
Оглавление

Предисловие	12
Введение	14
Что такое контракты данных	14
Как работать с книгой	15
Условные обозначения, используемые в книге	16
Использование примеров кода	17
Благодарности	18
От издательства	20
О научном редакторе русскоязычного издания	20

ЧАСТЬ I

ВВЕДЕНИЕ В АРХИТЕКТУРУ КОНТРАКТОВ ДАННЫХ

Глава 1. Почему контракты данных нужны именно сейчас	22
Цикл «мусор на входе — мусор на выходе»	23
Современное управление данными	23
Что такое долг по данным	25
Как работает принцип «мусор на входе — мусор на выходе»	27
«Смерть» хранилищ данных	30
Предшествующая эпоха	33
Программное обеспечение поглощает мир	34
Переход к микросервисам	36
Архитектура данных в плачевном состоянии	37
Развитие современного стека данных	39
Крупные игроки	39
Быстрый рост	41
Проблемы в раю	42
Дата-центричный ИИ и зарождение подхода «сдвиг влево» в работе с данными	45
Снижение рентабельности инвестиций при улучшении моделей машинного обучения	48
Стандартизация и массовая доступность разработки моделей в Data Science	49
Превосходство данных над машинным обучением в создании конкурентного преимущества	50
Рост популярности подхода «сдвиг влево»	51
Резюме	52

Глава 2. Качество данных — это не только их чистота	53
Определение качества данных	53
Сравнение OLTP и OLAP и их влияние на качество данных	56
Краткое описание OLTP и OLAP	56
Проблемы преобразования данных между OLTP и OLAP	58
Цена низкого качества данных	61
Оценка качества данных	63
Кто подвергается влиянию	68
Резюме	71
Глава 3. Проблемы масштабирования инфраструктуры данных	72
Чем дата-разработка отличается от разработки программного обеспечения	72
Как разработчики создают продукты	73
Как дата-разработчики создают продукты	74
Основные проблемы современных команд дата-разработчиков	76
Почему для дата-разработки нужна визуальная среда проектирования	82
Стоимость крупномасштабного рефакторинга	84
Особенности крупномасштабного рефакторинга	84
Пример из практики: крупномасштабный рефакторинг в компании Alan	85
Опасности миграции баз данных	88
Роль управления изменениями в обеспечении качества данных	91
Энтропийное поведение данных	91
Как данные отклоняются от установленной бизнес-логики	92
Управление изменениями должно соответствовать потребностям бизнеса	94
Как изменить инфраструктуру в масштабах компании	95
Число Данбара и закон Конвея	96
Пример из практики: команда разработчиков Atlassian	97
Как контракты данных обеспечивают масштабное управление изменениями	98
Резюме	100
Глава 4. Введение в контракты данных	101
Сотрудничество в сфере данных отличается от сотрудничества в других сферах	101
Участники процесса, с которыми вы будете работать	105
Роль производителей данных	105
Роль потребителей данных	107
Влияние производителей и потребителей	110
Трудности и препятствия потребителей данных на пути к качеству данных	113
Альтернатива: использование контрактов данных	117
Этапы внедрения контрактов данных	118
Где внедрить контракты данных	122

Кривая зрелости контрактов данных: осведомленность, ответственность и управление	125
Результаты внедрения контрактов данных	129
Контракты данных в сравнении с наблюдаемостью данных	131
Резюме	134

ЧАСТЬ II

РЕАЛИЗАЦИЯ АРХИТЕКТУРЫ КОНТРАКТОВ ДАННЫХ

Глава 5. Компоненты контрактов данных: дата-активы и определение контракта.....	136
Обзор компонентов	136
Дата-активы	138
Аналитические базы данных	139
Транзакционные базы данных	141
Журналирование событий и потоки событий	145
Собственные данные, сторонняя платформа	148
Определение контракта	150
Спецификация контракта данных	151
Бизнес-логика	156
Реестр схем и каталоги данных	158
Резюме	162
Глава 6. Компоненты контрактов данных: обнаружение изменений и предотвращение нарушений	164
Обнаружение	166
Карантин данных.....	170
Захват данных об изменениях, потоковая обработка и мониторинг данных в реальном времени.....	171
Сквозная история происхождения: графы потоков управления исходным кодом и происхождение данных.....	173
Статический анализ кода.....	177
Предотвращение.....	180
Контроль версий	181
CI/CD.....	181
Мониторинг нарушений и оповещения.....	184
Резюме	189
Глава 7. Внедрение контрактов данных.....	190
Обзор тестового сценария контракта данных	191
Набор данных	192
Приложение для музея	193
Обзор репозитория проекта	194

