

Посвящается Райнеру и Стефане.

– Майкл Саймонс

Посвящаю эту книгу всем тем, кто сегодня страдает от несправедливости, бедности, войн и болезней. Очень жаль, что человечество тратит свое будущее и планету на жажду богатства и власти, вместо того чтобы работать вместе ради созидания более приемлемого мира для всех и каждого. Жертвую бóльшую часть гонорара от этой книги в благотворительные организации, помогающие сделать мир лучше.

– Майкл Хунгер

Оглавление

1	■ Введение в DuckDB	22
2	■ Начало работы с DuckDB	36
3	■ Исполнение запросов на языке SQL.....	48
4	■ Продвинутая агрегация и анализ данных.....	94
5	■ Обследование данных без их сохранения в базе.....	142
6	■ Интеграция с экосистемой языка Python	173
7	■ DuckDB в облаке на основе платформы MotherDuck	201
8	■ Сборка конвейеров данных с помощью DuckDB.....	226
9	■ Сборка и развертывание приложений по работе с данными.....	262
10	■ Вопросы производительности работы с крупными наборами данных...	296
11	■ Заключение	342

Содержание

	<i>Вступительное слово</i>	12
	<i>Предисловие</i>	14
	<i>Признательности</i>	15
	<i>О книге</i>	16
	<i>Об авторах</i>	19
	<i>Об иллюстрации на обложке</i>	21
1	<i>Введение в DuckDB</i>	22
	1.1 Что такое DuckDB?	23
	1.2 Почему стоит обратить внимание на DuckDB?	25
	1.3 Когда следует использовать DuckDB?	26
	1.4 Когда не следует использовать DuckDB?.....	27
	1.5 Варианты использования	28
	1.6 Где используется DuckDB?.....	29
	1.7 Шаги потока обработки данных.....	30
	1.7.1 Форматы и источники данных.....	31
	1.7.2 Структуры данных	32
	1.7.3 Разработка запросов на языке SQL.....	32
	1.7.4 Использование или обработка результатов	33
	Краткий итог.....	35
2	<i>Начало работы с DuckDB</i>	36
	2.1 Поддерживаемые среды	36
	2.2 Установка интерфейса командной строки DuckDB.....	37
	2.2.1 macOS.....	37
	2.2.2 Linux и Windows.....	37
	2.3 Использование интерфейса командной строки DuckDB.....	38
	2.3.1 Инструкции языка SQL.....	38
	2.3.2 Точечные команды.....	39

2.3.3	Аргументы интерфейса командной строки	40
2.4	Система расширений DuckDB	40
2.5	Анализ CSV-файла с помощью интерфейса командной строки DuckDB	42
2.5.1	Режимы показа результатов	44
	Краткий итог	47
3	Исполнение запросов на языке SQL	48
3.1	Краткий обзор языка SQL	49
3.2	Анализ выработки электроэнергии	50
3.2.1	Скачивание набора данных	51
3.2.2	Целевая схема	52
3.3	Запросы на языке определения данных	54
3.3.1	Инструкция CREATE TABLE	54
3.3.2	Инструкция ALTER TABLE	57
3.3.3	Инструкция CREATE VIEW	58
3.3.4	Инструкция DESCRIBE	59
3.4	Запросы на языке манипулирования данными	60
3.4.1	Инструкция INSERT	61
3.4.2	Слияние данных	66
3.4.3	Инструкция DELETE	66
3.4.4	Инструкция SELECT	67
3.5	Специфичные для DuckDB расширения языка SQL	84
3.5.1	Работа с инструкцией SELECT	84
3.5.2	Вставка по имени	88
3.5.3	Доступ к псевдонимам повсюду	89
3.5.4	Группировка и упорядочивание по всем релевантным столбцам	90
3.5.5	Извлечение статистических выборок из данных	90
3.5.6	Функции с опциональными параметрами	91
	Краткий итог	92
4	Продвинутая агрегация и анализ данных	94
4.1	Предагрегация данных при их внесении в базу	95
4.2	Резюмирование данных	97
4.3	О подзапросах	99
4.3.1	Подзапросы как выражения	102
4.4	Группировочные множества	104
4.5	Оконные функции	108
4.5.1	Определение разделов	111
4.5.2	Кадрирование	113
4.5.3	Именованные окна	114
4.5.4	Доступ к предыдущим либо последующим строкам в разделе	117
4.6	Условия и фильтрация за пределами модификатора WHERE	119
4.6.1	Использование модификатора HAVING	120
4.6.2	Использование модификатора QUALIFY	121
4.6.3	Использование модификатора FILTER	122
4.7	Инструкция PIVOT	123

