

# **Contratos de Dados**

**Desenvolvendo pipelines  
escaláveis de dados**

**Chad Sanderson  
Mark Freeman  
B.E. Schmidt**

O'REILLY®  
Novatec

Authorized Portuguese translation of the English edition of Data Contracts ISBN 9781098157630 © 2025 Manifest Data Labs, Inc. 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 Contracts ISBN 9781098157630 © 2025 Manifest Data Labs, Inc. 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. [2026].

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: Daniel Vieira

ISBN do impresso: 978-85-7522-994-1

ISBN do ebook: 978-85-7522-995-8

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](mailto:novatec@novatec.com.br)

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

Twitter: [twitter.com/novateceditora](https://twitter.com/novateceditora)

Facebook: [facebook.com/novatec](https://facebook.com/novatec)

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

RNV20260330

**Dados Internacionais de Catalogação na Publicação (CIP)**  
**(Câmara Brasileira do Livro, SP, Brasil)**

Sanderson, Chad  
Contratos de dados : desenvolvendo pipelines  
escaláveis de dados / Chad Sanderson, Mark Freeman,  
B.E. Schmidt ; tradução Daniel Vieira. --  
São Paulo : Novatec Editora, 2026.

Título original: Data contracts.  
ISBN 978-85-7522-994-1

1. Ciência dos dados 2. Dados - Estruturas  
(Ciência da computação) 3. Infraestrutura de  
tecnologia da informação I. Freeman, Mark.  
II. Schmidt, B.E. III. Vieira, Daniel. IV. Título.

26-344387.1

CDD-005.73

**Índices para catálogo sistemático:**

1. Ciência de dados 005.73

Eliane de Freitas Leite - Bibliotecária - CRB 8/8415

# Sumário

Apresentação.....	9
Prefácio .....	11
<b>PARTE I: Introdução à arquitetura de contratos de dados .....</b>	<b>17</b>
<b>CAPÍTULO 1: Por que o setor agora precisa de contratos de dados.....</b>	<b>18</b>
Ciclo “garbage in, garbage out” .....	19
Gerenciamento moderno de dados.....	19
O que é dívida de dados? .....	20
Como o efeito garbage-in garbage-out se potencializa.....	23
A morte dos data warehouses .....	24
A era pré-moderna .....	27
O software engole o mundo .....	28
Uma mudança para microsserviços .....	30
Arquitetura de dados em deterioração.....	30
Ascensão da Modern Data Stack.....	32
Os grandes jogadores.....	32
Crescimento rápido .....	33
Problemas no paraíso .....	34
IA centrada em dados e o surgimento das práticas de shift left.....	37
Retorno do investimento decrescente na melhoria dos modelos de ML .....	40
Comoditização dos fluxos de trabalho de ciência de dados.....	41
Ascensão dos dados sobre o ML na criação de vantagem competitiva .....	42
Surgimento das práticas de shift left de dados.....	42
Conclusão .....	43

<b>CAPÍTULO 2: Qualidade dos dados não se refere a dados impecáveis .....</b>	<b>44</b>
Definição da qualidade dos dados .....	44
OLTP <i>versus</i> OLAP e suas implicações para a qualidade dos dados .....	46
Um breve resumo de OLTP e OLAP .....	47
Problemas de tradução entre visões de mundo de dados OLTP e OLAP .....	49
Custo da má qualidade dos dados .....	51
Medição da qualidade dos dados .....	52
Quem é impactado .....	57
Conclusão .....	59
<b>CAPÍTULO 3: Desafios da escalabilidade da infraestrutura de dados .....</b>	<b>61</b>
Diferenças entre desenvolvimento de dados e de software .....	61
Como os engenheiros de software criam produtos .....	62
Como os desenvolvedores de dados criam produtos .....	62
Desafios para equipes modernas de engenharia de dados .....	65
Por que o desenvolvimento de dados precisa de uma superfície de design .....	69
Custo de refatorações em larga escala .....	71
Considerações sobre refatoração em larga escala .....	72
Caso de uso: a refatoração em larga escala na Alan .....	73
Perigos das migrações de banco de dados .....	75
Papel do gerenciamento de mudanças na qualidade dos dados .....	77
Comportamento entrópico dos dados .....	78
Como os dados derivam da lógica de negócios estabelecida .....	78
A gestão de mudanças precisa alinhar-se às necessidades do negócio .....	80
Como as necessidades de infraestrutura mudam em escala .....	81
Número de Dunbar e Lei de Conway .....	82
Estudo de caso: equipe de engenharia da Atlassian .....	83
Como contratos de dados possibilitam o gerenciamento de mudanças em escala .....	84
Conclusão .....	85
<b>CAPÍTULO 4: Introdução aos contratos de dados .....</b>	<b>87</b>
A colaboração é diferente no mundo dos dados .....	87
Stakeholders com os quais você vai trabalhar .....	90
Papel dos produtores de dados .....	91
Papel dos consumidores de dados .....	92
Impacto de produtores e consumidores .....	95
Desafios e dificuldades dos consumidores de dados gerenciando a qualidade dos dados .....	97
Uma alternativa: o fluxo de trabalho do contrato de dados .....	101
Etapas do fluxo de trabalho do contrato de dados .....	102
Onde os contratos de dados são implementados .....	106
Curva de maturidade dos contratos de dados: conscientização, propriedade e governança ..	108
Resultados da implementação de contratos de dados .....	112
Contratos de dados <i>versus</i> observabilidade de dados .....	113
Conclusão .....	116

