado atualmente. Inspirados pela frase acima, de Linus Torvalds, criador do Git, vamos focar nos conceitos e comandos básicos desse sistema. Como sugere a frase, é importante dominar esses comandos antes de se aventurar no uso de comandos mais avançados. Caso o leitor não tenha conhecimento dos objetivos e vantagens proporcionados por um sistema de controle de versões, recomendamos primeiro a leitura da seção “Controle de Versões”, do Capítulo 10 deste livro.

A.1 Init & Clone

Para começar a usar o git para gerenciar as versões de um sistema devemos executar um dos seguintes comandos: **init** ou **clone**. O comando **init** cria um repositório vazio. O segundo comando — **clone** — primeiro chama **init** para criar um repositório vazio. Em seguida, ele copia para esse repositório todos os commits de um repositório remoto, passado como parâmetro. Seja, por exemplo, o seguinte comando:

```
git clone https://github.com/NOME-USER/NOME-REPO
```

Esse comando clona para o diretório corrente um repositório armazenado remo-

tamente no GitHub. Portanto, devemos usar clone quando vamos trabalhar em um projeto que já está em andamento e que já possui commits em um repositório central. No exemplo, esse repositório é disponibilizado pelo GitHub.

A.2 Commit

Commits são usados para criar snapshots (ou fotografias) dos arquivos de um sistema. Uma vez tiradas essas fotografias, elas são armazenadas no sistema de controle de versões, de forma compactada e otimizada, para não ocupar muito espaço em disco. Posteriormente, pode-se recuperar qualquer uma das fotografias, para, por exemplo, restaurar uma implementação antiga de um arquivo.

Recomenda-se que desenvolvedores realizem commits periodicamente, sempre que tiverem efetuado uma mudança importante no código. Em sistemas de controle de versões distribuídos, como o git, os commits são primeiramente armazenados no repositório local do desenvolvedor. Por isso, o custo de um commit é pequeno e, portanto, desenvolvedores podem realizar diversos commits ao longo de um dia de trabalho. Na verdade, o que não é recomendável é a realização de commits grandes, com modificações importantes em diversos arquivos. Também não recomenda-se que um commit inclua modificações relativas a mais de uma tarefa de manutenção. Por exemplo, não é recomendável corrigir dois bugs em um mesmo commit. Em vez disso, cada bug deve ser corrigido em um commit separado. Assim, facilita-se uma futura análise do código, caso, por exemplo, um cliente volte a reclamar que seu bug não foi corrigido.

Commits também possuem metadados, incluindo data, hora, autor e uma mensagem, que descreve a modificação realizada pelo commit. A próxima figura mostra uma página do GitHub que exibe os metadados principais de um commit do repositório google/guava. Pode-se observar que o commit refere-se a um refactoring, o que fica claro no seu título. Em seguida, o refactoring é explicado em detalhes na mensagem do commit. Na última linha da figura, podemos ver o nome do autor do commit e a informação de que ele foi realizado há 13 dias.

Na última linha da figura também podemos observar que todo commit possui um identificador único, no caso:

`1c757483665f0ba8fed31a2af7e31643a4590256`

Esse identificador possui 20 bytes, normalmente representados em hexadecimal. Esses bytes correspondem a uma verificação de consistência (*check sum*) do conteúdo do commit, conforme computado por uma função hash SHA-1.

The screenshot shows a GitHub pull request interface. At the top, there's a title bar with the text 'Refactor AbstractGraphTest to allow for tests with ImmutableGraph.' and a 'Browse files' button. Below the title, there's a detailed description of the reason for the change, mentioning that ImmutableGraph implementations are undertested compared to MutableGraph implementations. It also notes a bug introduced and a desire to share incident edge order tests. A note about design follows, explaining the choice of making an AbstractGraphTest subclass for mutable graphs instead of duplicating code. The commit message itself is empty ('RELNOTES=n/a'). Below the commit message, there's a note about migration: 'Created by MOE: <https://github.com/google/moe>' and 'MOE_MIGRATED_REVID=283035873'. The commit history shows a single commit from 'nymanjens' and 'nglorioso' 13 days ago, with a parent commit '2ee7f9d' and a commit hash '1c757483665f0ba8fed31a2af7e31643a4590256'. The commit message is empty.