Реализация контрактов данных	195
Обзор архитектуры контрактов данных	196
Компонент А: дата-активы	197
Компонент В: определение контракта	199
Компонент С: обнаружение изменений	202
Компонент D: предотвращение нарушений	205
Собираем все вместе	207
Сценарий нарушения контракта данных	210
Получен запрос на нормализацию таблицы <code>object_images</code>	210
Обновите файл миграции данных для новой схемы	210
Запуск модульных тестов и проверка на нарушение контракта	212
Отправьте изменения в удаленную ветку и просмотрите логи ошибок	213
Отмена изменений, связанных с миграцией базы данных	214
Обсудите с командой нижестоящего уровня их сценарий использования	215
Создайте новый контракт данных для таблицы <code>object_images_normalized</code>	216
Запустите модульные тесты и проверьте, нет ли ошибки из-за отсутствия актива в каталоге	218
Исправляйте миграцию базы данных до тех пор, пока все проверки не завершатся успешно	219
Отправьте изменения в удаленную ветку и убедитесь, что проверки пройденны	221
Резюме	221
Глава 8. Реальные примеры использования контрактов данных в продакшене	223
История внедрения контракта данных в компании Convooy	224
История внедрения контракта данных в компании Glassdoor	229
История внедрения контракта данных в компании Adevinta Spain	234
Резюме	242

ЧАСТЬ III

ПРИВЛЕЧЕНИЕ ВНИМАНИЯ РУКОВОДСТВА К АРХИТЕКТУРЕ КОНТРАКТОВ ДАННЫХ

Глава 9. Сдвиг влево: изменения в корпоративной культуре, необходимые для контрактов данных	246
Что должно случиться: упущения в концепции «сдвиг влево»	247
Понимание рисков: три реальных примера проблем, связанных с пробелами в восприятии	250
Пробелы в восприятии подпитывают ошибочные предположения	251
Разрывы в восприятии скрывают сложность реального мира	252
Различия в восприятии могут привести к противостоянию команд	253

Обеспечение внедрения: стратегический подход к преодолению разрыва между программным обеспечением и данными	257
Находим общий язык и укрепляем доверие	257
Прокладываем вектор развития с помощью объективных данных	259
Развитие подхода «сдвиг влево» в работе с данными в вашей организации	261
Оценка улучшения в работе разработчиков программного обеспечения в процессе внедрения подхода «сдвиг влево»	262
Что показала растущая потребность в DevSecOps	264
Увидеть (и принять) сдвиг влево в работе с данными таким, какой он есть на самом деле	267
Резюме	268
 Глава 10. Управление изменениями: суть взаимодействия людей, процессов и технологий	270
Важность управления изменениями	271
Данные как цепочки поставок	273
Уровни реализации контрактов данных	276
«Самолетные» и «авиационные» проекты	277
Проблемы перспективы и ретроспективы	279
Проблемы, связанные с внедрением технологий в первую очередь	280
Ловушка стандартов	283
Требования к успешному внедрению контракта данных	284
Разработка стратегии внедрения контрактов данных	286
Определите проблему	287
Структурируйте проблему	288
Расставьте задачи по приоритетности	290
Разработайте план анализа проблем и проведите анализ	290
Обобщите полученные данные	291
Разработайте рекомендации	292
Резюме	293
 Глава 11. Как добиться первых успехов в использовании контрактов данных	295
Определение первого сценария использования контрактов данных	296
Определение продуктов на основе данных	296
Определение ваших «стальных нитей»	298
Принятие окончательного решения	300
Определяем ваш бизнес-сценарий и требования	305
Соберите информацию и составьте план взаимодействия	305
Определите, какое место в жизненном цикле данных занимают ключевые заинтересованные стороны	307
Определите влияние, заинтересованность и риски ключевых заинтересованных сторон	309
Заручиться поддержкой руководства	312

Восемь шагов на пути к внедрению сдвига влево в масштабах всей организации	314
Создание ощущения срочности.....	315
Формирование руководящей коалиции	317
Разработка концепции и стратегии.....	317
Донесение концепции изменений до сотрудников	318
Расширение прав и возможностей сотрудников для масштабных действий.....	319
Достижение краткосрочных результатов.....	321
Закрепление достигнутых результатов и продвижение дальнейших изменений.....	322
Внедрение новых подходов в корпоративную культуру	323
Атрибуция в действии: важность материализации успеха.....	324
Резюме	326
Глава 12. Оценка эффективности контрактов данных.....	328
Что стоит измерить	328
С чего начать, если у вас почти нет готовых метрик	329
Минимально жизнеспособные деревья метрик	332
Измерения на трех этапах внедрения.....	334
Этап первый: измерение жизнеспособности	338
Этап второй: измерение воспроизводимости	339
Этап третий: измерение масштабируемости	341
Повествование для повышения эффективности измерений.....	344
Отправление	346
Трансформация.....	346
Возвращение.....	347
Собираем все вместе.....	348
Резюме	350
Заключение. ИИ и будущее контрактов данных.....	352
Будущие последствия сдвига влево в работе с данными.....	355
Данные как код.....	357
Влияние искусственного интеллекта	358
Роль агентов	359
Большие языковые модели и контекст.....	360
Будущее контрактов данных.....	361
Следующие шаги	362
Приложение. Дополнительные ресурсы.....	363
Об авторах	366
Иллюстрация на обложке	367