<b>PARTE II: Implementação da arquitetura de contrato de dados .....</b>	<b>117</b>
<b>CAPÍTULO 5: Componentes do contrato de dados: ativos de dados e definição de contrato .....</b>	<b>118</b>
Visão geral dos componentes.....	118
Ativos de dados.....	120
Bancos de dados analíticos .....	121
Bancos de dados transacionais.....	122
Sourcing e streams de eventos.....	125
Dados primários, plataforma de terceiros .....	129
Definição de contrato .....	130
Especificação do contrato de dados .....	131
Lógica de negócios.....	137
Schema Registry e catálogos de dados .....	138
Conclusão .....	142
<b>CAPÍTULO 6: Componentes do contrato de dados: detecção e prevenção .....</b>	<b>143</b>
Detecção .....	144
Quarentena de dados.....	148
Captura de dados de mudança, processamento de fluxos e monitoramento de dados em tempo real.....	149
Linhagem de ponta a ponta: grafos de fluxo de código-fonte e linhagem de dados.....	151
Análise estática de código.....	154
Prevenção .....	157
Controle de versão .....	157
CI/CD.....	158
Monitoramento e alertas de violações.....	160
Conclusão .....	164
<b>CAPÍTULO 7: Implementação de contratos de dados .....</b>	<b>166</b>
Introdução ao cenário simulado de contrato de dados .....	167
Conjunto de dados.....	168
Aplicação do museu.....	169
Visão geral do repositório do projeto.....	169
Implementação dos contratos de dados .....	171
Visão geral da arquitetura dos contratos de dados.....	171
Componente A: ativos de dados.....	172
Componente B: definição do contrato .....	174
Componente C: detecção .....	177
Componente D: prevenção.....	180
Juntando tudo.....	182
Cenário de violação do contrato de dados.....	183
Solicitação recebida para normalizar a tabela object_images.....	184
Atualize o arquivo de migração de dados para o novo esquema.....	185
Verifique os testes unitários e acione a violação contratual .....	187

Envie para o branch remoto e visualize os logs de erro .....	188
Reverta as alterações de migração do banco de dados .....	189
Converse com a equipe downstream para entender o caso de uso .....	189
Crie novo contrato de dados para a tabela object_images_normalized .....	190
Execute testes unitários e veja o erro de ativo ausente no catálogo .....	192
Corrija a migração do banco de dados até que todas as verificações sejam aprovadas.....	193
Envie para o branch remoto e veja se a mudança passa .....	195
Conclusão .....	196

## **CAPÍTULO 8: Estudos de caso do mundo real: contratos de dados em produção..... 197**

História do contrato de dados da Convoy.....	198
História do contrato de dados da Glassdoor .....	203
História do contrato de dados da Adevinta Espanha .....	206
Conclusão .....	213