4.8	Использование модификатора ASOF JOIN.....	128
4.9	Использование табличных функций	133
4.10	Использование соединений LATERAL.....	137
	Краткий итог.....	141
5	Обследование данных без их сохранения в базе	142
5.1	Зачем использовать базу данных без сохранения в ней данных?.....	143
5.2	Выявление типа файла и схемы	144
5.2.1	<i>Замечание по синтактико-структурному разбору данных CSV.....</i>	<i>145</i>
5.3	Дробление вложенных данных JSON.....	147
5.4	Перевод формата CSV в формат Parquet	154
5.5	Анализ Parquet-файлов и запросы к ним	161
5.6	Запросы к SQLite и другим системам управления базами данных.....	165
5.7	Работа с файлами Excel.....	170
	Краткий итог.....	172
6	Интеграция с экосистемой языка Python	173
6.1	Начало работы.....	174
6.1.1	<i>Установка пакета Python.....</i>	<i>174</i>
6.1.2	<i>Открытие подключения к СУБД</i>	<i>175</i>
6.2	Использование реляционного API.....	176
6.2.1	<i>Загрузка данных CSV с помощью API на языке Python</i>	<i>177</i>
6.2.2	<i>Составление запросов.....</i>	<i>180</i>
6.2.3	<i>Исполнение запросов на языке SQL.....</i>	<i>186</i>
6.3	Выполнение запросов к кадрам данных pandas	187
6.4	Пользовательские функции	190
6.5	Операционная совместимость с таблицами Apache Arrow и кадрами данных Polars	196
	Краткий итог.....	199
7	DuckDB в облаке на основе платформы MotherDuck	201
7.1	Введение в платформу MotherDuck	202
7.1.1	<i>Принцип работы</i>	<i>202</i>
7.1.2	<i>Зачем использовать платформу MotherDuck?.....</i>	<i>204</i>
7.2	Начало работы с платформой MotherDuck	205
7.2.1	<i>Использование платформы MotherDuck посредством пользовательского интерфейса.....</i>	<i>206</i>
7.2.2	<i>Подключение к платформе MotherDuck с DuckDB посредством аутентификации на основе токена</i>	<i>207</i>
7.3	Максимально эффективное использование платформы MotherDuck	209
7.3.1	<i>Закачивание баз данных на платформу MotherDuck</i>	<i>210</i>
7.3.2	<i>Создание баз данных на платформе MotherDuck</i>	<i>212</i>
7.3.3	<i>Коллективизация доступа к базам данных.....</i>	<i>214</i>

7.3.4	Управление секретами S3 и загрузка данных из корзин S3.....	216
7.3.5	Оптимизация внесения данных в базу и использования платформы MotherDuck	217
7.3.6	Выполнение запросов к данным с помощью искусственного интеллекта.....	219
7.3.7	Интеграции.....	224
	Краткий итог.....	225

