

Padrões de Design de Engenharia de Dados

Receitas para resolver os problemas mais comuns de engenharia de dados

Bartosz Konieczny

O'REILLY®
Novatec

Authorized Portuguese translation of the English edition of Data Engineering Design Patterns ISBN 9781098165819 © 2025 Bartosz Konieczny. This translation is published and sold by permission of O'Reilly Media, Inc., the owner of all rights to publish and sell the same.

Tradução em português autorizada da edição em inglês da obra Data Engineering Design Patterns ISBN 9781098165819 © 2025 Bartosz Konieczny. Esta tradução é publicada e vendida com a permissão da O'Reilly Media, Inc., detentora de todos os direitos para publicação e venda desta obra.

© Novatec Editora Ltda. [2025].

Todos os direitos reservados e protegidos pela Lei 9.610 de 19/02/1998. É proibida a reprodução desta obra, mesmo parcial, por qualquer processo, sem prévia autorização, por escrito, do autor e da Editora.

Editor: Rubens Prates

Tradução: Cláudio José Adas

Revisão gramatical: Denise Elisabeth Himpel

ISBN do impresso: 978-85-7522-966-8

ISBN do ebook: 978-85-7522-967-5

Novatec Editora Ltda.

Rua Luís Antônio dos Santos 110

02460-000 – São Paulo, SP – Brasil

Tel.: +55 11 2959-6529

Email: novatec@novatec.com.br

Site: <https://novatec.com.br>

Twitter: twitter.com/novateceditora

Facebook: facebook.com/novatec

LinkedIn: <https://linkedin.com/company/novatec-editora/>

GRA20250828

Sumário

Prefácio	9
Capítulo 1: Introduzindo Padrões de Design de Engenharia de Dados	15
O que são padrões de design?	15
Mais padrões de design?	17
Padrões comuns de engenharia de dados	17
Estudo de caso usado neste livro	19
Resumo	21
Capítulo 2: Padrões de Design de Ingestão de Dados.....	22
Carga total	23
Padrão: Carregador Completo	23
Carga incremental	26
Padrão: Carregador Incremental	27
Padrão: Captura de Dados Alterados.....	31
Replicação.....	35
Padrão: Replicador de Passagem	35
Padrão: Replicador de Transformação.....	38
Compactação de dados.....	41
Padrão: Compactador	41
Prontidão de dados	44
Padrão: Marcador de Prontidão.....	44
Ativado por eventos	47
Padrão: Gatilho Externo	47
Resumo	51

Capítulo 3: Padrões de Design de Gerenciamento de Erros	52
Registros não processáveis	53
Padrão: Registro não Entregue (Dead-Letter)	53
Registros duplicados	59
Padrão: Desduplicador com Janela.....	59
Dados atrasados.....	63
Padrão: Detector de Dados Atrasados	63
Padrão: Integrador Estático de Dados Atrasados	69
Padrão: Integrador Dinâmico de Dados Atrasados	74
Filtragem.....	80
Padrão: Interceptador de Filtros	80
Tolerância a falhas	84
Padrão: Ponto de Verificação (Checkpoint).....	84
Resumo	88
 Capítulo 4: Padrões de Design de Idempotência	 89
Substituição	90
Padrão: Limpeza Rápida de Metadados.....	90
Padrão: Substituição de Dados.....	95
Atualizações	98
Padrão: Fusão	98
Padrão: Fusão com Estado.....	103
Banco de dados	109
Padrão: Idempotência com Chave.....	109
Padrão: Gravador Transacional	114
Conjunto de dados imutável	118
Padrão: Proxy	118
Resumo	122
 Capítulo 5: Padrões de Design de Valor de Dados	 123
Enriquecimento de Dados.....	124
Padrão: Junção Estática	124
Padrão: Junção Dinâmica	129
Decoração de Dados	133
Padrão: Encapsulamento	133
Padrão: Decorador de Metadados	137
Agregação de Dados.....	141
Padrão: Agregador Distribuído.....	141
Padrão: Agregador Local	145
Sessionalização.....	148
Padrão: Sessionalizador Incremental	149
Padrão: Sessionalizador com Estado	154

Ordenação de dados.....	160
Padrão: Classificador de Recipientes.....	160
Padrão: Classificador FIFO.....	164
Resumo	168
Capítulo 6: Padrões de Design de Fluxo de Dados.....	169
Sequência.....	170
Padrão: Sequenciador Local.....	170
Padrão: Sequenciador Isolado.....	174
Fan-In.....	178
Padrão: Fan-In Alinhado	178
Padrão: Fan-In Desalinhado	182
Fan-Out.....	185
Padrão: Divisão Paralela	186
Padrão: Escolha Exclusiva.....	189
Orquestração.....	194
Padrão: Executor Único	194
Padrão: Executor Simultâneo	196
Resumo	198
Capítulo 7: Padrões de Design de Segurança de Dados	199
Remoção de dados	200
Padrão: Particionador Vertical.....	200
Padrão: Substituição no Local.....	205
Controle de Acesso.....	209
Padrão: Acesso Refinado a Tabelas.....	209
Padrão: Acesso Refinado a Recursos.....	212
Proteção de dados.....	216
Padrão: Criptografia	217
Padrão: Anonimizador.....	220
Padrão: Pseudoanonimizador	223
Conectividade	227
Padrão: Ponteiro de Segredos.....	227
Padrão: Conector sem Segredo	229
Resumo	232
Capítulo 8: Padrões de Design de Armazenamento de Dados.....	234
Particionamento	235
Padrão: Particionador Horizontal	235
Padrão: Particionador Vertical.....	241
Organização de Registros	244
Padrão: Bucket.....	244
Padrão: Classificador	246

Otimização de Desempenho de Leitura	250
Padrão: Aprimorador de Metadados.....	250
Padrão: Materializador de Conjunto de Dados	254
Padrão: Manifesto	256
Representação de Dados.....	259
Padrão: Normalizador.....	259
Padrão: Desnormalizador	265
Resumo	270
Capítulo 9: Padrões de Design de Qualidade de Dados.....	271
Garantir a Qualidade	272
Padrão: Auditar-Gravar-Auditar-Publicar	272
Padrão: Aplicação de Restrições	279
Consistência do Esquema.....	282
Padrão: Aplicação de Compatibilidade de Esquema.....	283
Padrão: Migrador de Esquema.....	288
Observação da Qualidade	291
Padrão: Observador Off-Line.....	291
Padrão: Observador On-Line.....	295
Resumo	300
Capítulo 10: Padrões de Design de Observabilidade de Dados.....	301
Detectores de Dados	302
Padrão: Detector de Interrupção de Fluxo	302
Padrão: Detector de Distorção.....	306
Detectores de Tempo	310
Padrão: Detector de Atraso	310
Padrão: Detector de Descumprimento do SLA	313
Linhagem de Dados	317
Padrão: Rastreador de Conjunto de Dados.....	317
Padrão: Rastreador Refinado.....	321
Resumo	324
Posfácio	325
Apêndice: Resumo de padrões	326
Índice remissivo	335