## **PARTE III: Obtendo o apoio da liderança para a arquitetura de contratos de dados.....215**

### **CAPÍTULO 9: Shift left: a mudança cultural necessária para os contratos de dados ..... 216**

O que precisa ser verdade: o que deixamos passar sobre o shift left dos dados.....	217
Entendendo os riscos: três exemplos reais de complicações da lacuna de percepção.....	220
Lacunas de percepção alimentam falsas suposições .....	220
Lacunas de percepção obscurecem a complexidade do mundo real .....	221
As lacunas de percepção podem colocar equipes umas contra as outras.....	222
Possibilitando a adoção: abordagem estratégica para fechar a lacuna entre software e dados .....	225
Encontrando uma ponte e promovendo a confiança.....	226
Traçando o rumo com informações objetivas.....	227
Evoluindo o shift left dos dados em conjunto com sua organização .....	228
Medindo melhorias na qualidade de vida do engenheiro de software durante a adoção do shift left dos dados .....	229
O que a crescente necessidade de DevSecOps demonstrou .....	231
Ver (e abraçar) o shift left dos dados pelo que realmente é .....	233
Conclusão .....	234

### **CAPÍTULO 10: Gestão de mudanças:O ponto crucial de pessoas, processos e tecnologia ..... 236**

Importância da gestão de mudanças.....	237
Dados como cadeias de suprimentos.....	238
Níveis de implementação dos contratos de dados .....	241
Projetos de aviões e de companhias aéreas.....	241
Problemas orientados ao futuro e ao passado.....	244
Desafios de implementações focadas primeiro na tecnologia .....	245
A armadilha dos padrões .....	247
Requisitos para a implementação bem-sucedida do contrato de dados.....	247
Desenvolvendo estratégia para introdução de contratos de dados.....	249

Definir o problema .....	250
Estruturar o problema.....	251
Priorizar as questões.....	252
Desenvolver um plano de trabalho de análise de questões e realizar as análises .....	253
Sintetizar as descobertas .....	254
Desenvolver recomendações .....	255
Conclusão.....	256
<b>CAPÍTULO 11: Obtendo suas primeiras vitórias com contratos de dados .....</b>	<b>257</b>
Determinando seu primeiro caso de uso de contrato de dados.....	258
Definição de produtos de dados .....	258
Identificando seus fios de aço .....	259
Tomando sua decisão final .....	261
Determinando seu caso de negócios e requisitos.....	265
Colete informações e planeje o engajamento.....	265
Mapeie onde os principais stakeholders se enquadram no ciclo de vida dos dados.....	267
Determine a influência, interesse e riscos dos principais stakeholders .....	268
Obtenha o patrocínio da liderança executiva .....	271
Oito passos para a adoção do shift left na organização inteira .....	273
Estabelecer um senso de urgência .....	274
Criar a coalizão orientadora .....	275
Desenvolver uma visão e uma estratégia .....	276
Comunicar a visão de mudança.....	276
Empoderar colaboradores para a ação de base ampla .....	277
Gerar vitórias de curto prazo .....	279
Consolidar ganhos e produzir mais mudanças.....	279
Fixar as novas abordagens na cultura.....	280
Atribuição em ação: a importância de tornar o sucesso tangível.....	281
Conclusão .....	283
<b>CAPÍTULO 12: Medindo o impacto dos contratos de dados.....</b>	<b>285</b>
O que vale a pena medir? .....	285
Por onde começar se você tiver poucas medições existentes .....	286
Árvores métricas mínimas viáveis.....	288
Medindo ao longo das três fases da adoção .....	290
Meta da fase 1: medição da viabilidade.....	293
Meta da fase 2: medição da repetibilidade.....	295
Meta da fase 3: medição da escalabilidade .....	297
Narrativas para comunicar o impacto da medição .....	298
Partida .....	300
Transformação.....	301
Retorno.....	302
Juntando tudo .....	302
Conclusão .....	304

<b>Conclusão: IA e o futuro dos contratos de dados.....</b>	<b>306</b>
Implicações futuras do shift left dos dados.....	308
Dados como código.....	310
Impacto da IA .....	311
Papel dos agentes .....	312
LLMs e contexto.....	313
Futuro dos contratos de dados.....	314
Próximos passos .....	315
<b>APÊNDICE: Recursos adicionais .....</b>	<b>316</b>
<b>Índice remissivo .....</b>	<b>321</b>

#### **ISENÇÃO DE RESPONSABILIDADE**

Todos os esforços foram feitos para assegurar o fornecimento de informações as mais precisas, completas e exatas possíveis neste livro. Contudo, tais informações são fornecidas “como estão” e sem nenhuma garantia, seja expressa, seja implícita. O autor, a editora, os distribuidores e qualquer entidade envolvida direta ou indiretamente na sua comercialização não assumirão responsabilidade alguma por eventual prejuízo ou dano, direto ou indireto, em razão dos dados contidos neste livro.