Figura A.1: Commit no GitHub

A.3 Add

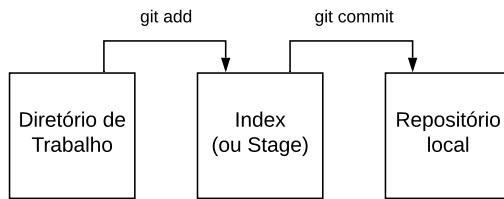
Na máquina local, o sistema git manipula três áreas distintas:

- Um **diretório de trabalho**, onde devemos salvar os arquivos que pretendemos versionar. Às vezes, essa área é chamada também de árvore de trabalho (*working tree*).
- O **repositório** propriamente dito, que armazena o histórico de commits.
- Uma área intermediária, chamada de **index** ou **stage**, que armazena temporariamente os arquivos que se pretende versionar. Tais arquivos são ditos rastreáveis (**tracked**).

Dentre essas três áreas, o desenvolvedor acessa apenas o diretório de trabalho, que funciona como um diretório comum do sistema operacional. As duas outras áreas são internas ao git e manipuladas exclusivamente por ele. Como qualquer diretório, o diretório de trabalho pode conter diversos arquivos. Porém, apenas aqueles adicionados ao index, por meio de um **add**, serão gerenciados pelo git.

Além de armazenar a lista de arquivos versionados, o index também armazena o conteúdo deles. Por isso, antes de fazer um **commit** devemos executar um **add**, para salvar o conteúdo do arquivo no index. Feito isso, podemos usar um **commit** para salvar no repositório local a versão adicionada ao index. Esse fluxo é ilustrado na próxima figura.

Exemplo: Suponha o seguinte arquivo simples, mas suficiente para explicar os comandos **add** e **commit**.

Figura A.2: Comandos `add` e `commit`

```
// arq1
x = 10;
```

Após criar esse arquivo, o desenvolvedor executou o seguinte comando:

```
git add arq1
```

Esse comando adiciona o arquivo `arq1` no index (ou stage). Porém, logo em seguida, o desenvolvedor modificou de novo o arquivo:

```
// arq1
x = 20; // novo valor de x
```

Feito isso, ele executou:

```
git commit -m "Alterando o valor de x"
```

A opção `-m` informa a mensagem que descreve o commit. Porém, o ponto que queremos ressaltar com esse exemplo é o seguinte: como o usuário não executou um novo `add` após mudar o valor de `x` para 20, a versão mais recente do arquivo não será salva pelo commit. Em vez disso, a versão de `arq1` que será versionada é aquela onde `x` tem o valor 10, pois ela é a versão que consta do index.

Para evitar o problema descrito nesse exemplo, é comum usar um `commit` da seguinte forma:

```
git commit -a -m "Alterando valor de x"
```

A opção `-a` indica que antes de executar o commit queremos adicionar no index todos os arquivos rastreados (*tracked*) que tenham sido modificados desde o último commit. Portanto, a opção `-a` não elimina a necessidade de usar `add`. O uso desse

comando continua sendo necessário, pelo menos uma vez, para indicar ao git que desejamos tornar um determinado arquivo rastreável.

Da mesma forma que existe um `add`, também existe uma operação para remover um arquivo de um repositório git. Um exemplo é dado a seguir:

```
git rm arq1.txt  
git commit -m "Removendo arq1.txt"
```

Além de remover do repositório git local, o comando `rm` também remove o arquivo do diretório de trabalho.

A.4 Push & Pull

O comando **push** copia os commits mais recentes do repositório local para o repositório remoto. Portanto, ele é uma operação mais lenta, pois envolve comunicação pela rede. Um **push** deve ser usado quando o desenvolvedor deseja tornar uma modificação visível para os demais desenvolvedores. Para atualizar seu repositório local, os outros desenvolvedores do time devem executar um comando **pull**. Esse comando realiza duas operações principais:

- Primeiro, um **pull** copia os commits mais recentes do repositório central para o repositório local do desenvolvedor. Essa operação inicial é chamada de **fetch**.
- Em seguida, o comando **pull** atualiza os arquivos do diretório de trabalho. Essa operação é chamada de **merge**.

A próxima figura ilustra o funcionamento dos comandos **push** e **pull**.