8	Сборка конвейеров данных с помощью DuckDB.....	226
8.1	Конвейеры данных и роль DuckDB	227
8.2	Внесение данных с помощью инструмента dlt.....	229
8.2.1	Установка поддерживаемого источника	230
8.2.2	Сборка конвейера	231
8.2.3	Обследование метаданных конвейера	235
8.3	Преобразование и моделирование данных с помощью dbt	236
8.3.1	Настройка проекта dbt.....	237
8.3.2	Определение источников	238
8.3.3	Описание преобразований с помощью моделей	239
8.3.4	Тестирование преобразований и конвейеров	242
8.3.5	Преобразование всех CSV-файлов.....	245
8.4	Оркестровка конвейеров данных с помощью Dagster.....	249
8.4.1	Определение активов.....	250
8.4.2	Выполнение конвейеров.....	252
8.4.3	Управление зависимостями в конвейере	254
8.4.4	Продвинутое вычисление в активах.....	257
8.4.5	Закачивание данных на платформу MotherDuck	259
	Краткий итог.....	261

9	Сборка и развертывание приложений по работе с данными	262
9.1	Создание собственного приложения по работе с данными с помощью Streamlit.....	263
9.1.1	Что такое Streamlit?	264
9.1.2	Сборка приложения.....	265
9.1.3	Использование компонентов библиотеки Streamlit.....	267
9.1.4	Визуализация данных с помощью библиотеки plot.ly	272
9.1.5	Развертывание приложения на облачной платформе Community Cloud.....	275
9.2	Сборка информационной панели деловой аналитики с помощью платформы Apache Superset.....	278
9.2.1	Что такое Apache Superset?	278
9.2.2	Рабочий процесс платформы Superset.....	281
9.2.3	Создание первой информационной панели	282
9.2.4	Создание набора данных из SQL-запроса	289
9.2.5	Экспорт и импорт информационных панелей	293
	Краткий итог.....	294

10	Вопросы производительности работы с крупными наборами данных	296
10.1	Загрузка полной базы данных Stack Overflow и выполнение запросов к ней.....	297
10.1.1	Выгрузка и извлечение данных.....	297
10.1.3	Обследование данных CSV-файла.....	302
10.1.4	Загрузка данных в DuckDB.....	305
10.1.5	Быстрые разведывательные запросы к крупным таблицам.....	308
10.1.6	Распределение постов по дням недели.....	312
10.1.7	Использование перечислений для тегов.....	314
10.2	Планирование и исполнение запросов.....	319
10.2.1	Планировщик и оптимизатор.....	320
10.2.2	Среда выполнения и векторизация.....	321
10.2.3	Визуализация планов запросов с помощью инструкций Explain и Explain Analyze.....	322
10.3	Экспорт данных Stack Overflow в Parquet-файлы.....	326
10.4	Обследование набора данных о нью-йоркском такси прямо из Parquet-файлов.....	329
10.4.1	Конфигурирование учетных данных для доступа к S3.....	330
10.4.2	Автовывявление типов файлов.....	331
10.4.3	Обследование схемы Parquet-файлов.....	332
10.4.4	Создание представлений.....	333
10.4.5	Проведение анализа данных.....	334
10.4.6	Применение набора данных о такси.....	339
	Краткий итог.....	341
11	Заключение	342
11.1	Чему вы научились в этой книге.....	342
11.2	Предстоящие стабильные версии DuckDB.....	343
11.3	Аспекты, которые не были рассмотрены.....	343
11.4	Где можно узнать больше?.....	344
11.5	Каково будущее инженерии данных с использованием DuckDB?.....	345
	Приложение А. Клиентские API для DuckDB.....	347
	Предметный указатель.....	366

Вступительное слово

Дорогой читатель, добро пожаловать в эту книгу о DuckDB. Кажется немного сюрреалистичным осознавать, что я пишу предисловие к этой книге о DuckDB ввиду быстроты произошедших событий на фоне общеизвестного факта, что мир систем управления данными движется очень медленно – программные проекты, начатые в 70-х годах, до сих пор занимают прочные позиции на рынке.

Прошло всего несколько лет с тех пор, как одним вечером 2018 года мы сидели в баре Joost в Амстердаме и решили разработать новую систему. Мы и раньше задумывались над этим, но сомневались, так как понимали, что это безрассудная идея. Банальный здравый смысл подсказывает, что для успешного создания новой системы управления базами данных требуется «десять миллионов долларов». Но мы остановились на не менее безрассудном плане: создать новый тип системы управления данными – такой, каких еще не было, – внутри-процессную аналитическую систему, и, возможно, к этому новому типу систем обычные правила не будут применимы. После еще нескольких кружек пива мы практически определились с первым черновым вариантом архитектуры DuckDB, а уже на следующий день приступили к первым пробам.

Только год спустя, в 2019 году, мы открыли репозиторий и начали рассказывать о нем людям. Мы продемонстрировали первую разработку DuckDB на конференции SIGMOD 2019, которая, по счастливой случайности, проходила в Амстердаме. Поскольку мы были одними из организаторов конференции, мы подсовывали стикеры в пакеты с вкусняшками, проводя одну из первых попыток вирусного маркетинга. Примерно в то же время для общественности был открыт репозиторий исходного кода. Так сказать, «утка выскочила из мешка».

Надо признать, каждый день запускаются тысячи проектов с открытым исходным кодом, и подавляющее большинство из них, к сожалению или нет, никогда не получают развития. Этого ожидали и мы, дескать, скорее всего, никому не будет дела до нашей «DuckDB». Но случилось удивительное: мало-помалу звезды в репозитории на GitHub начали накапливаться. Мы думаем, что это произошло из-за другой конструктивной цели DuckDB: простоты использования. Мы заметили, что преобладающее настроение в области систем

управления данными, по-видимому, опиралось на понимание, что мир, по всей видимости, будет благодарен за то, что ему разрешат использовать с огромным трудом добытые результаты исследований в области систем управления базами данных и систем, которые мы разрабатываем. Однако мы наблюдали тревожный эффект: результаты десятилетий исследований порой просто игнорировались по той причине, что их трудно использовать. В некотором роде меня парадигму систем управления данными, одна из целей разработки DuckDB состояла в том, чтобы сделать ее использование максимально простым и устранить хотя бы некоторые самые большие нарекания, которые мы слышали от практиков.

По всей видимости, люди это заметили. Благодаря активности в социальной сети, ранее именуемой Twitter, и, что особенно важно, благодаря регулярному появлению на Hacker News, был достигнут большой прирост популярности. Теперь DuckDB скачивается миллионами раз в месяц и используется повсеместно – как в крупнейших компаниях, так и в самых маленьких встроенных устройствах. Платформа MotherDuck предлагает практикующим пользователям хостинговую версию в стиле DuckDB с сильным локальным компонентом. О DuckDB даже пишутся книги.

Мы рады, что именно Марк и два Майкла предлагают вам эту книгу. Для нас большая честь, что ее написал такой замечательный коллектив. Являясь экспертами в области сложных технологий работы с данными, они подают материал разработчикам в непринужденной и увлекательной манере, но при этом в глубоко компетентной форме. Будем надеяться, что эта книга вам понравится и, конечно же, понравится работать с DuckDB.

*– Марк Раасвельдт и Ханнес Мюхляйзен,
создатели DuckDB, 2023 год*

Предисловие

Эта книга посвящена DuckDB – современной, быстрой, встроенной аналитической системе управления базами данных (СУБД). Она работает на вашей машине и может легко обрабатывать многочисленные гигабайты данных из различных источников, включая JSON, CSV, Parquet, SQLite и Postgres. СУБД DuckDB хорошо интегрируется в экосистемы языков Python и R и позволяет выполнять запросы к резидентным кадрам данных с нулевым копированием данных. Более того, отпадает необходимость в развертывании облачных складов данных для повседневной обработки данных; вместо этого появляется возможность выполнять DuckDB на своих данных, локально либо в облаке.