Exemplo: Suponha que no repositório git central de um projeto exista o seguinte arquivo:

```
void f() {  
    ...  
}
```

Suponha que dois desenvolvedores, chamados Bob e Alice, realizaram um **pull** e, portanto, copiaram esse arquivo para o repositório local e para o diretório de trabalho de suas máquinas. A sintaxe desse comando é a seguinte:

```
git pull
```

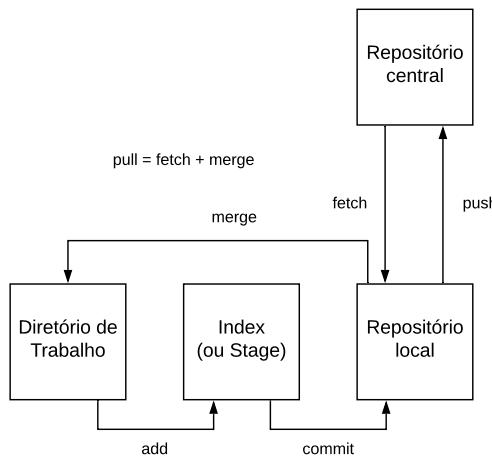


Figura A.3: Comandos push e pull

No mesmo dia, Bob implementou uma segunda função g no arquivo:

```

void f() {    // antiga
    ...
}

void g() {    // implementada por Bob
    ...
}
  
```

Em seguida, Bob realizou um `add`, um `commit` e um `push`. Esse último comando tem a seguinte sintaxe:

```
git push origin master
```

O parâmetro `origin` é um valor default, usado pelo git, para indicar o repositório remoto, por exemplo, o repositório GitHub. Já o parâmetro `master` indica o branch principal. Iremos estudar mais sobre branches daqui a pouco.

Após executar o comando `push` acima, a nova versão do arquivo estará salva também no repositório remoto. Alguns dias depois, Alice decidiu que precisa alterar esse mesmo arquivo. Como ela ficou um tempo sem trabalhar no sistema, o recomendado é que ela execute primeiro um `pull`, para atualizar seu repositório local e seu diretório de trabalho com as mudanças ocorridas nesse período, como

aquela realizada por Bob. Assim, após esse `pull`, o arquivo em questão será atualizado na máquina da Alice, para incluir a função `g` implementada por Bob.

A.5 Conflitos de Merge

Conflitos de merge acontecem quando dois desenvolvedores alteram o mesmo trecho de código ao mesmo tempo. Para entender melhor essa situação, nada melhor do que usar um exemplo.

Exemplo: Suponha que Bob implementou o seguinte programa:

```
main() {
    print("Hello, world!");
}
```

Concluída a implementação, Bob realizou um `add`, seguido de um `commit` e um `push`.

Em seguida, Alice realizou um `pull` e obteve a versão do arquivo implementada por Bob. Então, Alice resolveu traduzir a mensagem do programa para Português:

```
main() {
    print("Olá, mundo!");
}
```

Enquanto Alice fazia a tradução, Bob percebeu que escreveu Hello de forma errada, com apenas uma letra l. Porém, Alice foi mais rápida e realizou a trinca de comandos `add`, `commit` e `push`.

Bob, após corrigir o erro de ortografia, salvou o arquivo e também executou um `add`, seguido de um `commit`. Por fim, ele executou `push`, mas o comando falhou com a seguinte mensagem de erro:

Updates were rejected because the remote contains work that you do
not have locally. This is usually caused by another repository
pushing to the same ref. You may want to first integrate the
remote changes (e.g., git pull ...) before pushing again.

A mensagem é bem clara: Bob não pode executar um `push`, pois o repositório remoto possui conteúdo novo, no caso, gerado por Alice. Antes de executar um `push`, Bob precisa executar um `pull`. Porém, ao fazer isso, ele recebe uma nova mensagem de erro:

CONFLICT (content): Merge conflict in arq2

Automatic merge failed; fix conflicts and then commit the result.

Essa nova mensagem é também clara: existe um conflito de merge no arquivo arq2. Ao editar esse arquivo, Bob vai perceber que ele foi modificado pelo git, para destacar as linhas que geraram o conflito:

```
main() {
    <<<<< HEAD
    print("Hello, world!");
    =====
    print("Olá, mundo!");
    >>>>> f25bce8fea85a625b891c890a8eca003b723f21b
}
```

As linhas inseridas pelo git devem ser entendidas da seguinte forma:

- Entre <<<<< HEAD e ===== temos o código modificado por Bob, isto é, pelo desenvolvedor que não conseguiu dar um `push` e teve que dar um `pull`. `HEAD` designa que o código foi modificado no último commit realizado por Bob.
- Entre ===== e >>>>> f25bce8 ... temos o código modificado por Alice, isto é, pela desenvolvedora que executou com sucesso seu `push`. f225bce8... é o ID do commit no qual Alice modificou essa parte do código.

Cabe então a Bob resolver o conflito, o que é sempre uma tarefa manual. Para isso, ele tem que escolher o trecho de código que vai prevalecer — o seu código ou o da Alice — e editar o arquivo de acordo com tal escolha, para remover os delimitadores inseridos pelo git. Vamos supor que Bob decida que o código de Alice é o certo, pois agora o sistema está usando mensagens em Português. Logo, ele deve editar o arquivo, de forma que fique assim:

```
main() {
    print("Olá, mundo!");
}
```

Veja que Bob removeu os delimitadores inseridos pelo git (<<<<< HEAD , ===== e >>>>> f25bce8...). E também o comando `print` com a mensagem em inglês. Após deixar o código da forma correta, Bob deve executar novamente os comandos `add`, `commit` e `push`, que agora serão bem sucedidos.

Nesse exemplo, mostramos um conflito simples, que ficou restrito a única linha de um único arquivo. No entanto, um `pull` pode dar origem a conflitos mais

complexos. Por exemplo, um mesmo arquivo pode apresentar vários conflitos. E também podemos ter conflitos em mais de um arquivo.

A.6 Branches

O git organiza o diretório de trabalho em “diretórios virtuais”, chamados de **branches**. Até agora, não precisamos comentar sobre branches porque todo repositório possui um branch default, chamado de **master**, criado pelo comando `init`. Se não nos preocuparmos com branches, todo o desenvolvimento ocorrerá no master. Porém, em alguns casos, é interessante criar outros branches para melhor organizar o desenvolvimento. Para descrever o conceito de branches, vamos de novo usar um exemplo.

Exemplo: Suponha que Bob é responsável por manter uma determinada funcionalidade de um sistema. Para simplificar, vamos assumir que essa funcionalidade é implementada em uma única função `f`. Bob teve a ideia de mudar completamente a implementação de `f`, de forma que ela passe a usar algoritmos e estruturas de dados mais eficientes. Para isso, Bob vai precisar de algumas semanas. No entanto, apesar de estar otimista, Bob não tem certeza de que a nova implementação vai proporcionar os ganhos que ele imagina. Por fim, mas não menos importante, durante a implementação do novo código, Bob pode precisar do código original de `f`, para, por exemplo, corrigir bugs reportados pelos usuários.

Esse é um cenário interessante para Bob criar um branch para implementar e testar — de forma isolada — essa nova versão de `f`. Para isso, ele deve usar o comando:

```
git branch f-novo
```

Esse comando cria um novo branch, chamado `f-novo`, supondo que esse branch ainda não existe.

Para mudar do branch corrente para um novo branch, deve-se usar `git checkout [nome-branch]`. Para descobrir qual o nome do branch corrente, basta usar `git branch`. Na verdade, esse comando lista todos os branches e indica qual deles é o corrente.

Branches podem ser entendidos como "sub-diretórios virtuais" do diretório de trabalho. A principal diferença é que branches são gerenciados pelo git e não pelo sistema operacional. Por isso, optamos por chamá-los de virtuais. Explorando mais essa comparação, podemos pensar que o comando `git branch [nome]` equivale ao comando `mkdir [nome]`, com a diferença que o git não apenas cria um branch mas copia para ele todos os arquivos do branch pai. Por outro lado, diretórios são

criados vazios pelo sistema operacional. Já o comando `git checkout [nome]` lembra o comando `cd [nome]`. E `git status` lembra um misto de comandos `ls` e `pwd`. Também para reforçar essa comparação, existem certos comandos que permitem adicionar ao prompt do sistema operacional não apenas o nome do diretório corrente mas também o nome do branch corrente. Assim, o prompt pode ser exibido como `c:\projetos\sistema1\master>`.

Por outro lado, existe também uma diferença importante entre branches e diretórios. Um desenvolvedor somente pode alterar o branch corrente de A para B se as modificações que ele fez em A estiverem salvas. Isto é, se ele tiver realizado antes um `add` e `commit`. Caso ele tenha esquecido de chamar esses comandos, um comando `git checkout B` irá falhar com a seguinte mensagem de erro:

```
Your local changes to the following files would be overwritten by checkout:  
[list of files]
```

Please commit your changes or stash them before you switch branches.

Voltando ao exemplo, após Bob ter criado o seu branch, ele deve proceder do seguinte modo. Quando ele quiser trabalhar na nova implementação de `f`, ele deve primeiro mudar o branch corrente para `f-novo`. Por outro lado, quando ele precisar modificar o código original de `f` — aquele que está em produção — ele deve se certificar de que o branch corrente é o `master`. Independentemente do branch em que estiver, Bob deve usar `add` e `commit` para salvar o estado do seu trabalho.

Bob vai continuar nesse fluxo, alternando entre os branches `f-novo` e `master`, até que a nova implementação de `f` esteja concluída. Quando isso acontecer, Bob vai precisar copiar o novo código de `f` para o código original. No entanto, como está usando branches, ele não precisa realizar essa operação de forma manual. O git oferece uma operação, chamada `merge`, que realiza exatamente essa cópia. A sintaxe é a seguinte:

```
git merge f-novo
```

Esse comando deve ser chamado no branch que irá receber as modificações realizadas em `f-novo`. No nosso caso, no branch `master`.

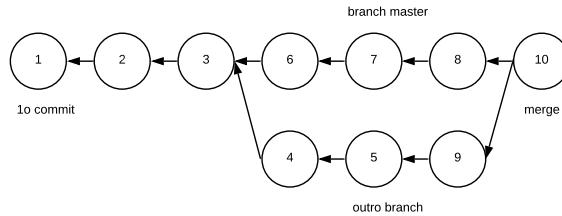
Como o leitor já deve estar pensando, um merge pode gerar conflitos, chamados também de **conflitos de integração**. No caso específico de merge de branches, esses conflitos vão ocorrer quando tanto o branch que está recebendo as modificações (`master`, no nosso exemplo) como o branch que está sendo integrado (`f-novo`, no exemplo) tiverem alterado os mesmos trechos de código. Conforme discutido na Seção A.6, o git irá delimitar os trechos com conflitos e caberá ao desenvolvedor que chamou o `merge` resolvê-los. Isto é, escolher o código que deve prevalecer.

Por fim, após realizar o merge, Bob pode remover o branch `f-novo`, caso não seja importante manter o histórico dos commits realizados para implementar a nova versão de `f`. Para deletar `f-novo`, ele deve executar o seguinte comando no master:

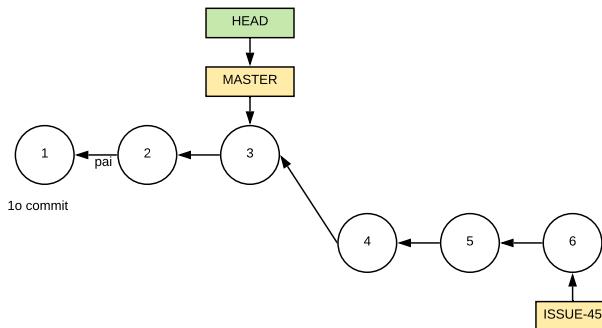
```
git branch -d f-novo
```

A.6.1 Grafo de Commits

Commits podem possuir zero, um ou mais pais (ou antecessores). Como ilustra a próxima figura, o primeiro commit de um repositório não possui pai. Já um commit de merge possui dois ou mais pais, que representam os branches que foram unidos. Os demais commits possuem exatamente um pai.

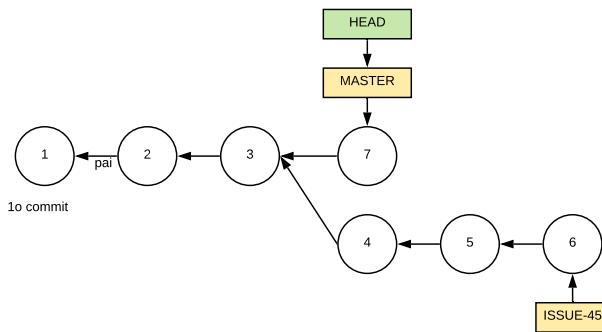


Um branch nada mais é do que uma variável interna do git que contém o identificador do último commit realizado no branch. Existe ainda uma variável chamada `HEAD`, que aponta para a variável do branch atual. Ou seja, `HEAD` contém o nome da variável que contém o identificador do último commit do branch atual. Um exemplo é mostrado a seguir:



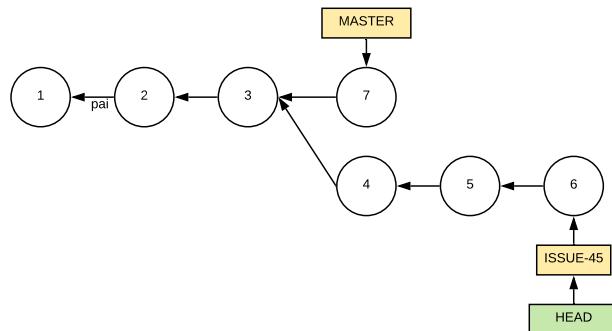
Nesse exemplo, existem dois branches, representados pelas variáveis `MASTER` e `ISSUE-45`. Cada uma delas aponta para o último commit de seu respectivo branch.

A variável `HEAD` aponta para a variável `MASTER`. Isso significa que o branch atual é o `MASTER`. Se realizarmos um commit, a configuração mudará para:



O novo commit tem identificador 7. Ele foi realizado no `MASTER`, já que `HEAD` apontava para a variável desse branch. O pai do novo commit pai é o antigo `HEAD`, no caso o commit 3. A variável `MASTER` avançou e passou a apontar para o novo commit. Isso significa que, se não mudarmos de branch, o pai do próximo commit será o commit 7.

Porém, se mudarmos para o branch `ISSUE-45`, a configuração será a seguinte:



A única mudança é que a variável `HEAD` passou a apontar para a variável do novo branch, isto é, para a variável `ISSUE-45`. Isso é suficiente para fazer com que o próximo commit seja realizado nesse branch, isto é, que ela tenha o commit 6 como pai.

A.7 Branches Remotos

Até esse momento, trabalhamos com branches localmente, isto é, os branches que discutimos existem apenas no repositório local. No entanto, é possível realizar o `push` de um branch local para um repositório remoto. Para ilustrar esse recurso, vamos usar um exemplo semelhante ao da seção anterior.

Exemplo: Suponha que Bob criou um branch, chamado `g-novo`, para implementar uma nova funcionalidade. Ele realizou alguns commits nesse branch e agora gostaria de compartilhá-lo com Alice, para que ela implemente parte da nova funcionalidade. Para isso, Bob deve usar o seguinte `push`:

```
git push -u origin g-novo
```

Esse comando realiza o `push` do branch corrente (`g-novo`) para o repositório remoto, chamado pelo git de `origin`. O repositório remoto pode, por exemplo, ser um repositório do GitHub. O parâmetro `-u` indica que, no futuro, vamos querer sincronizar os dois repositórios por meio de um `pull` (a letra do parâmetro vem da palavra *upstream*). Essa sintaxe vale apenas para o primeiro `push` de um branch remoto. Nos comandos seguintes, pode-se omitir o `-u`, isto é, usar apenas `git push origin g-novo`.

No repositório remoto, será criado um branch `g-novo`. Para trabalhar nesse branch, Alice deve primeiro criá-lo na sua máquina local, mas associado ao branch remoto, por meio dos seguintes comandos, que devem ser executados no master:

```
git pull  
git checkout -t origin/g-novo
```

O primeiro comando é necessário para tornar o branch remoto visível na máquina local. Já o segundo comando cria um branch local, chamado `g-novo`, que Alice vai usar para rastrear mudanças no branch remoto `origin/g-novo`, conforme indica o parâmetro `-t`, que vem da palavra *tracking*. Em seguida, Alice pode realizar commits nesse branch. Por fim, quando estiver pronta para publicar suas mudanças, ela deve executar um `push`, com a sintaxe normal, isto é, sem o parâmetro `-u`.

Agora, Bob pode realizar um `pull`, concluir que a implementação da nova funcionalidade está finalizada e, portanto, pode ser integrada no master, por meio de um `merge`. Bob pode também deletar os branches local e remoto, usando os comandos:

```
git branch -d g-novo  
git push origin --delete g-novo
```

E Alice também pode deletar seu branch local, chamando apenas:

```
git branch -d g-novo
```

A.8 Pull Requests

Pull requests é um mecanismo que viabiliza que um branch seja revisado e discutido antes de ser integrado no branch principal. Quando se usa pull requests, um desenvolvedor sempre implementa novas funcionalidades em um branch separado. Concluída a implementação, ele não integra imediatamente o novo código no branch principal. Antes que isso ocorra, ele abre uma solicitação para que seu branch seja revisado e aprovado por um segundo desenvolvedor. Essa solicitação para revisão e integração de código é chamada de pull request. Trata-se de um mecanismo mais comum no GitHub, mas que possui equivalente em outros sistemas de controle de versões.

Modernamente, o processo de revisão e integração do código de um pull request ocorre via interface Web, provida, por exemplo, pelo GitHub. Porém, se essa interface não existisse, o revisor teria que começar o seu trabalho realizando um `pull` do branch para sua máquina local. Daí então a origem do nome, isto é, pull request é uma solicitação (*request*) para que um segundo desenvolvedor revise e integre um determinado branch. Para atender a essa solicitação, ele deve começar realizando um `pull` do branch.

A seguir, vamos detalhar o processo de submissão e revisão de pull requests por meio de um exemplo.

Exemplo: Suponha que Bob e Alice são membros de uma organização que mantém um repositório chamado awesome-git, com uma lista de links interessantes sobre git. Os links ficam armazenados no arquivo README.md desse repositório, isto é, na sua página principal, cujo endereço é github.com/aserg-ufmg/awesome-git. Qualquer membro da organização pode sugerir a adição de links nessa página. Mas veja que estamos usando a palavra sugerir. Isto é, eles não podem fazer um `push` diretamente no branch master. Em vez disso, a sugestão de link precisa ser revisada e aprovada por um outro membro do time.

Bob resolveu então sugerir a adição, nessa página, de um link para o presente apêndice do livro Engenharia de Software Moderna. Para isso, ele primeiro clonou o repositório e criou um branch, chamado `livro-esm`, por meio dos seguintes comandos:

```
git clone https://github.com/aserg-ufmg/awesome-git.git  
git checkout livro-esm
```

Em seguida, Bob editou o arquivo README.md, adicionando a URL do apêndice. Por fim, ele realizou um `add`, um `commit` e fez um `push` do branch para o GitHub:

```
git add README.md  
git commit -m "Livro ESM"  
git push -u origin livro-esm
```

Na verdade, esses passos não são novidade em relação ao que vimos na seção anterior. No entanto, as diferenças começam agora. Primeiro, Bob deve ir na página do GitHub e selecionar o branch `livro-esm`. Feito isso, o GitHub mostrará um botão para criação de pull requests. Bob deve clicar nesse botão e descrever o seu pull request, como mostra a próxima figura.

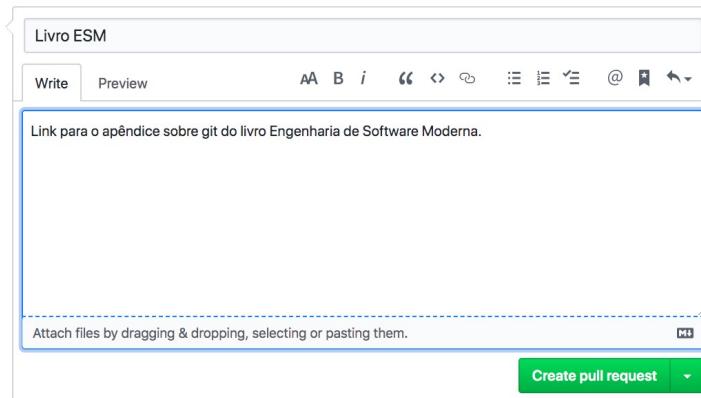


Figura A.4: Exemplo de pull request

Um pull request é uma solicitação para que um outro desenvolvedor revise e, se for o caso, realize o merge de um branch que você criou. Consequentemente, pull requests são um recurso para que uma organização passe a adotar **revisões de código**. Ou seja, desenvolvedores não integram diretamente o seu código no master do repositório remoto. Em vez disso, eles solicitam, via pull requests, que outros desenvolvedores revisem primeiro esse código e então façam o merge.

Na página do GitHub para criação de pull requests, Bob pode informar que deseja que seu código seja revisado pela Alice. Ela será então notificada que existe um pull request esperando sua revisão. Também via interface do GitHub, Alice pode revisar os commits do pull request criado por Bob, inclusive por meio de um diff entre o código novo e o código antigo. Se for o caso, Alice pode trocar mensagens com Bob, para esclarecer dúvidas sobre o novo código. Mais ainda, ela pode solicitar mudanças no código. Nesse caso, Bob deve providenciar as

mudanças e realizar um novo `add`, `commit` e `push`. Então, os novos commit serão automaticamente anexados ao pull request, para que Alice possa conferir se o seu pedido foi atendido. Estando a modificação aprovada, Alice pode integrar o código no master, bastando para isso clicar em um dos botões da página de revisão de pull requests.

A.9 Squash

Squash é um comando que permite unir diversos commits em um único commit. É uma operação recomendada, por exemplo, antes de submeter pull requests.

Exemplo: No exemplo anterior, suponha que o pull request criado por Bob tivesse cinco commits. Mais especificamente, ele está sugerindo o acréscimo de cinco novos links no repositório awesome-git, os quais foram coletados por ele ao longo de algumas semanas. Após a descoberta de cada link, Bob executou um commit na sua máquina local. E deixou para realizar o pull request apenas após acumular cinco commits.

Para facilitar a revisão de seu pull request por parte de Alice, Bob pretende unir esses cinco commits em um único commit. Assim, em vez de analisar cinco commits, Alice vai ter que analisar apenas um. Porém, a modificação submetida será exatamente a mesma, isto é, ela consiste na inclusão de cinco novos links na página. Porém, em vez de a solicitação estar distribuída em cinco commits (cada commit, adicionando um único link), ela estará concentrada em apenas um commit (adicionando cinco links).

Para realizar um squash, Bob deve chamar:

```
git rebase -i HEAD~5
```

O número 5 significa que pretende-se unir os cinco últimos commits do branch atual. Um editor de textos será aberto com uma lista contendo o ID e a descrição de cada um, como mostrado a seguir:

```
pick 16b5fcc Incluindo link 1
pick c964dea Incluindo link 2
pick 06cf8ee Incluindo link 3
pick 396b4a3 Incluindo link 4
pick 9be7fdb Incluindo link 5
```

Bob deve então usar o próprio editor para substituir a palavra `pick` por `squash`, exceto aquela da primeira linha. O arquivo ficará então assim:

```
pick 16b5fcc Incluindo link 1
squash c964dea Incluindo link 2
squash 06cf8ee Incluindo link 3
squash 396b4a3 Incluindo link 4
squash 9be7fdb Incluindo link 5
```

Bob deve então salvar o arquivo. Automaticamente, um novo editor será aberto, para ele informar a mensagem do novo commit — isto é, do commit que junta os cinco commits listados. Uma vez informada a mensagem, Bob deve salvar o arquivo e, então, o squash estará finalizado.

A.10 Forks

Fork é o mecanismo que o GitHub oferece para clonar repositórios remotos, isto é, repositórios armazenados pelo próprio GitHub. Um fork é realizado via interface do GitHub. Na página de qualquer repositório, existe um botão para realizar essa operação. Se fizermos um fork do repositório torvalds/linux será criado uma cópia desse repositório na nossa conta do GitHub, chamado, por exemplo, mtov/linux.

Como fazemos sempre, vamos usar um exemplo para explicar essa operação.

Exemplo: Suponha o repositório github.com/aserg-ufmg/awesome-git, usado no exemplo sobre pull requests. Suponha ainda uma terceira desenvolvedora, chamada Carol. Porém, como Carol não é membro da organização ASERG/UFMG, ela não tem permissão para realizar push nesse repositório, como fez Bob no exemplo anterior. Apesar disso, Carol acha que na lista atual falta um link importante e interessante, cuja inclusão ela gostaria de sugerir. Mas relembrando: Carol não pode seguir os mesmos passos usados por Bob no exemplo anterior, pois ela não tem permissão para dar push no repositório em questão.

Para resolver esse problema, Carol deve começar criando um fork do repositório. Para isso, basta clicar no botão fork, que existe na página de qualquer repositório no GitHub. Assim, Carol terá na sua conta do GitHub um novo repositório, cujo endereço será o seguinte: github.com/carol/awesome-git. Ela poderá clonar esse repositório para sua máquina local, criar um branch, adicionar o link que deseja na lista de links e realizar add, commit e push. Essa última operação será realizada no repositório resultante do fork. Por último, Carol deve ir na página do seu fork no GitHub e solicitar a criação de um pull request. Como o repositório é um fork, ela terá agora uma opção extra: destinar o pull request para o repositório original. Assim, caberá aos desenvolvedores do repositório original, como Bob e Alice, revisar e, se for o caso, aceitar o pull request.

Portanto, fork é um mecanismo que, quando combinado com pull requests, viabiliza que um projeto de código aberto receba contribuições de outros desenvolvedores. Explicando um pouco melhor, um projeto de código aberto pode receber contribuições — mais especificamente, commits — não apenas de seu time de desenvolvedores (Bob e Alice, no nosso exemplo), mas de um outro desenvolvedor com conta no GitHub (como é o caso de Carol).

Bibliografia

- Scott Chacon; Ben Straub. Pro Git. 2a edição, Apress, 2014.
- Rachel M. Carmena. How to teach Git. Blog post ([link](#)).

Exercícios de Fixação

Neste apêndice, mostramos diversos exemplos. Tente reproduzir cada um deles. Nos exemplos que envolvem repositórios remotos, a sugestão é usar um repositório do GitHub. Nos exemplos que envolvem dois usuários (Alice e Bob, por exemplo), a sugestão é criar dois diretórios locais e usá-los para reproduzir os comandos de cada usuário.