С помощью СУБД DuckDB вы сможете без особых затруднений решать свои задачи по анализу реляционных данных. Ее очень удобно использовать и легко осваивать. А самое главное, ее можно использовать подобно СУБД SQLite в средах и приложениях Python. Мы уверены, что при изложении материала о DuckDB мы попали в самую точку, рассказав о ее режиме встроенного интерфейса командной строки, интеграции с Python, возможностях сборки конвейеров данных, а также о безболезненном погружении в современный язык SQL с помощью DuckDB.

Все мы давно являемся экспертами-практиками и преподавателями в области обработки данных. Однако мы пришли из разных уголков этого спектра – графовых, реально-временных столбчатых и реляционных систем управления базами данных, – и при всем при этом именно в DuckDB мы нашли нечто ценное, о чем, на наш взгляд, стоит рассказать. Нам очень нравится использовать DuckDB как вне нашей компетенции, так и в качестве полезного инструмента в соответствующих областях деятельности.

Признательности

Благодарим Джордана Тигани, соучредителя и генерального директора MotherDuck, который был нашим техническим редактором. Благодаря его кропотливой работе и отзывам наши примеры и текст стали намного лучше. Также большое спасибо всем остальным техническим рецензентам, которые усердно работали над главами и оставляли свои отзывы, в особенности Маркосу Ортису, Георгу Хайлеру и Джейкобу Мэтсону за рассмотрение первоначального предложения, а также Дирку Гомесу за проверку всего исходного кода.

Нам очень помогли статьи и объяснения Марка Раасвельдта, Ханнеса Мюхляйзена и Алекса Монахана, которые рассказали о внутреннем устройстве СУБД DuckDB, идеях, лежащих в ее основе, и некоторых упражнениях с языком SQL, о которых мы раньше даже не подозревали. Благодарим вас!

Большое спасибо Райану Бойду и Мехди Кваззе. Они не только дали ценные замечания, но и очень помогли привлечь внимание к книге.

Мы также благодарим редакционный коллектив издательства Manning, в особенности Ронду Мейсон и Джону Генник, за столь позитивное отношение к нам, а также Кристиана Берка за его усердную правку и быструю обратную связь. Майкл Хунгер и Майкл Саймонс рады за наших коллег из Neo4j, которые, возможно, были удивлены, и не раз, но достаточно уверенно признают возможность существования графового и реляционного подходов. СУБД DuckDB оказала на нас большое влияние, когда мы задумались об эмпатии к пользователям нашего программного обеспечения.

Всем рецензентам: Андрею Абрамушичу, Абхилашу Бабу Дьотиндре Бабу, Анджану Бакчу, Крису Больярду, Тьяго Бритто Борхесу, Нандиру Доктору, Дидье Донсезу, Дирку Гомесу, Саймону Хьюитту, Эндрю Джадду, Мадихе Халид, Симеону Лейзерзону, Ноэлю Ллеваресу, Себастьяну Майеру, Эли Майосту, Сумиту Палу, Анупу К. Париху, Саше Санковой, Уильяму Ямир Сильве, Ганешу Сваминатану, Мэри Энн Тигесен, Рохини Уппулури, Анкиту Вирмани, Вонди Вольде и Хэну Чжану – ваши предложения помогли сделать эту книгу лучше.

И конечно же, подобного рода книга писалась в дополнение и параллельно с обязанностями по работе, что отнимает много свободного времени и личной жизни. Поэтому мы благодарим наши семьи за постоянную поддержку наших безумных идей.

О книге

Мы не хотели писать справочник (для этого есть документация). Наоборот, нашим желанием было поделиться тем волнением и радостью, которые мы испытывали при работе с DuckDB, и сделать это таким образом, чтобы вы узнавали что-то новое на каждой странице и получали то же удовольствие, что и мы при написании. Материал данной книги излагается динамично, насыщен информацией, имеет практическую направленность и поучительные, легкоусвояемые примеры, которые несложно применять на практике.

Кто должен прочитать эту книгу

Идеальным читателем этой книги будет инженер данных, исследователь данных или разработчик, которые заинтересованы в эффективном анализе существующих структурных данных без необходимости организовывать инфраструктуру. Он должен быть хорошо знаком с инструментами командной строки и желательно знать немного Python. Мы рассмотрим большое количество исходного кода на языке SQL, начиная с элементарных модификаторов SQL и продвигаясь к продвинутым аналитическим инструкциям. СУБД DuckDB доступна на всех основных операционных системах и не нуждается в установке; достаточно скачать и запустить исполняемый файл. Единственно, в главе о бессерверной аналитической платформе MotherDuck потребуется создать учетную запись, если вам захочется ее опробовать.

Как эта книга организована: путеводитель

В главах 1 и 2 мы начинаем с щадящего введения в DuckDB, рассказывая о ее использовании и месте в современных конвейерах данных. Первым делом мы научим вас использовать интерфейс командной строки DuckDB и только в главе 3 перейдем к введению в язык SQL. Там мы рассмотрим базовые модификаторы и инструкции, а затем в главе 4 погрузимся в мир расширенного анализа данных с помощью языка SQL, используя расширенные агрегации, оконные

функции, рекурсивные инструкции SQL и многое другое. Разумеется, в этот раздел будут включены специфичные для конкретного поставщика и удобные для разработчиков расширения, которые DuckDB привносит для всеобщей пользы.

У СУБД DuckDB есть много достоинств, и одно из них заключается в том, что она не навязывает свое хранилище данных. Всю главу 5 мы посвятили обсуждению вопроса о том, как использовать механизм SQL поверх множества различных форматов файлов для своих собственных целей, не помещая данные в таблицы.

В главе 6 мы подробно рассмотрим интеграцию DuckDB с языком Python, а в главе 7 перейдем к облачным технологиям с использованием платформы MotherDuck.

После этого у нас будут готовы все инструменты для сборки эффективных конвейеров данных (глава 8) и развертывания приложений по работе с данными (глава 9).

В главе 10 мы отступим на шаг назад и обсудим некоторые соображения, касающиеся крупных наборов данных, а также применим все то, чему вы научитесь к тому времени.

СУБД DuckDB предлагает не только интерфейс командной строки и фантастическую интеграцию с языком Python, но и интеграцию с языками Java, C, C++, Julia, Rust и многими другими. Мы их рассмотрим в конце книги, в приложении А, в особенности способы использования DuckDB из Java.

Об исходном коде

Данная книга содержит много примеров исходного кода как в пронумерованных листингах, так и в виде обычного текста. В обоих случаях исходный код отформатирован вот таким шрифтом фиксированной ширины. Иногда исходный код также выделяется **жирным шрифтом**, чтобы подчеркнуть ту его часть, которая изменилась по сравнению с предыдущими шагами главы, например когда к существующей строке исходного кода добавляется новая функциональность.

Во многих случаях изначальный исходный код был переформатирован; мы добавили разрывы строк и переработали отступы, чтобы они соответствовали доступному пространству книжной страницы. В редких случаях даже этого было недостаточно, и листинги содержали маркеры продолжения строки (➡). В таких случаях, возможно, придется удалять лишний пробел, введенный этим маркером, чтобы исходный код заработал, или исправлять длинные URL-адреса.

Вдобавок при описании исходного кода в тексте комментарии к исходному коду нередко удалялись из листингов. Многие листинги сопровождаются аннотациями к исходному коду, выделяя важные концепции.

Исполняемые фрагменты исходного кода можно получить из liveBook-версии (онлайновой версии) этой книги по адресу <https://livebook.manning.com/book/duckdb-in-action>. Полный исходный код примеров книги доступен для скачивания с веб-сайта издательства Manning по адресу <https://www.manning.com/books/duckdb-in-action> и с репозитория книги на GitHub по адресу <https://github.com/duckdb-in-action/examples>.

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте www.dmkpress.com, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу dmkpress@gmail.com; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу http://dmkpress.com/authors/publish_book/ или напишите в издательство по адресу dmkpress@gmail.com.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг, мы будем очень благодарны, если вы сообщите о ней главному редактору по адресу dmkpress@gmail.com. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательства «ДМК Пресс» и Manning Publications очень серьезно относятся к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу электронной почты dmkpress@gmail.com.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

Об авторах



Марк Нидхэм работает инженером по маркетингу продуктов в ClickHouse, где он создает короткие видеоролики и пишет посты о реально-временных складах данных. Он также трудится над удобством взаимодействия разработчиков с продуктами, упрощая процесс запуска, внося изменения в продукт и улучшая документацию к нему.

В течение последнего десятилетия Марк работал в области инфраструктуры данных, сначала в Neo4j над графовыми базами данных, а затем в StarTree над реально-временной аналитикой с помощью Apache Pinot. Последние 15 лет он ведет блог на веб-сайте markhneedham.com о своем опыте написания программного обеспечения и за это время создал немало коротких образовательных видеороликов по темам, касающимся данных и ИИ на Youtube-канале по адресу <https://www.youtube.com/@learndatawithmark>.

Он пишет в твиттере под именем [@markhneedham](https://twitter.com/markhneedham).



Майкл Хунгер уже более 35 лет увлечен разработкой программного обеспечения. Последние 14 лет он работает над базой данных Neo4j с открытым исходным кодом, выполняя различные функции, в том числе в качестве руководителя отдела инноваций и стратегии продуктов для разработчиков. До прихода в Neo4j он консультировал крупные проекты Java и написал немало SQL-кода для баз данных Oracle, Informix и MySQL. В 2006 году он также создал предметно-ориентированный язык SQL под названием Jequel, который впоследствии был объединен с аналогичными проектами.

Как разработчик Майкл наслаждается многими аспектами языков программирования, инструментов и технологий, каждый день узнавая что-то новое, участвуя в захватывающих и амбициозных проектах с открытым исходным кодом, а также в написании книг и статей, связанных с программным обеспечением. В сферу его интересов входят Java, Kotlin, GraphQL, графовые базы данных, генеративный

искусственный интеллект и современная аналитика данных. Майкл выступал на многих конференциях и помогал в организации нескольких из них. Благодаря его усилиям он был принят в программу Java Champions; уже более 12 лет он раз в два месяца пишет колонку «Эффективный Java» для печатного журнала Java SPEKTRUM. Майкл помогает детям учиться программировать, проводя еженедельные занятия по программированию только для девочек в местных школах.

Более подробную информацию о писательской деятельности и проектах Майкла можно найти в его блоге по адресу <https://www.jexp.de>.



Майкл Саймонс является участником программы Java Champions и работает старшим сотрудником отдела программного обеспечения в Neo4j. Он профессионально трудится разработчиком уже более 20 лет. В Neo4j он играет важную роль в интеграции Neo4j в более широкую экосистему Java. До прихода в графовое пространство он работал в немецком коммунальном секторе, используя язык SQL для вычисления и предсказания энергопотребления для крупных немецких операторов транспортных сетей и производителей энергии, задолго до того, как аналитические базы данных получили более широкое распространение. И по сей день ему нравится использовать декларативную природу языка SQL (и, конечно, Cypher), чтобы запрашивать у машин ответы, а не просто давать им инструкции, чтобы те производили результат.

Майкл – известный докладчик на конференциях, много лет соединяющий Java и базы данных – как реляционные, так и графовые. Майкл является автором бестселлера «Spring Boot 2» и соавтором книги «Arc42 на примерах» (Arc42 by Example), посвященной документированию архитектуры программного обеспечения. Он также ведет блог по адресу info.michael-simons.eu.

В свободное время Майкл все еще грезит стать спортсменом-любителем, и когда он не тренируется, готовится к очередному марафону, он использует DuckDB, чтобы документировать свои успехи по адресу biking.michael-simons.eu/history.

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru